University of Trento

Doctoral Thesis

On the effect of experience: An experimental approach to delegation and tax compliance

Author:
Viola Saredi

Supervisors:
Prof. Luigi Mittone
Dr. Matteo Ploner

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Economics and Management

School of Social Sciences

December, 2016

Supervisors

Prof. Luigi Mittone
Università degli Studi di Trento

Dr. Matteo Ploner
Università degli Studi di Trento

Evaluation Committee

Dr. Marco Faillo
Università degli Studi di Trento

Prof. Roberto Gabriele
Università degli Studi di Trento

Prof. Enrico Zaninotto
Università degli Studi di Trento

UNIVERSITY OF TRENTO

Abstract

School of Social Sciences

Doctor of Philosophy in Economics and Management

On the effect of experience:
An experimental approach to delegation and tax compliance

by Viola Saredi

In the field of decision-making under risk, researchers have started to focus on the effect of information acquisition modality on people's decisional process, by means of a comparison between Decision from Description (DfD) and Decision from Experience (DfE). A literature review on the topic is provided in Chapter 1, which analyzes the determinants of the so-called description-experience gap and its translation into the planning-ongoing gap, according to which people tend to overweight rare events under description (or planning) and underweight them under experience (or ongoing decision-making).

In such a framework, Chapter 2 experimentally investigates delegation in risky choices, in a three-party agency framework. Agents build a portfolio for their principals by selecting among prospects that are either fully described or experienced. Nevertheless, principals are given the opportunity to take over control and build their own portfolio by paying a fee. Principals are more efficient and ambitious than agents. Such a higher quality of principals' portfolios is associated to a higher effort exerted in collecting information on risky options. Principals anticipate this performance difference, but pay a control fee that is generally excessive and negatively impacts on their final earnings.

Chapter 3 and Chapter 4 study tax compliance, by providing a comparison between the two information acquisition modalities. Specifically, Chapter 3 serves as an introduction to Chapter 4, as it reviews the main theoretical and experimental literature on tax compliance, by referring to the role of objective, perceived, and weighted probabilities in compliance decisions. Besides this, it provides a novel methodological analysis that justifies the adoption of laboratory experiments as an externally valid tool if sustained by agent-based simulations in the field of tax compliance. Chapter 4 reports on a laboratory experiment designed to explore the presence of the planning-ongoing gap in taxpayers' behavior, by means of a (self) commitment system for compliance. In line with overweighting of rare events - i.e., fiscal audits-, planning induces the majority of people not only to opt for a commitment to tax compliance, but also to actually comply.

Keywords: experimental economics; delegated choices under risk; decision from experience; tax compliance; agent-based simulations; external validity; plannin-ongoing gap

Contents

List of Figures

List of Tables

Introduction and Overview of included papers

Behavioral Economics (BE) emerges from the realization that traditional economic models are too unrealistic. Standard economic theory deals with rational agents. These are supposed to be able to fully identify the available alternatives, their own needs, and the optimal ways in which they can satisfy them, by means of a sometimes mathematically complex cost-benefit analysis. In such an ideal framework, defaults, frames, subjective probability weighting, or learning mode are supposed to have no effect on human choices. Individuals are able to assess probabilities in a consistent manner, are perfectly aware of their stable preferences, rationally update their beliefs, and thus are always capable of maximizing their own interest (Becker, 1976). According to Fischhoff (1988), neoclassical economists have traditionally taken for granted that people optimize their decisions; therefore, their research has never been intended to test the hypothesis that people actually optimize, rather to identify *what* people optimize.

However, in the 1950s researchers started to observe that, in many circumstances, economic agents do not adopt those behaviors that standard economic theory prescribes. This raised the question whether it is correct to assume that individuals are actually able to solve complex decision-making problems by means of sophisticated algorithms. This problem was immediately faced by Herbert Simon, who proposed his hypothesis of *bounded rationality* (Simon, 1955; Simon, 1956). In this way, Behavioral Economics began to emerge as an interdisciplinary approach to human behavior. Nevertheless, this rise was not so immediate and simple. As a matter of fact, in the same period, another proposal was advanced by the economist Friedman (1953): although individuals do not have the formal tools necessary for optimization in decision-making, they behave as if they do so, just like a billiard player makes 'his shot as if he knew the formulas' (Friedman and Savage, 1948). Friedman was skeptical about the possibility of observing how people face decision problems, because individuals could be unaware of the mental processes involved in their actions. Departures from rational maximization could be considered as random errors, which however tend to disappear in the long run. In fact, those who fail to conform to rational behavior are gradually excluded from markets: only rational operators can survive. It is, therefore, evident that in Friedman's view it was pointless to investigate cognitive the aspects of decision-making.

Nevertheless, through the years much research has been devoted to the psychological analysis of the realism and plausibility of standard economics. Probably, the most significant impulse to the development of BE was the emergence in the 1970s of a new branch of psychology called 'behavioral decision making' (BDM). In fact, according to Dawes (1998), what distinguishes BDM from other existing approaches to human decision-making resides in the fact that BDM adopts theories of rational decisions as a reference point from which individuals' observed behavior systematically departs. It took the work of the two psychologists Amos Tversky and Daniel Kahneman to bring BDM to the attention of mainstream economists. They are probably best known for the development of Prospect Theory, presented in the Econometrica paper 'Prospect Theory: An analysis of decision under risk' (Kahneman and Tversky, 1979), and deepened in 'The framing of decisions and the psychology of choice' (Tversky and Kahneman, 1985). The two authors point out that human choices are not always optimal: in this respect, their theory suggests that individuals' willingness to take risk is highly reference-dependent, and affected by probability weighting. Thanks to this, PT has proved its success in accommodating a wide range of anomalies of decision-making (Kahneman and Tversky, 2000), by recognizing that decision-makers assess lotteries in terms of deviations from a reference point and not of absolute wealth state, for instance.

More recently, Prospect Theory has been challenged in the study of uncertainty (for applications, see Hertwig et al., 2004; Hau et al., 2008; Hertwig and Erev, 2009; Gonzalez and Dutt, 2011). Building on the continuum of uncertainty introduced by Knight (1921) in his seminal book 'Risk, Uncertainty, and Profit', an experience-based approach has been proposed to investigate how people make decisions when they are not provided with a full description, but are induced to experience the initially unknown available alternatives. In fact, in general, Decisions from Description (DfD) have been operationalized in the laboratory as choices between monetary gambles: precisely, the PT approach is adopted with non-trivial choice problems that explicitly describe outcomes and associated probabilities (as in Kahneman and Tversky, 1979). In contrast, Decision from Experience (DfE) has been defined by Hadar and Fox (2009) as a situation in which individuals derive an incomplete knowledge of possible outcomes and probabilities from a sampling procedure. Decision from Experience is characterized by repeated decisions for which no objective prior information on payoff distributions is provided. Decision-makers have to rely on the information collected during the iterated trials. In order to investigate the difference between a description-based and an experience-based task, Barron and Erev (2003) explore five different experimental situations on small feedback-based decisions, and find that experience can lead to the opposite results of Decision from Description. This gives birth to the so-called description-experience gap. Nevertheless, it is necessary to consider that DfE should not be considered as an actual challenge to PT, since it may provide a beautiful tool to study not only risk, but also ambiguity - even if its empirical sophistication is still missing in modern ambiguity theory.

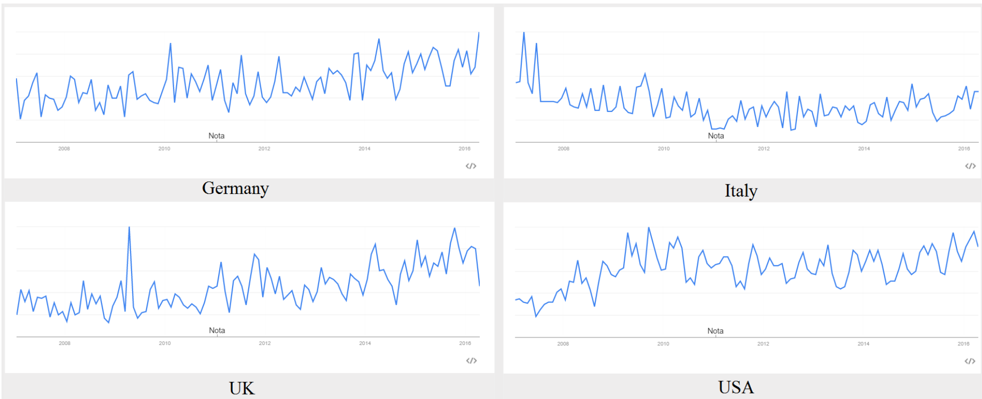
Since then, a number of both psychological and economic contributions have entered the always growing body of behavioral economics literature (see Camerer, Loewenstein, and Rabin, 2011), and this field has become very popular nowadays especially for its evident policy implications. In his article 'The End of Rational Economics', Dan Ariely argues that the 2007-09 financial crisis has negatively affected people's trust in traditional economic theory:

> We are now paying a terrible price for our unblinking faith in the power of the invisible hand. We're painfully blinking awake to the falsity of standard economic theory - that human beings are capable of always making rational decisions and that markets and institutions, in the aggregate, are healthily self-regulating. [...] We are finally beginning to understand that irrationality is the real invisible hand that drives human decision making. It's been a painful lesson, [...]. Armed with the knowledge that human beings are motivated by cognitive biases of which they are largely unaware (a true invisible hand if there ever was one), businesses can start to better defend against foolishness and waste.

In fact, during the last few years, in an increasing number of countries (such as the USA, and the UK), new government institutions devoted to the application of behavioral science in order to redesign and improve public services have begun to appear. Building on both psychological and economic evidence, Behavioral Economics relies on the hypothesis that cognitive biases may prevent people's rational decisions. BE focuses on the real decisions people make, such as deciding whether to save for retirement, whether to cheat and to what extent, and so on.

Nevertheless, some researchers are more skeptical about the efficacy of BE in preventing people from making wrong decisions: as pointed out by Loewenstein and Ubel (2010), sometimes the behavioral sciences have been used 'as a political expedient, allowing policymakers to avoid painful but more effective solutions rooted in traditional economics'. In order to support this claim, the two authors report the failure of the American law on posting calories in restaurants in order to address the obesity issue (see also Hartocollis, 2009).

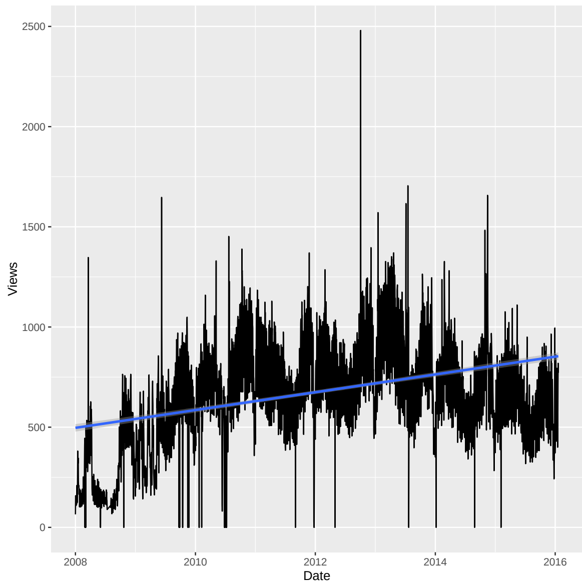
FIGURE 1: Behavioral Economics - Analysis on Google Trends



An analysis of what people are looking for on Google seems to support what Dan Ariely claims: since 2008, in the USA and the UK there has been a remarkably growing interest in understanding what this field is. Figure 1 shows the relative frequency of researches of 'Behavioral Economics' on Google over the last nine years in four different countries, chosen among the most active in this academic field.[1] As the result of Google Trends suggests, there is an undeniable increasing, though oscillatory, frequency (with respect to the overall quantity of researches on Google) in the USA, Germany, and the UK; in contrast, in Italy this trend is not so clear, but, at the same time, we observe a recent renovated interest after an apparent past depression from 2008 to 2011.

Similarly, Figure 2 reports an overview of how many people have been looking for the meaning of BE on Wikipedia over the last nine years in the world. Also in this case, it is evident that, overall, people's interest in understanding the fundamentals of Behavioral Economics has been continuously growing.

FIGURE 2: Behavioral Economics on Wikipedia



The same spread consensus is not always found for experimental economics, which however is one of the main data source of BE. In fact, many economists are skeptical about the external validity and the generalizability of conclusions and inferences about data collected in the laboratory. A laboratory is an artificial context: time is compressed, subjects are asked to make unrealistic and unnatural repeated decisions, and the framing is manipulated by the experimenter. Nevertheless, experimental studies, in the laboratory, as well as in the field, are possible and valuable. In fact, surveys and questionnaires are not always reliable, especially with respect to illegal activities (Slemrod

[1]For the sake of completeness, for the Italian and the German case, the research key 'Behavioral Economics' has also been translated into Italian and German, respectively. Nevertheless, for none of the two countries a difference in searching trends has emerged with respect to the English key.

and Weber, 2012); on the contrary, experiments allow the combination of economic and psychological theory with evidence. Therefore, experiments play an important role concerning research on tax evasion. Anyway, experimental results should not be taken as empirical evidence for people's behavior in real-life situations. As a matter of fact, people behave differently when they know to be observed (Levitt, 2006; Levitt and List, 2007). Field experiments might overcome these limits of the laboratory, but they are not always easily implementable. Similarly, Agent-based Computational Economics (ACE) may contribute to the development of realistic decision-making models: simulations allow researchers to observe how a system evolves when heterogeneous, and sometimes boundedly rational agents are free to interact.

In any case, laboratory experiments may provide a valuable support in gaining some clear and controlled insights on individuals' behavior in clandestine activities (defined as any hidden activity, either legal or not, which could be hardly identified, investigated, and quantified in the field), such as cheating in general, tax evasion, bribing, or even agents' behavior in agency dilemmas. Such experiments allow to carefully study simple dynamics and individual behaviors in a highly controlled setting. In this sense, laboratory findings may contribute to the understanding of the main causes of clandestine activities, and offer a starting point for new policies fighting against these activities with the help of a behavioral economics solid background.

Outline of the Thesis

The present thesis relies on laboratory experiments in order to shed new light on the role played by experience in different decision contexts, also involving clandestine activities. Experience represents the history of a person: an individual undergoing a certain situation, might learn and gather information on the features of that specific situation. In this sense, experience can be thought of as a source of information acquisition, opposed to a descriptive full knowledge. As pointed out by Hertwig et al. (2004), in our everyday life most of us are involved in situations and interpersonal interactions in which parties' information and past experience play a fundamental role. One example is the possible disagreement between doctors and patients, which could be due to the fact that their opinions and decisions regarding risky situations are based on information coming from different sources. Patients can learn from description, meaning that they can search for and read even detailed information on the Web, find official statistics on surgeon consequences, and side effects; the same statistics are available to doctors. However, doctors can also rely on their previous training and personal experience: if few doctors have encountered one of the uncommon side effects, then they might underestimate the probability of such rare events. Situations of this kind have been experimentally studied by Hertwig et al. (2004): the effect of rare events on decision-making under risk depends on how knowledge about their probability is obtained. In line with this, Weber et al. (1993) demonstrate experimentally that, when asked to generate a diagnostic hypothesis, physicians rely on their memory, and recent experience. In fact, the experimentally manipulated availability of a certain hypothesis induces doctors to generate that hypothesis more frequently. More in general, it has been shown that decision-making in repeated games is highly sensitive to recent feedbacks (among others, Erev and Roth, 1998; Cheung and Friedman, 1998; Camerer and Ho, 1999).

However, according to traditional economics, no behavioral difference is expected to emerge if decision-makers personally experience the outcomes of a decision problem with respect to the case of a pure description of these outcomes. As exemplified by Malmendier and Nagel (2016), the effect of living through a depression on financial investment should not differ from the effect of reading about it; or alternatively, the effect of having experienced unemployment on consumption from the effect of knowing your risk of future unemployment. Nevertheless, the authors analyze individuals' expectations about future inflation, and interestingly find that differences in life-time experiences strongly correlate with differences in inflation expectations. To a similar purpose, Simonsohn et al. (2008) experimentally test whether the same piece of information is weighted differently only because it is learned from direct experience: they find that participants' decisions are influenced more heavily by the behavior of players they interact with than by others' behavior that they simply observe.

Building on such an evidence, this thesis investigates how learning from experience might influence individuals' decisions, and aims to provide policy advice.

Chapter 1 introduces the main literature on experience-based decisions, and draws a comparison between Decision from Description and Decision from Experience, which are applied in *Chapter 2-4*. In line with this, we provide a review on the description-experience gap concept (and its more recent translation into the planning-ongoing gap), and on how experience, as a knowledge source, affects individuals' behavior. This chapter is intended to provide an overview of the main experimental literature critically investigating the nature of the gap. It is shown how the existence of such a gap has been constantly challenged. At the same time, however, the relevance of the laboratory adoption of an experience-based paradigm to study real-world phenomena is explained.

Chapter 2 is based on the working paper 'Taking Over Control: An Experimental Analysis of Delegation Avoidance in Risky Choices', joint with Matteo Ploner (CEEL - University of Trento). It reports on a three-party agency problem embedded in delegated risky decisions, by analyzing discrepancies in risk-taking and decision quality. The usual agency dilemma involves the interaction between an agent and a principal: the *agent* is motivated to behave according to his own best interest, and, at the same time, makes decisions on behalf of the principal, who acts as both *employer* and *recipient* of the agent himself. In our study we distinguish between the two roles of the principal and thus identify three parties: the experimenter act as an employer, while the role of agent and recipient are played by experimental subjects. Consider a bank (employer) where a financial advisor (agent) is hired to provide investment services to the bank's clients (recipients), or, alternatively, a hospital where a doctor has to choose the medical treatment on behalf of his patient, for instance.

The experiment studies such an agency conflict by providing a better insight into real world decision-making: besides description-based decisions, in which subjects receive a full description of the decision problem in terms of outcomes and probability distribution, experience-based tasks are introduced, since they present relevant similarities with the settings people encounter outside the laboratory. The experience condition is characterized by an initial lack of information concerning the decision problem; however, subjects have the chance to collect information through sampling (Barron and Erev, 2003;

Hertwig et al., 2004), in order to make better decisions. Such an experimental exploration of delegated risk-taking captures important components of everyday decisions, such as costly acquisition and collating of payoff information (Rakow and Newell, 2010).

Hence, the aim of the paper is twofold: on the one hand, we verify whether the way in which the decision-maker collects information (description vs. experience) affects the outcome of the decision process and a principal's (or recipient's) willingness to delegate; on the other hand, we test whether choices differ systematically according to whether they have direct consequences for oneself or for someone else (self vs. other).

We find that subjects deciding on behalf of others tend to make inefficient investment decisions: principals are more ambitious, and make fewer and less dominated choices, irrespective of the process of information acquisition. While principals adapt their effort to the complexity of the situation, agents are reluctant to collect information to evaluate prospects. Principals predict agents' poor performance and are ready to pay a substantial fee to avoid delegation. However, the fee is generally excessive and negatively impacts on final earnings. This research shows that in making decisions on behalf of others agents' performance might be very poor: not only, agents' choices are not in line with principals' preferences, but they are also characterized by a high degree of inefficiency, maybe due to agents' sloppiness and lack of effort. However, principals' evaluation of agents' performance is even worse: they are ready to give up part of their final earnings and they end up paying an excessive fee in order to avoid delegation.

Chapter 3 and *Chapter 4* add to the literature on tax compliance. Specifically, *Chapter 3* is based on the book chapter 'Taxpayer's Behavior: from the Laboratory to Agent-Based Modeling', joint with Luigi Mittone (CEEL - University of Trento). It serves not only as an introduction to *Chapter 4*, but also provides a novel methodological analysis that justifies the adoption of laboratory experiments as an externally valid tool which might need to be sustained by agent-based simulations. In this respect, the chapter offers a review of the main theoretical and experimental literature on tax compliance, by specifically referring to the role of objective, perceived, and weighted probabilities in compliance decisions. Then, in the field of tax compliance, I recall and extend the mediation approach proposed by Guala and Mittone (2005), by specifically addressing experiments' external validity, intended to bridge the gap between experimental systems and the real domain of application. By relying on human-calibration for the implementation of realistic taxpayers, agent-based simulations help test and investigate human behavior. They tackle the limits of rationality and behavioral homogeneity, which traditionally characterize theoretical and experimental claims. In order to analyze the role of simulations in supporting tax experiments' external validity, three different approaches are identified and discussed. Firstly, I consider models focusing on the macro dynamics among heterogeneous behavioral types identified in the laboratory. From this viewpoint, agent-based simulations allow researchers to implement and manipulate population heterogeneity in a highly controlled manner, so that this, and its interaction with other variables, can be analyzed as a determinant of policies' efficacy. Secondly, I present models analyzing micro behavioral patterns observed in the laboratory. Such a micro perspective allows researchers not only to test and validate experimental findings, but it also helps uncover and understand human cognitive processes and psychological drivers, which cannot be fully investigated in a purely human setting. Lastly, a combined approach is considered: this is intended to address the complexity of the decision environment outside the laboratory, and helps understand the interaction between micro and macro factors. Hence, this chapter provides a possible guidance for the adoption of such a human-agent combination to a policy purpose.

Chapter 4 is based on the journal paper 'Commitment to Tax Compliance: Timing Effect on Willingness to Evade', joint with Luigi Mittone (CEEL - University of Trento). It builds on the experimental evidence of the existence of a significant difference between planning and ongoing decisions in the context of tax compliance (see Mittone, 1997). When asked to plan their actions, people often overweight events with small probabilities while in ongoing (i.e. real time and, in general, repeated) decisions, they tend to underweight these events and thus behave as if they ignored them.

More precisely, the chapter tests experimentally the robust presence of such an inconsistency in tax-payers' behavior, and, based on this, proposes the introduction of a gentle rule of enforcement to sustain tax compliance. In fact, the availability of a mechanism of partial commitment to compliance is experimentally tested in a repeated measurements setting: the inconsistency between planning and ongoing decisions is investigated by offering tax payers the possibility to commit to automatically pay half of the due tax and then to decide whether to pay the remaining part with a discount (in return for the automatic declaration). This allows to investigate whether a (partial and appealing) enforcement system may induce people to commit to tax-compliance in the long-term and to stick with their compliance plan. According to previous results found in the literature, we expect that tax payers, when asked to decide between (immediate full) compliance and (immediate full) evasion, underweight the probability of being audited, while they overweight this event when they need to plan their behavior.

In our research, we identify two main treatments: The Ongoing Treatment - in which subjects can decide every round both whether to adopt the commitment mechanism for that round, and whether to declare their round-earnings - and the Planning Treatment - in which subjects can decide every ten rounds whether to adopt the commitment mechanism for the following rounds, but every round whether to declare their round-earnings. Experimental results confirm that ex-ante evaluation and planning induce the majority of people to adopt a long-lasting commitment - i.e. to choose the condition under which compliance is more profitable - which, in turn, actually fosters tax compliance. In contrast, in the case of ongoing decision-making, we find a less frequent adoption of the (short-term) commitment, and a significantly lower rate of compliance, as if tax-payers perceived evasion as less risky.

Finally, the Section *Concluding Remarks* summarizes the results of the experimental studies, identifies their main limitations and proposes lines of future research.

Chapter 1

Literature Review

1.1 Experience as a learning mode

In our everyday life, we have to face choices: some of them are simple, and require only a negligible amount of information; others have a big impact on our life and, therefore, require much more time and effort. In many circumstances, we make decisions in uncertain conditions: we are not necessarily provided with all the information we need to choose in an accurate way. In this respect, not only the amount of available information, but also the information acquiring mode is fundamental in determining the decision outcome, which, in turn, affects our future satisfaction.

Nevertheless, over the last decades, *Decision from Description* (DfD) has been the most common pardigm applied to the study of decision-making under risk. Decision-makers are usually provided with a description of all outcomes and corresponding probabilities of two (or more) different options; according to these pieces of information, they are asked to select the option they prefer. In a very famous example, Kahneman and Tversky (1979) propose the following problem: people can choose between (A) a 100% chance of winning 3, and (B) an 80% chance of winning 4. Thanks to this, the two authors identify the so-called ‘certainty effect‘: Contrary to EUT, 80% of decision-makers prefer the sure option A over the risky one B, despite the lower expected value. The two authors also claim that many people, by doing so, tend to overweight small probabilities, and underweight moderate and large probabilities. This finding is one of the main behavioral patterns that Prospect Theory can accommodate (Tversky and Kahneman, 1992). Nevertheless, as mentioned in the Introduction, more recent evidence has led to question the reliability and the applicability of Prospect Theory to real life decisions, and a different and more ecological paradigm based on experience has been proposed (Barron and

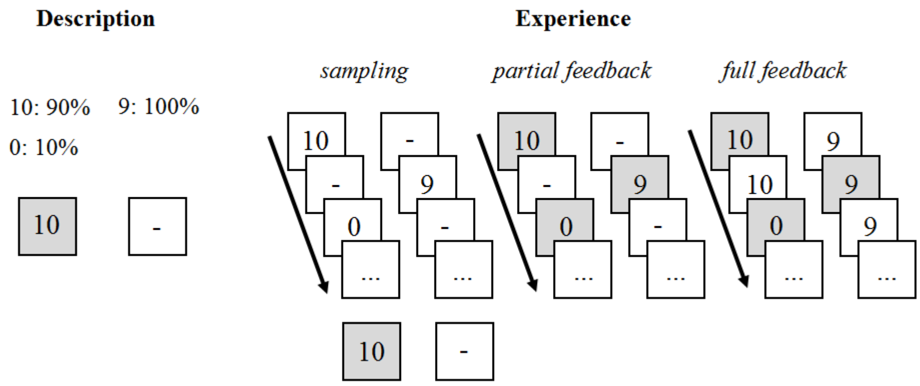
Erev, 2003). Building on this, many experimental works explore situations in which decision-makers face small decision problems called 'small feedback-based decisions'. These problems are characterized by three main properties: (i) decisions are repeated (or experienced repeatedly), (ii) alternatives have small and similar expected values, and (iii) decision-makers do not have any prior information about the payoff distributions, which are, however, simple - i.e., each option contains only one or two possible outcomes, and are kept constant during the experiment. Due to this, such problems are different from those considered in standard decision theory (DfD): as a matter of fact, according *Decision from Experience* (DfE), decisions are not one-shot, outcomes and probabilities are not perfectly known; information is limited to feedbacks concerning the outcomes of previous decisions or explorative trials.

For the experimental implementation, decision-makers face the same problem many times, and, for their decisions, they have to rely on the feedback they collect through sampling. Each available option is represented by a button: by clicking on it, individuals sample an outcome from the underlying distribution, with replacement. In this respect, two different main paradigms have been used (see Figure 1.1). According to the *feedback paradigm* (Barron and Erev, 2003), decision-makers sample from all the available options, in order to gather information on outcomes and associated probabilities. Each sampled outcome contributes to determine the final payoff, which is constantly updated and known to the decision-maker. Therefore, in such a process, the individual has to balance the objectives of exploration - in terms of information gathering - and exploitation - in terms of payoff maximization.[1] These two goals can be distinguished thanks to the alternative *sampling paradigm* (Hertwig et al., 2004), which distinguishes the sampling phase from the choice phase. During the former, the decision-maker is free to repeatedly sample from the available options, but none of the sampled outcomes has actual monetary consequences. The only purpose of this phase is collecting information on the options, so that the individual can make an informed decision. Whenever the decision-maker feels ready, he can stop sampling, and move to the choice phase, in which he

[1] A further specification is required: it is possible to implement either a *partial feedback paradigm* - the decision-maker learns the outcome distribution corresponding to the selected button - or a *complete feedback paradigm* - the decision-maker learns about the outcome distributions of the all the options, but the actual payoff is determined according to the outcome of the selected button.

selects the option he prefers. He discovers his payoff only at the end of the experiment.

FIGURE 1.1: Overview of the choice paradigms designed after Camilleri (2011)



This figure considers two lotteries (one with 90% chance of winning 10, and the other with 100% chance of winning 9), and provides a simple graphical representations of four paradigms, which can be distinguished according to outcome distribution, number of choices involved, and feedback. In the Description condition, full information is provided about the distributions, only one decision is carried out, and the feedback is incomplete, as the individual does not know the outcome of the non selected lottery. As for the Experience condition, outcome distributions are never known, since lotetries are represented by blank buttons; the choice is single only in the sampling paradigm, and the feedback is incomplete in both the sampling and the partial feedback paradigm.

Camilleri and Newell (2011b) provide an interesting experimental test of these paradigms: specifically, they draw a detailed comparison among description, sampling, partial- and full-feedback paradigm, in order to study the relevance of making repeated decisions, and the effect of the exploration-exploitation tension. They find a big difference between the sampling and the feedback paradigm, proving the relevance of consequential actual choices, compared to the pure exploration characterizing the first phase of the sampling paradigm.

Anyway, regardless of the specific paradigm adopted, decision-makers have to explore probabilities and outcomes, by means of repeated draws with replacement from an a priori unknown probability distribution. In fact, according to Camilleri and Newell (2013), a decision from experience is defined as 'a choice situation in which the alternative decision outcomes and their associated probabilities are learned from observing a sequential sample of outcomes over time'. In contrast, a decision from description is 'a choice situation in which the alternative decision outcomes and their associated probabilities are learned from a summary description'.

Over the past few years, such a distinction has become of great interest: researchers have collected substantial evidence showing that decisions differ according to the process of information acquisition. This is the so-called *description-experience gap*, which was initially identified by Barron and Erev (2003), and replicated with a number of different decision problems (Hertwig et al., 2004; Hertwig and Erev, 2009; Rakow and Newell, 2010; Hau et al., 2008; Hau, Pleskac, and Hertwig, 2010; Rakow, Demes, and Newell, 2008; Ungemach, Chater, and Stewart, 2009): under some conditions, preferences elicited by applying DfE contradict PT predictions.

1.2 The Description-Experience Gap

During their first investigation, Barron and Erev (2003) replicate the experiment by Kahneman and Tversky (1979), though moving from description to experience. The basic task is a binary choice between two lotteries referred to as H (= higher expected value) and L (= lower expected value). This task has to be performed 400 times with immediate feedback, and the final payment is determined by the accumulated payoffs (feedback paradigm). Following the experimental structure proposed by Kahneman and Tversky, subjects are given two different problems, yet with no prior descriptive information. In one decision problem, subjects have to choose between 'L: 3 with certainty' and 'H: 4 with probability 0.8; 0 otherwise'. In another problem, which is created by multiplying the probability of winning in the previous problem by 0.25, they have to choose between 'L: 3 with probability 0.25; 0 otherwise' and 'H: 4 with probability 0.2; 0 otherwise'. In DfD, Kahneman and Tversky (1979) identify the *common-ratio effect*, which is an example of the certainty effect: subjects tend to overweight small probabilities; therefore, a nonlinear probability weighting function has to be adopted in order to correctly predict people's decisions. In the first problem, 80% of the subjects prefer option L because they overweight the small probability of the bad outcome 0. Similarly, in the second problem, most subjects overweight the probability of the good outcome 4, and thus only 35% of them prefer option L over H. In contrast, when DfE is applied, in the first problem, the mean proportion of H choices is

63%, but it decreases to 51% in the second problem. Barron and Erev (2003) explain that this result might be due to the *payoff variability effect* that they identify in diverse decision problems. Payoff variability - defined according to expected payoff difference - is higher in the first problem; therefore, we observe a decrease in the payoff variability when moving to the second problem, which impairs expected value maximization. In fact, the proportion of subjects choosing option H is higher in the first than in the second problem.

In line with this, the experimental application of DfE shows that subjects tend to underweight small probabilities in both gain and loss domains: rare events - i.e., those associated to relatively low probabilities - receive less impact than they deserve according to their objective probabilities. Taleb (2007) refers to the extreme case, in which the possibility of rare events is completely ignored, as the 'Black Swan effect'. But how does direct experience lead to probability underweighting? In order to answer this question, it is necessary to take into account two issues: the limited information search (reliance on small samples), and the recency effect.

As for the first issue, the smaller the number of draws from a payoff distribution, the larger the probability that the subject will not encounter the rare event. More in general, small samples cause the rare event to be encountered less frequently than expected (given its objective likelihood), since the binomial distribution for the number of times a particular outcome will be observed in n independent trials when both p (probability of the rare event) and n (= number of draws) are small, is skewed. In fact, in their experimental test which relies on the sampling paradigm, Hertwig et al. (2004) observe that 78% of subjects sample the rare event less than expected. This seems to have a systematic effect on subsequent decisions.

As for the second issue, rare events are less likely to occur right before the actual decision and therefore less likely to affect the final choice. In contrast, common events will tend to be overweighted because they are more likely to be encountered. Hertwig et al. (2004) find that, in general, the second half of the sample sequence offers a better prediction of subjects' subsequent decisions.

As it will be treated more extensively in *Chpater 3* and *Chapter 4*, this finding of probability underweighting can be of great interest in order to analyze and explain subjects' behavior in tax evasion experiments in which a sequence of small feedback-based decisions is involved. Indeed, fiscal audits can be considered as rare events, whose probability is in general underweighted by decision-makers since the audit experience is not frequent. In contrast, when the perception of fiscal audits is artificially manipulated, so that subjects do not consider them as rare events (for instance, because they are experienced rather frequently during the first periods of the experiment), tax evasion is reduced.

However, despite these evident biases affecting decisions, people tend to rely on small samples for three important reasons, analyzed by Marchiori, Di Guida, and Erev (2013).

1. People have cognitive limitations: as replicated in the laboratory, people are presented information for a short period, and they are not able to store and perfectly recall the whole amount of information they collect through sequential sampling. Therefore, in repeated decisions people might be highly sensitive to recent feedbacks (among others, Erev and Roth, 1998; Camerer and Ho, 1999; Cheung and Friedman, 1998).

2. In order to gather information, people have to bear a psychological cost. In this sense, reliance on a small sample might be optimal, because it allows to minimize the searching effort. This factor is studied in *Chapter 2*, in which DfE is intended to replicate an agency dilemma in delegated decision-making under risk.

3. When evaluating the probability that a certain event occurs, people usually look for a subset of previous experiences, which are comparable to the situation they are facing.

Referring back to the characteristics of the description-experience gap, experimental evidence seems to confirm the emergence of loss aversion under both DfD and DfE: risky options minimizing the probability of losses are more attractive than options maximizing expected payoffs. Nevertheless, Erev, Ert, and Yechiam (2008) find contradictory results. In some cases, subjects deviate from maximization because of loss aversion, while in many other cases subjects' choices are led by reliance on small samples and diminishing payoff sensitivity. Subjects do not make cautious decisions because they have an insufficient sensitivity to large potential losses.

As this review shows, over the last years, a growing number of researchers have tested the description-experience gap in a variety of decision contexts. For instance, in a recent contribution, Wulff, Hills, and Hertwig (2015) report on a laboratory experiment testing the effect of two different learning modes people can face when trying to acquire information on online products. People can either read a summary *description* consisting of the mean of the available consumer ratings, or sample individual reviews. Formally, both formats present identical distributional information, but they differ in the way users experience that information. In fact, summaries present complete information in one descriptive format (DfD), while individual ratings require a sequential search (DfE). The authors prove the existence of a substantial decision gap due to the difference between the two learning modes. In fact, they observe that, when searching through individual reviews, people tend to rely on small samples of ratings, and, at the same time, they overweight recently sampled information.

To a similar purpose, together with prof. Matteo Cantamesse (Catholic University of the Sacred Heart), I am currently conducting a research that applies a similar comparative approach to the investigatigation of the interaction between summary rates and sample numerosity in the process of evaluating a product or a service from online reviews. In fact, not only the number of reviews a user actually reads, but also the numerosity of available reviews can have a relevant impact: if fewer reviews are available, then it is more likely to observe skewed distributions of ratings, or, at least, less representative distributions. This implies that some products or services may exhibit a higher ranking, only because they also have a lower number of reviews. This kind of investigation would provide relevant insights for a better understanding of how people integrate these two pieces of information, i.e. reviews' numerosity and average rating.

Despite the interesting issue posed by the previous experimental example (see Wulff, Hills, and Hertwig, 2015), in which two learning modes are available, it is important to highlight that, in many situations, we cannot rely on a full description of the outcomes of the problem. On the contrary, we need to rely on our empirical previous observations, i.e. on our or others' experience. In this respect, examples provided in the Introduction justify the research interest in systematically studying the existence of an experience effect on decisions, compared to the traditional descriptive setting. As pointed out, the fact of not having a solid, common, and reliable informational background has relevant consequences: namely, (i) according to the specific experience, people can rely on misleading samples (i.e., observed samples are not representative of the true underlying outcome distribution); (ii) most recent experience can be overweighted; (iii) people might not be able to build an accurate representation of the frequencies of possible outcomes.

For all these reasons, some authors have extensively studied and tested in the laboratory the effect of sampling biases, the role of repeated choices, and the role of probabilistic representation. Hence, their first aim is to verify whether the description-experience gap can be entirely due to reliance on *misleading samples*, and thus it can be simply explained as a statistical phenomenon. In order to understand to what extent the description-experience gap can be caused by a sampling rather than different probability weighting, Fox and Hadar (2006) reconsider and replicate the experimental analysis conducted by Hertwig et al. (2004). They find that, when controlling for misleading sampling, results can be explained by Prospect Theory. To the same purpose, Rakow, Demes, and Newell (2008) introduce a condition in which subjects' sampling process is not free but yoked in order to control for sampling bias and recency effects. According to their results, authors reject the necessity of two separate theoretical models, one for description- and the other for experience-based choices. This is also supported by Camilleri and Newell (2009) and Camilleri and Newell (2011a): they investigate the effect of sampling variability, which is inevitably implied by the DfE paradigm. The two authors focus their analysis only on those trials in which subjects experience a distribution resembling the true probability distribution underlying the available options: in these cases, no difference emerges across modes of information acquisition. Therefore, the two authors point out that the difference across decision patterns is mainly due to non-equivalent information at the point of choice - i.e., to the specific sample a decision-maker experiences before making a decision. Nevertheless, Camilleri and Newell (2011a) also admit that description-based models are insufficient to investigate many real-world decision problems that involve searching, learning, and iterative beliefs updating, for instance.

Besides this, Ungemach, Chater, and Stewart (2009) present an experimental analysis showing that the description-experience gap cannot be simply regarded as a statistical phenomenon. In fact, they force participants to sample 40 times from each option; participants are free to choose the sampling order, and also the outcome order is random. However, the authors build the sampling process such that each participant can rely on a sample that is representative of the underlying distribution. They report that, even under these conditions, underweighting emerges, and it is due neither to recency nor to judgment error. A similar conclusion is drawn by Hau et al. (2008) and Camilleri and Newell (2011b), who test whether the reliance on very large samples - i.e., on samples that are likely to resemble the true underlying distributions - cancels out the description-experience gap. Interestingly, they find that decision-makers treat and weight probabilities in a different manner according to the information gathering mode.

In summary, experimental evidence seems to show that, if people receive perfectly equivalent information under the two conditions, then the gap disappears. However, at the same time, this seems not to be always true. Furthermore, description-based choices sometimes are highly artificial: in real world decisions, people might not have full information on a given problem, and they might need to rely on their previous, and inevitably potentially biased, experience.

Finally, besides this investigation of the effect of sampling biases, researchers have also addressed the issue of information storage, and the issue of probability representation. As for *memory*, Hertwig et al. (2004) show that, according to the 'narrow window hypothesis' (Kareev, 1995; Kareev, 2000), people tend to rely on a limited number of sampled outcomes, which are condensed among the latest sampled items. Based on this, the memory order effect seems to be relevant in explaining the gap. However, in contrast to this, Camilleri and Newell (2011a), Hau et al. (2008), and Ungemach, Chater, and Stewart (2009) find that the gap emerges even when memory can play no role at all. Therefore, it seems that memory might be relevant, but it cannot be named among the main determinants of the gap.

As for the role of *probabilistic representation*, Fox and Hadar (2006) suggest that probabilities might be estimated according to the information format. For this reason, Camilleri and Newell (2009) run an experiment investigating probability judgment. Participants are asked to estimate the probability of each outcome of the different options they face: they find that people tend to overestimate probabilities of rare events, especially in a description-based task. In a further experimental exploration, Camilleri and Newell (2011a) confirm these findings: they reject the hypothesis that the effect of presentation format on choice is directly mediated by its effect on probability estimates. These results are in line with many previous observations: people tend to exhibit conservatism, which implies overestimation of small probabilities (Kahneman and Tversky, 1979). Fischhoff et al. (2000) report that teenagers overestimate the chance of death in near future; similarly, Americans overestimate the probability that a smoker develop a lung cancer (Viscusi, 1992).

Nevertheless, at the same time, in the experimental and empirical literature, it is possible to identify underweighting of rare events as a robust phenomenon, because of people reliance on small samples. One of the first empirical evidence is provided by the Hungarian physician Semmelweis, who demonstrated that routine hand washing can substantially reduce the death rate of maternity patients and of their doctors. Although his experimental findings were published, no increase in hand washing was observed. In fact, in many circumstances people behave as if they ignore rare events, because they believe that such events cannot happen to them. Therefore, as pointed out by Cooper and Kagel (2003), this evidence suggests that experience might reduce event weighting even when subjects, in order to make their decisions, can rely on a summary description (decision under risk); therefore, in such a case, feedback on previous occurrences - i.e., independent realizations of a similar event - do not actually add any relevant information (Yechiam, Barron, and Erev, 2005; Jessup, Bishara, and Busemeyer, 2008).

In this respect, Barron and Yechiam (2009), and Marchiori, Di Guida, and Erev (2013) focus on the distinction between judgment and choice, and provide experimental evidence on the so-called *overestimating-underweighting paradox*. According to this, in an experience-based judgment and decision, people overestimate rare events, but then they behave as if they underweighted them, exactly like smokers exhibit a low sensitivity to the risk of developing a lung cancer, but, according to their elicited subjective estimates, they actually tend to overestimate this probability (Viscusi and Hakes, 2008). Marchiori, Di Guida, and Erev (2013) show that, even when people are provided with full and certain information on a given decision problem (for instance, in terms of objective probabilities), they exhibit a systematic underweighting, mainly because their decisions depend on the information collected by means of small and noisy samples of past experiences.

1.3 The Planning-Ongoing Gap

A closely related phenomenon is the so-called *planning-ongoing gap*, which can be somehow considered as a translation of the description-experience gap.[2] According to this, when asked to plan and evaluate in advance their future actions, people often overestimate (and overweight) rare events; while in ongoing - i.e. real time, and, in general, repeated - decisions, they tend to underweight these events, and thus to behave as if these could be completely ignored (Yechiam, Barron, and Erev, 2005; Schurr, Rodensky, and Erev, 2014b). According to Schurr et al. (2012), people are more risk seeking and spontaneous when asked to make sequential and distinct evaluations or decisions (ongoing); in contrast, their decisions are less affect-driven, when a series of problems are evaluated jointly (planning).

This underweighting tendency has been identified in many diverse contexts: drivers pass on a two-lane road (Harris, 1988); people back up their computers less frequently than suggested; people buy radios with a detachable front panel, but then they do not detach it after parking. Zohar and Erev (2006) analyze the problem of rare event underweighting by explicitly referring to the spread lack of compliance with safety rules in many organizations. As a matter of fact, this kind of setting can be considered as an appropriate field representation of experience-based decisions: employees repeatedly make the same decision, and, thanks to the feedback they receive, learn how dangerous a certain behavior might be. The two authors point out that employees often seem to be willing to endanger their own well-being, and thus organizations

[2]People overweight rare events when relying only on description (Kahneman and Tversky, 1979), but the opposite pattern emerges when they rely on experience (Barron and Erev, 2003; Hertwig et al., 2004).

need to invest resources to promote employees' compliance with safety rules. In fact, when actually facing the compliance decisions, employees might be affected by the melioration bias (Herrnstein et al., 1993) - the immediate benefit of an option is emphasized - and the rare event bias (Hertwig and Erev, 2009) - events associated to a negligible probability are disregarded. According to this, employees tend to opt for the unsafe behavior: on the one hand, they observe the immediate and tangible benefit of non complying with rules (such as, wearing safety glasses, helmets, or ear plugs might require time, and reduce the comfort); on the other hand, they underweight the probability of an accident. In addition, this unsafe behavior is reinforced by positive feedback, in case no accident occurs while not complying with safety rules.

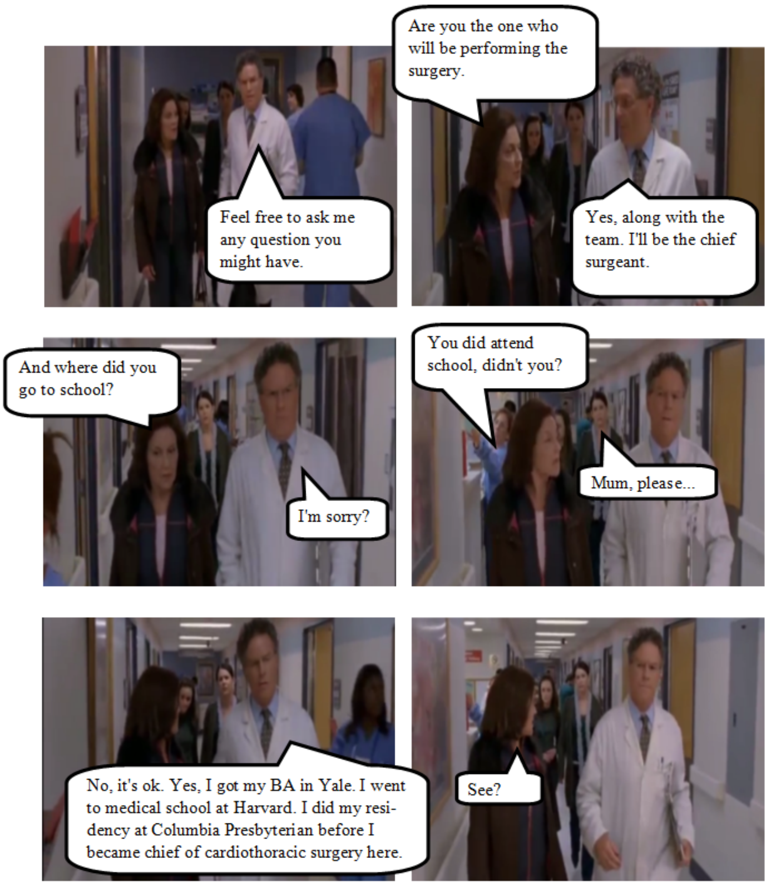
This evidence shows how experience might lead to inefficiency: in fact, according to the previous example, though knowing the huge cost of not complying with safety rules, employees prefer to adopt a reckless behavior. In lottery terms, this seems to imply that experience causes a deviation from the maximization of expected returns (Erev and Roth, 2014). In this respect, an interesting way of addressing this issue is proposed by Schurr, Rodensky, and Erev (2014a), and Erev et al. (2010): in their experimental research, they study problems related to not obeying safety rules in the workplace. They observe that workers are aware of the likelihood of rare events (i.e. accidents), and thus of the importance of such rules, but that, at the same time, they actually tend not to comply with safety rules, as if they underweighted rare events. Building on this, they suggest how the planning-ongoing gap can be exploited in order to offer a solution to employees' unsafe behavior: it is possible to build an enforcement system on the fact that workers, when asked to plan and think about the relevant risks in their workplace, state they want to behave safely. Specifically, the research by Erev et al. (2010) focused on doctors and nurses' frequent violation of the safety rule 'use protective gloves while drawing blood from patients'. The authors reported that many workers that violated the rule were planning to obey it. In this framework, a gentle enforcement program was introduced to help workers behave in accordance with their plans: workers were asked to remind the rule to their colleagues. The implementation of this program led to an increase in the use of protective gloves. Another demonstration of the potential of this approach can be found in the study by Schurr, Rodensky, and Erev (2014a) on the use of protective gear in 11 midsize factories in Israel. Also in this case, a gentle enforcement system is used: Supervisors are asked to approach those workers who do not obey the safety rules, remind them how risky this behavior is, and record violations. Results show an increase in the rate of safe behaviors from 60% to more than 90%.

In a similar manner, *Chpaer 4* provides an experimental investigation of the planning-ongoing gap in the research field of tax compliance; at the same time, it tests the introduction of an enforcement system, intended to fight against evasion temptation, by relying on planned compliance.

Chapter 2

Taking Over Control: An Experimental Analysis of Delegation Avoidance in Risky Choices

joint with Matteo Ploner



2.1 Introduction

Traditionally, economic research on risky choices focused on the direct consequences for the decision-maker. Only more recently has attention shifted to contexts in which an agent chooses on behalf of a principal. From a review of the relevant literature on delegated decision-making a complex picture emerges: in general, individuals choose differently when deciding on others' money rather than on their own, but no clear and consistent tendency in the shift of risk propensity has been identified. Discrepancies across studies could be due to a mixture of idiosyncratic experimental factors - such as the risk preferences elicitation method, individual risk preferences, or the incentives structure - which prevent from drawing clear conclusions about the effect of delegation on risk-taking. Pollmann, Potters, and Trautmann (2014) and Chakravarty et al. (2011) find that individuals exhibit less risk aversion when choosing for an anonymous stranger than for themselves. Though opting for a different elicitation method, Agranov, Bisin, and Schotter (2014) and Hsee and Weber (1997) provide the same evidence of a decrease in risk aversion prompted by delegation. Andersson et al. (2014) argue that such a higher risk propensity when choosing for others may be prompted by a decrease in loss aversion. On the other hand, Eriksen and Kvaløy (2009) and Kvaløy, Eriksen, and Luzuriaga (2014) find that people are more risk averse when dealing with other people's money. This evidence is confirmed by Bolton, Ockenfels, and Stauf (2015) and Pahlke, Strasser, and Vieider (2015), who however also observe that individuals tend to be less risk averse when managing other people's money in the loss domain.

This research interest is motivated by the fact that in several domains individuals make choices whose consequences are barely predictable, but which nonetheless have important implications for others. Choices of this kind do not always imply a direct link between the outcomes of the delegated choice and the economic returns of the delegated agent. At the same time, however, the immediate recipient of the delegated agent's decisions might not coincide with the agent's employer. As an example, doctors working in a hospital choose medical treatments on behalf of their patients, and bank financial advisors define the composition of their clients' portfolios. In such a context, how does a recipient's willingness to delegate is affected by an agent's source of knowledge?

In our research, we adopt the perspective of the aforementioned studies: delegated agents make decisions for others without a direct economic stake in the decision. Nevertheless, previous studies consider only the point of view of the delegated agent and compare decision-making for oneself and for others only in terms of risk propensity. Generally, no assessment of the efficiency of the decision in terms of the trade-off between risk and returns is made. To the best of our knowledge, our paper is the first to explore whether principals - i.e., passive recipients of agents' decisions - perceive delegation as efficient and whether they are actually willing to delegate.[1] Arora and McHorney (2000) and Levinson et al. (2005) study people's preferences for delegation and participation in medical decisions, and find that patients' demographic variables are among the main determinants of delegation. Nevertheless, they do not focus on how the doctors' source of knowledge (previous academic training vs. direct working experience) might affect patients' willingness to delegate.

The present experimental research investigates the agency problem that characterizes delegated decisions, in which three parties are involved and the principal affected by agent's performance differs from the one who has hired the agent and set the monetary incentive structure to be adopted. [2] In a context where recipients and agents' incentives are not aligned and the source of knowledge might be different, we consider both the agent's behavior, in terms of risk-taking and decision quality, and the principal's expectations about the delegation. The experiment allows us to study conflicts in agency by providing a better insight into real world decision-making (Koritzky and Yechiam, 2010): it combines the research on three-party agency dilemmas and the comparison

[1] Such a willingness is also investigated in the questionnaire and laboratory study by Botti and Iyengar (2004). Participants evaluate two different sets of imaginary dishes: one consisting of four sumptuous entrées, and one of four revolting dishes. They observe that, in the case of more attractive entrées, choosers (i.e., those selecting a dish for themselves) show a higher anticipated satisfaction than non-choosers (i.e., those asked to imagine to eat a dish chosen by someone else) do. In contrast, in the case of less attractive entrées, choosers' anticipated satisfaction is lower than non-choosers'. Nevertheless, this result holds only when the choice options are more differentiated Botti and McGill (2006), or when actions have an hedonic goal (i.e., they are made for one's own pleasure) and not a utilitarian goal Botti and McGill (2011).

[2] Following, the term *principal* will be used to refer to the principal-recipient and not to the principal-employer, whose role is played by the experimenter.

between *description-based*[3] and *experience-based*[4] tasks, which present relevant similarities with decision settings people encounter in the real world.[5] The aim of the paper is twofold: on the one hand, we verify whether the way in which the decision-maker collects information (description vs. experience) affects the outcome of the decision process and a principal's willingness to delegate;[6] on the other hand, we test whether choices differ systematically according to whether they have direct consequences for oneself or for someone else (self vs. other).

In many instances, people facing a decision problem may rely both on their knowledge from previous training and on their experience. Benjamin and Budescu (2015) investigate how the learning mode (either by description or by experience) affects advice giving and taking: they find that advisers learning from description are more confident in providing information, and, in line with this, advice from description is, in general, preferred by decision-makers. According to the risk/uncertainty taxonomy by Knight (1921),[7] both description- and experience-based choices can ultimately be considered as decision-making under risk: even if probabilities are not explicit, they are still measurable (Hau, Pleskac, and Hertwig, 2010). Risky decisions from experience occupy a middle ground (Rakow and Newell, 2010). Initially, the probability distribution is not known, but it can be determined through sampling. In this context, the degree of experience is defined by the size of the experienced sample: despite practical difficulties in computations, if people decrease the degree of uncertainty

[3]This approach is based on prospects explicit and full description (in terms of outcomes and associated probabilities), and it is the most commonly used for problems involving monetary gambles.

[4]This approach (Barron and Erev, 2003; Hertwig et al., 2004; Rakow and Newell, 2010) is characterized by repeated decisions on monetary gambles, and lack of objective prior information on outcome distributions. Decision-makers have to rely on the information they collect during the iterated trials.

[5]To list just a few examples: vaccinations recommendations (Hertwig et al., 2004), daily decisions to use safety devices (Yechiam, Erev, and Barron, 2006; Erev, 2007), evaluation of innovation (Rakow and Miler, 2009), and reaction to possible disasters (Yechiam, Barron, and Erev, 2005; Weber, 2006).

[6]In both Description and Experience, people usually choose between two (risky) options. Previous studies comparing these two learning modes in self decision-making show that, in descriptive settings, people tend to overweight rare outcomes, while they underweight them in experiential settings (e.g., Gonzalez and Dutt, 2011; Hertwig et al., 2004; Weber, Shafir, and Blais, 2004; Hertwig and Erev, 2009).

[7]Knight (1921) introduces a continuum of types of uncertainty/probability, characterized by different *degrees of uncertainty*: risky situations where probabilities are defined precisely are opposed to situations where only estimation can form one's beliefs.

sufficiently, they can determine a priori probabilities with precision.[8]

We present a detailed exploration of the principal-agent relationship in the context of delegated risk-taking that captures important components of everyday decisions, such as costly acquisition and collating of payoff relevant information (Rakow and Newell, 2010). If no specific assumption about agents' qualification and trustworthiness is made, principals' expectations can play a fundamental role: the lack of trust in agents' commitment may hamper the emergence of potentially fruitful delegation relationships.

In the experiment, principals build a portfolio of prospects for themselves (*Self*) and, simultaneously, agents build a portfolio for their principals (*Other*), choosing from the same set of lotteries. Prospects are either presented to participants in a conventional way, as distributions of probabilities over outcomes (*Description*), or are experienced by participants (*Experience*). As the study is intended to purely address the issue of decision-making on behalf of someone else and not to investigate the role of monetary incentives, agents have no stake in the choice they make on behalf of their principals: they earn a fixed amount irrespective of their actual decisions. Therefore, the agents' choices we observe are not affected by any monetary concern. Under the canonical assumption of selfish maximization, the choices of the agents are expected to be quite erratic. Agents have no pecuniary incentive to implement a coherent choice plan and this would reflect in mindless choices, especially when the choice process is cognitively more demanding (i.e., *Experience*). However, previous evidence about delegated risky decisions (see the review above) and the documented existence of widespread other-regarding preferences (e.g., Camerer, 2013) suggest that concerns for principals' welfare would encourage agents to make choices that, at the very least, do not explicitly harm the principal. Andersson et al. (2013) find that a pro-social orientation moderates agents' propensity to risk others' money, even when incentives to increase the risk on behalf of others are introduced.

Within this setting, we analyze the effort of principals and agents in reducing the degree of uncertainty of prospects and the efficiency of portfolios

[8]Some studies explore the relationship between the size of the Description-Experience Gap and the extent of the information gathering process (Hau et al., 2008; Hau, Pleskac, and Hertwig, 2010; Ungemach, Chater, and Stewart, 2009).

in terms of mean/variance.[9] Furthermore, we allow principals to retain full control over the portfolio composition and we measure their willingness to pay to avoid delegation. More precisely, our research is structured around the following research questions:

Q1 - Does the risk content of portfolios change across experimental conditions that differ in roles and information acquiring process?

Q2 - Does the efficiency of portfolios change across experimental conditions that differ in roles and information acquiring process?

Q3 - Do principals exert more effort in reducing uncertainty than agents?

Q4 - How do principals' expectations and attitudes affect the willingness to delegate? What drives principals' desire to take over control?

We find that (1) portfolios built by principals are more ambitious in terms of mean/variance irrespective of the learning mode. In general, (2) participants build more efficient portfolios under description; this is especially true for principals, whose portfolios are caharecterized by a higher degree of efficiency. In addition, (3) principals adapt their effort to the complexity of the task more than agents. The lack of effort of agents and the inferior quality of portfolio delivered is anticipated by the principals, who pay excessively large fees to retain control over their earnings. They exhibit the strongest willingness to retain their control when they decide under description and agents under experience (4).

2.2 Methodology

We observe how principals and agents build a risky portfolio under different decision settings (*Description vs. Experience*), and we measure principals' willingness to pay in order not to delegate by means of a random price mechanism (Becker, DeGroot, and Marschak, 1964).

[9]To define the efficiency of a portfolio we refer to a mean/variance dominance criterion. A portfolio is more efficient than another if for a given expected return it has a lower variance or, alternatively, if for a given variance it delivers higher expected returns.

2.2.1 Experimental Task

Subjects can play either the role of principal (*Self* decision-making) or the role of agent (*Other* decision-making). At the start of the experiment, each participant is informed about his role and is randomly and anonymously matched with another participant.

Principals are asked to build a portfolio for themselves, while agents are asked to perform the same task for their principals. Before choices are implemented, every principal states his willingness to pay to retain his portfolio, instead of replacing it with the one built by their agent (see Figure 2.3).[10] Subjects build portfolios by selecting risky options from three multiple price lists (MPLs).[11] Each list involves 10 decisions between a Leftward and a Rightward prospect, with the former being safer than the latter. The general structure of each prospect is $P = [L, p; H, 1-p]$ with $0 < L < H$ and $p > 0$. Expected values do not vary across lists, while probability distributions go from $p = 0.5$ gambles (less risky and very easy to understand) to $p = 0.7$, and finally to $p = 0.9$ gambles (characterized by higher degree of risk and rare events).[12] Participants in both roles are asked to build two 30-prospect portfolios labeled A and B. Prospects in portfolio B are characterized by larger differences in the standard deviations of the Leftward and Rightward prospects. In the following, we adopt the letter of the portfolio and the probability p as labels to identify blocks of 10 prospects. As an example, $A.5$ uniquely identifies prospects of portfolio A that assign probability 50% to the lowest outcome. Tables 2.10 and 2.11 in the Appendix contain a detailed description of the prospects.

The aforementioned task is common to all the experimental treatments: the manipulation refers to the way in which prospects are presented. Under

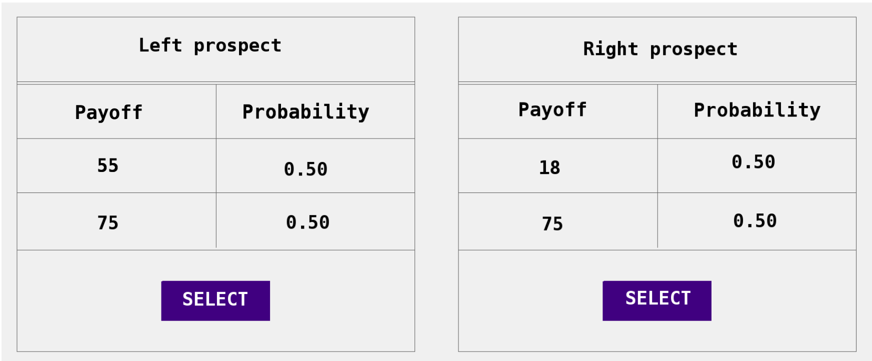
[10]Bids are collected via a standard BDM procedure: bids must lie between 0 and 1000 ECU, to be deducted from principals earnings. The BDM screen can be found in Appendix A.

[11]The choice structure of the present experiment can be considered as a version of the multiple price list (MPL) format. The only difference is represented by the fact that our participants do not go through a sequence of three prospect list screens. Instead, they are displayed a screen for each couple of prospects they are sequentially asked to evaluate. This was necessary for the implementation of the treatments involving experience. Nevertheless, couples are always presented in the same order.

[12]The presence of rare events (i.e. events associated to small probabilities) is relevant for the investigation of the Description-Experience gap in the context of self-other decision-making: it is characterized by overweighting of rare events in case of description - i.e., a higher degree of risk-taking (Tversky and Kahneman, 1992) - and underweighting of rare events in case of experience - i.e., a lower degree of risk-taking (Hertwig et al., 2004).

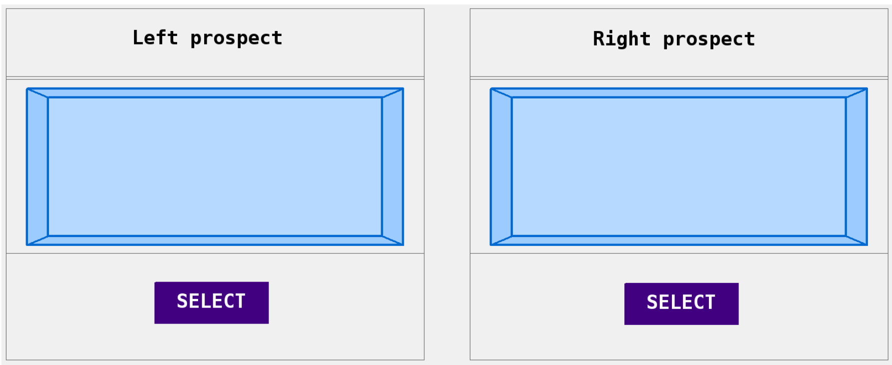
Description, the typical decision screen (Figure 2.1) shows only two gambles and no reference to previous decisions; it includes all the relevant information on the two prospects, so that participants knowing both outcomes and probability distributions can compare them.

FIGURE 2.1: Typical Decision Screen - Description



Under *Experience* (Figure 2.2), subjects can collect information on each couple of prospects and select the one that they prefer. The well-established sampling paradigm is adopted (Barron and Erev, 2003; Hertwig et al., 2004).

FIGURE 2.2: Typical Decision Screen - Experience



Each prospect is represented by a button: by clicking on it, participants sample an outcome from the underlying distribution, with replacement. They can sample in whatever order and as many times as they like. When confident enough to evaluate the prospects, they select the one from which the actual payoff will be drawn. By means of this paradigm, we can investigate the role of experience on subjects' decisions, and, more interestingly, introduce an agency problem, as the clicking task requires agents' effort in reducing the degree of uncertainty when making decisions. Hence, on the one hand, we can observe whether including this task in a principal's decision problem affects their willingness to pay in order not to delegate. On the other hand, we can observe how much the principal and the agent are actually interested in reducing uncertainty. Therefore, the treatments are motivated by our interest in understanding not only how agent's behavior is affected by the risk exposure of a (passive) principal, but also how principal's expectations and behavior affect delegation.

2.2.2 Experimental Design and Session Structure

The experiment is based on a 2×2 factorial design. On the one hand, we manipulate the way in which the principal receives information on prospects to build his own portfolio, since he can either receive a full description of prospects or discover these, by experiencing the lotteries. On the other hand, we experimentally manipulate the way in which the agent receives information about prospects when building their principal's portfolios.

As shown in Table 4.1, we combine these two factors to obtain four experimental treatments:

EE - Both the principal and the agent decide under experience;

DD - Both the principal and the agent decide under description;

DE - The principal decides under description, while the agent decides under experience;

ED - The principal decides under experience, while the agent decides under description.

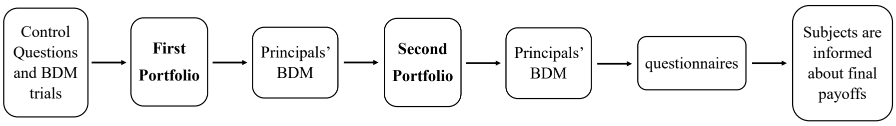
TABLE 2.1: Treatments

		Agent	
		Description	*Experience*
Principal	*Description*	DD	DE
	Experience	ED	EE

Specifically, the role and mode of information collection are experimentally manipulated in a within-subjects fashion. Indeed, Figure 2.3 shows that each session consists of two distinct yet identical parts for the structure, but not for the specific prospects: if a participant decides under Description in the first part (either treatment DD or DE), then he decides under Experience (either treatment EE or ED, respectively) in the second part, and vice versa. Because of this within-subject manipulation, two different sets of prospects are implemented (A and B): in the first part of the experiment, subjects build a portfolio from one set; in the second part, they build a second portfolio from the other set (see Table 2.10 and Table 2.11).

Every session includes two questionnaires measuring subjects' locus of control and risk attitudes, and a questionnaire for demographics.[13] Questionnaires are administered at the end of the session, before subjects are made aware of their final payoff. The first questionnaire consists of eight questions from the Levenson's IPC (Internal, Powerful Others, and Chance) scale[14] (Levenson, 1972), while the second is composed of seven questions from the 30-item version of the DOSPERT (Domain-Specific Risk-Taking) scale (Blais and Weber, 2006).[15]

FIGURE 2.3: Overview of the experimental structure



Each session ends with feedback about participants' final payoffs. A principal's payoff is determined as the sum of the payoffs he gets in the two parts of the experiment; it is computed at the end of the sessioh. Payoffs depend first on the BDM procedure (Becker, DeGroot, and Marschak, 1964): if the principal's bid (i.e. the willingness to pay in order not to delegate) is higher

[13]An English version of the questionnaires is in Appendix A.

[14]This scale of internal control products a measure of individual belief in chance as separate from belief in powerful others: it allows us to determine to which extent subjects believe that events in their life directly depend on their own actions. Therefore, such a scale is relevant for the present experimental study, since we observe subjects' willingness to 'control' decisions on risky events affecting their earnings. We built a sample questionnaire consisting of five questions on chance and three questions on internal control. Following Sapp and Harrod (1993), we relied on the Lumpkin (1988) validated brief version of the Levenson (1972) scale, though excluding the powerful others dimension.

[15]This psychometric scale allows us to introduce an additional control for participants' risk taking in specific domains. Given the focus of our research, the sample questionnaire consists of four financial and three social questions on work situations. Questions regarding ethical, health/safety, or recreational issues have been neglected.

than the randomly generated number, then the principal keeps his own portfolio. According to the BDM procedure, the principals pay a fee equal to the randomly drawn number. If the bid is lower, then their payoff for that part is determined by the portfolio built by their agent. At the end of the session, all selected prospects are played out and the principal is paid according to the outcomes of the gambles. The agent's payoff is fixed: it does not depend on the decisions made on behalf of their principal and it is line with the usual average payment that subjects receive in our laboratory.

2.2.3 Participants and Procedure

The experiment was run at CEEL (Cognitive and Experimental Economics Laboratory) of the University of Trento (Trento, Italy), and participants were recruited among undergraduate students or recent graduates (of the same university), who previously subscribed to CEEL's database. The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). Overall, we conducted eight sessions.

In total, 156 participants took part in the experiment: 78 subjects (43 males, and 35 females) in both treatment EE and DD, while the remaining 78 subjects (40 males and 38 females) in ED and DE. The average age was 22.10 (s.d. = 2.565). Most of the participants (72%) were students of Economics, 4% of Law, 5% of Engineering, 5% of Humanities, 9% of Social Sciences, 2.5% of Mathematics and Hard Sciences, and 2.5% had recently graduated. None of the participants was informed about the purpose of the experiment and every subject was allowed to participate only once.

Upon their arrival in the laboratory, participants were randomly assigned to a computer, and asked to sit in cubicles. They were provided with the instructions of the entire experiment and were informed that the experiment was composed of two independent parts. Subjects were given time to read the instructions individually. Then, instructions were read aloud by one of the experimenters. Before the experiment started, participants answered a few questions about the experimental rules, and were given the opportunity to play three trial BDMs, which did not affect final payoffs. This was intended to check participants' comprehension, both of the experimental instructions and of the bidding mechanism.

In both the software and the instructions we employed non-loaded terminology, such as “Participant 1” (for principals), “Participant 2” (for agents), and “prospect”. This is intended to rule out any context-related effect, and make our results more generalizable and valid in a variety of frameworks involving delegated risk-taking.

Each session lasted about 50 minutes. As for the payoffs, Participants 1 received 3 Euros as a show-up fee, plus a sum that varied according to their decisions (or, as appropriate, according to the decisions of their agent). In the end, this sum was converted into Euro and rounded up or down to the nearest ten euro-cent (1000 ECU = 2 Euros). On average, these participants earned 14 Euros (with a maximum of 16.90 Euros, and a minimum of 9.70 Euros, show-up fee included). Participants 2 earned a fixed amount of 13 Euros (show-up fee included).

2.3 Results

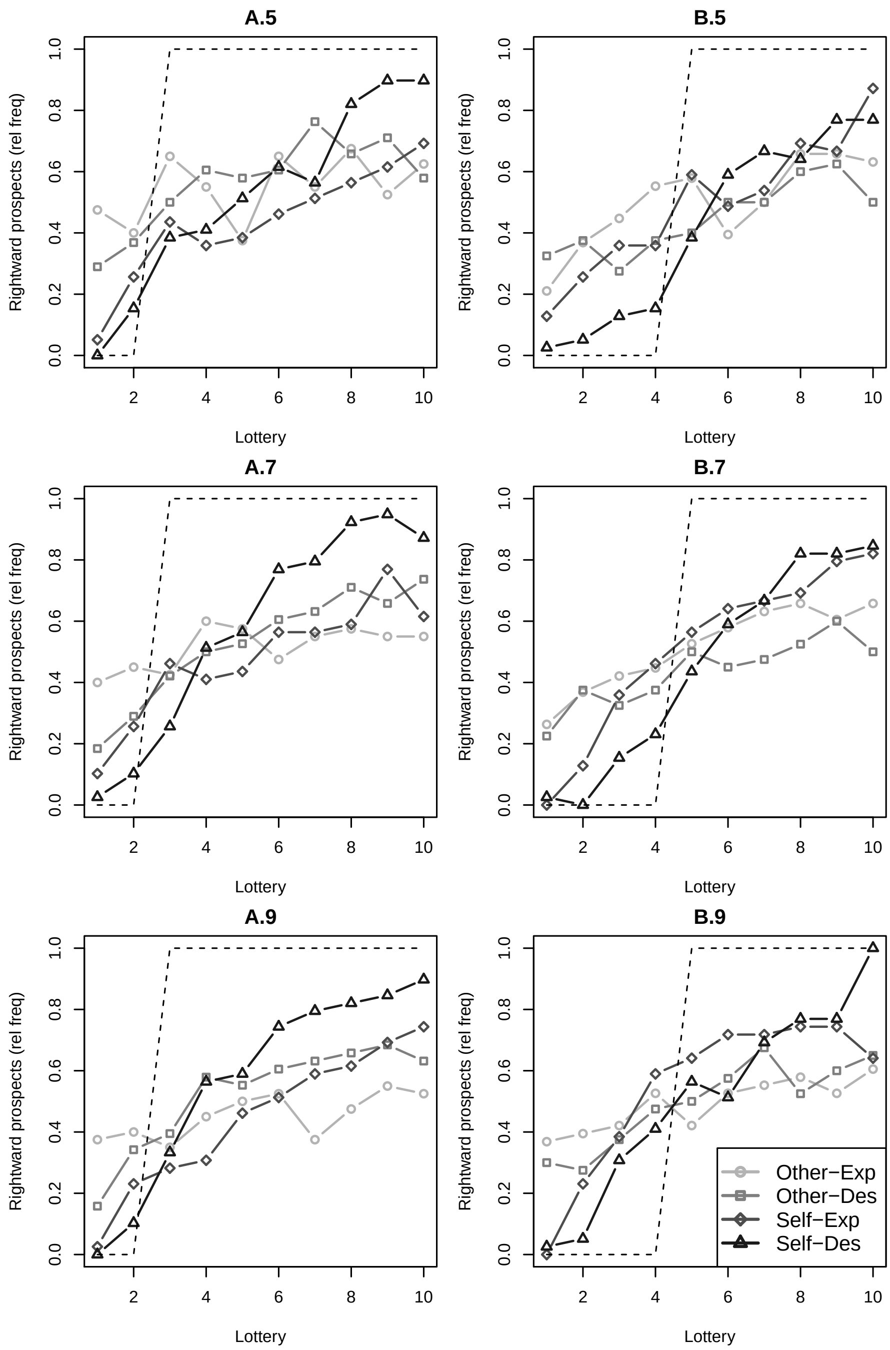
Experimental data are presented in two steps. First (in Sections 3.1-3.3), we present a statistical descriptive analysis of participants' portfolios: we compare subjects' choices across treatments and roles. Also, we consider the level of effort expended by both principals and agents in the process of information gathering, as well as principals' desire for control. Second (in Section 4.5.2), we present a regression analysis.

2.3.1 Analysis of Risk and Efficiency in Decisions

Figure 2.4 reports participants' portfolio decisions, keeping distinct both the role and the information process. Each panel reports the frequency of choice for the Rightward prospect in each of the 10-prospect Multiple Price Lists (see Tables 2.10 and 2.11 in the Appendix for more details). The dashed line shows the choice pattern of a risk-neutral decision maker.

The distribution of agents' decisions in each MPL is systematically flatter and more stable. Compared to principals, agents choose fewer “ambitious” Rightward prospects beyond the risk-neutrality switching point and are more likely to choose dominated prospects before this. In this respect, an analysis of individual frequencies of Rightward prospects selected in Prospect # 1 (where the Leftward prospect always stochastically dominates the Rightward

FIGURE 2.4: Prospect Choice



prospect) shows interesting differences. Agents choose more frequently the dominated Rightward prospect than principals, both in Experience (WRT, $p-value < 0.001$) and in Description (WRT, $p-value < 0.001$).[16] Overall, compared to description, experience seems to lead to a more frequent selection of dominated prospects (Table 2.2). This is especially true for agents (WST,

[16] All tests are two-sided, if not specified. WRT stands for Wilcoxon Rank Sum Test, while WST stands for Wilcoxon Signed Rank Test.

$p - value = 0.005$), while only a marginally significant difference emerges for principals (WST, $p - value = 0.078$).

TABLE 2.2: Relative frequency of dominated choices - Prospect # 1

	Other	Self
Description	0.248	0.017
Experience	0.350	0.051

Result 1a - *Agents tend to make systematically more dominated choices than principals. Agents perform significantly worse under Experience than under Description.*

Building on this evidence, we perform a more detailed analysis of risky choices, measuring both the expected return (μ_{PF}) and the standard deviation (σ_{PF})[17] of every 10-prospect portfolio in a MPL.[18]

A summary of average expected portfolio returns and standard deviations across 10-prospect MPL portfolios is reported in Table 2.3.[19] Principals tend to build more ambitious portfolios (higher return μ and degree of risk σ) under Description than under Experience: this is more evident in Set A (WST on μ: $p - value < 0.001$; WST on σ: $p - value < 0.001$) than in Set B. In contrast, no clear tendency emerges for agents' portfolios: in Set A, they are slightly more ambitious under Description than under Experience (WST on μ: $p - value = 0.064$; WST on σ: $p - value = 0.202$), while the opposite is observed in Set B, yet not systematically.

The identification of these measures (μ and σ) also allows us to draw a comparison among portfolios according to the mean-variance efficiency criterion.[20]

[17] The portfolio's expected return $\mu_{P_i} = \sum_{i=1}^{30} w_{P_i}\mu_{P_i}$ is defined as the weighted average of expected returns of every prospect P_i selected from the three MPLs. The portfolio's standard deviation $\sigma_{P_i} = \sqrt{\sum_{i=1}^{30} w_{P_i}^2 \sigma_{P_i}^2}$ is determined as the square root of the weighted average of variances of every prospect P_i (since the covariance across prospects is assumed to be null, the component $\sum_{P_i}\sum_{P_j \neq P_i} w_{P_i} w_{P_j} \sigma_{P_i}\sigma_{P_j}\rho_{P_i P_j}$ is omitted).

[18] We focus on portfolio choices at the MPL level instead of focusing on 30-prospect portfolios A and B, to account for the nature of the experimental task: participants were choosing in a MPL without knowing the nature of prospects in the following MPL. Because of this they could not develop a global portfolio strategy.

[19] The same analysis for each MPL can be found in Appendix B (Table 2.7).

[20] Every portfolio $\hat{PF}$ is non-dominated by another portfolio PF either when $\sigma_{\hat{PF}} < \sigma_{PF}$ or when $\sigma_{\hat{PF}} \geq \sigma_{PF}$ and $\mu_{\hat{PF}} \geq \mu_{PF}$.

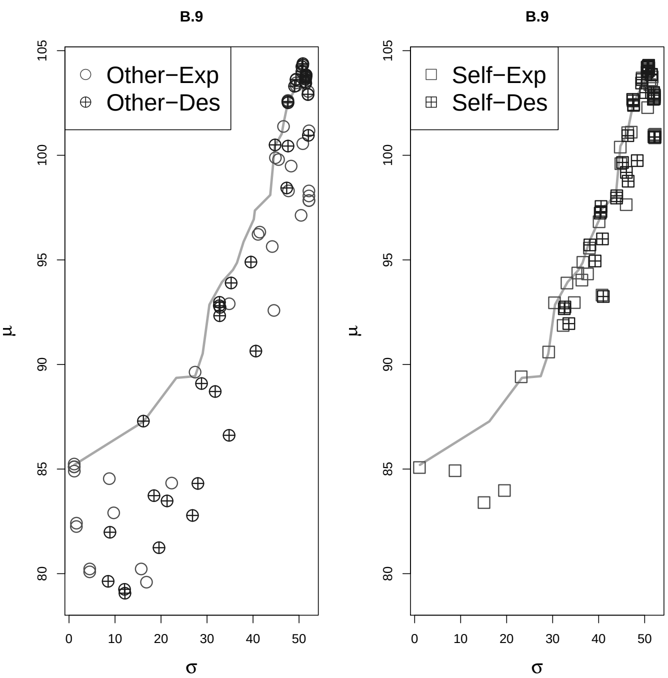
TABLE 2.3: Portfolios' Average Expected Returns and Standard Deviations

	Other-Des	Other-Exp	Self-Des	Self-Exp
		A		
μ	78.973 (9.454)	76.162 (8.467)	83.385 (5.788)	78.652 (7.028)
σ	24.015 (14.27)	22.049 (12.535)	28.87 (12.102)	24.567 (12.769)
		B		
μ	93.824 (8.995)	94.483 (8.019)	99.514 (5.213)	97.906 (6.000)
σ	24.892 (13.767)	24.535 (14.759)	30.188 (13.019)	28.413 (12.976)

Notes: For every set, role and mode of information acquisition, we compute the portfolio's average expected returns (μ) and standard deviations (σ). Corresponding standard deviations are in parentheses.

In our analysis we consider separately each of the 10-prospect sub-portfolios, as exemplified in Figure 2.5.[21]

FIGURE 2.5: Portfolios in the mean/variance space - MPL B.9



Beside confirming that principals' portfolios are characterized by higher returns and risk, especially under Description, the analysis shows that these are also closer to the efficiency frontier (gray line) of observed non-dominated choices.

Table 2.4 reports the overall frequency of dominated and non-dominated

[21] A complete graphical representation of all MPLs can be found in Appendix B (Figure 2.24).

choices in all 10-prospect sub-portfolios.[22] It is evident that experience leads to a lower degree of efficiency:[23] irrespective of the role, the proportion of dominated portfolios is significantly higher under Experience than under Description (WST on principals: $p - value < 0.001$; WST on agents: $p - value < 0.001$).[24] Nevertheless, it emerges, at the same time, that the proportion of dominated portfolios is systematically higher for agents than for principals (WRT under Description: $p - value < 0.001$; WRT under Experience: $p - value < 0.001$). Principals are generally able to build an efficient portfolio under Description, while seem to face some difficulties in doing so under Experience. In contrast, the majority of agents' portfolios is not efficient even under Description, where the prospect evaluation process is assumed to be simpler. The degree of inefficiency is dramatically high under Experience.

Table 2.4: Overall Portfolio Efficiency

%	Dominated Portfolios	Non-Dominated Portfolios
Other-Des	**55.1**	44.9
Other-Exp	**73.5**	26.5
Self-Des	31.2	**68.8**
Self-Exp	**52.6**	47.4

Result 1b - *Principals choosing under Description build more ambitious and efficient portfolios. The majority of portfolios built by agents are not efficient.*

2.3.2 The Portfolio Building Process: Effort Analysis

The clicking task can be intended to mimic the effort that a decision-maker exerts to perform an informed prospect selection. Figure 2.6 provides us with a visual analysis of both principals and agents' average effort in reducing the degree of uncertainty across MPLs.

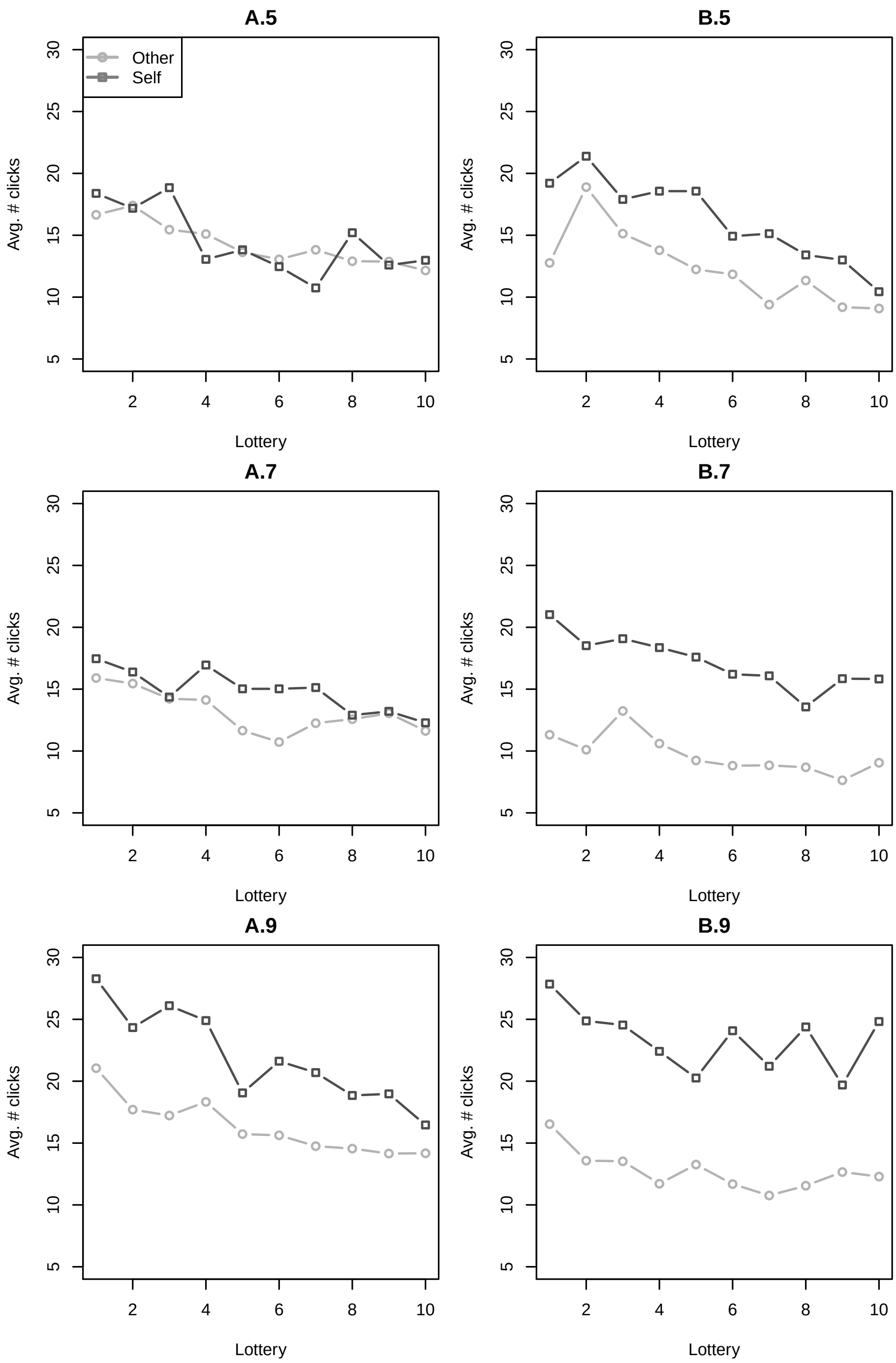
AS expected, if no monetary incentive is involved, agents invest significantly less in exploring lotteries than principals do (WRT on list B.5: $p - value = 0.035$; WRT on list B.7: $p - value < 0.001$; WRT on list B.9: $p - value < 0.001$; WRT on list A.9: $p - value = 0.037$). Specifically, the

[22] Table 2.8 in the Appendix reports the frequency of dominated/non-dominated choices in each of the 10-prospect sub-portfolios.

[23] A more detailed analysis can be found in Appendix B (Table 2.8)

[24] All tests are performed on averages at the individual level to preserve statistical independence.

FIGURE 2.6: Clicking Process - Effort Task



higher the degree of heterogeneity in the probability distribution, the more the difference becomes evident. Indeed, the average level of effort exerted by agents is quite stable and similar across MPLs. In contrast, principals' effort gradually increases when moving from the first to the third MPL of the same set (see Figure 2.25 in Appendix B), i.e. when inferring the right underlying probability distribution becomes more complex because of the presence of rare events ($p = .9$). Furthermore, the difference in effort is stronger in Set B than in Set A, with the former displaying larger differences in the standard deviation of the Rightward and Leftward prospects than the latter. When pooling data irrespective of the set and the probability associated to the lowest outcome, Agents click on average 6.6 times for each choice, while Principals click 9 times, on average. A test on individual-level data shows that the overall difference in clicking between the two types is highly significant (WRT, p-value< 0.001)

Result 2 - *Principals exert higher effort than agents do. The difference in effort is significantly larger for the set with higher variance and for prospects characterized by rare events.*

2.3.3 Investment in Delegation Avoidance

Now we consider principals' willingness to pay (WTP) to avoid delegation, which, as shown by Table 2.5, is generally very high.

In general, principals prefer to make decisions on their own. One possible explanation is that they correctly predict the degree of risk and inefficiency characterizing agents' portfolios. Accordingly, they are willing to pay a substantial fee, which decreases their potential earnings, in the end. This is true both when the information process is asymmetric (Treatment DE and ED) and when it is symmetric (Treatment EE and DD), even if Table 2.5 shows the willingness to avoid delegation is lower in this case.

Table 2.5: WTP to avoid delegation

Treatment	Mean	St. Dev.
EE	444.10	260.39
DD	503.33	231.49
DE	615.89	225.34
ED	517.17	231.25

Principals' WTP reveals a strong distaste for the Experience condition relative to the Description condition. This is true not only in case of principals' description, but also in case of agents' description. In fact, at an aggregate level, principals' willingness to pay in DE is higher than in EE (WRT: p-value = 0.0014) and in DD (WRT: p-value = 0.0096). Similarly, their willingness to pay is higher in ED than in EE, yet with no statistical significance. This allows us to rule out the notion that principals prefer to keep their portfolios only because they have exerted a positive and substantial effort in gathering information to reduce uncertainty. Principals do not feel too confident in making decisions under Experience, and thus they are less willing to pay in order to retain their portfolios. Since they cannot rely on objective distributions but on evaluations, they are more ready to incur the risk embedded in agents' portfolios. For the same reason, their desire to take over control increases when agents face the Experience condition. As a result of the combination of these two effects, principals' willingness to pay is systematically higher in treatment DE than in ED (WRT: p-value = 0.0089), and no significant difference emerges when comparing DD and ED.

Result 3 - *Principals reveal the strongest desire to take over control when they decide under Description and agents decide under Experience. The desire to take over control is at its minimum level when both decide under Experience.*

With respect to Result 3, it is interesting to note that the lowest proportion of inefficient portfolios is actually identified among principals' portfolios built under Description, while the highest proportion of inefficient portfolios among those built by agents is identified under Experience (see Table 2.4). This means that the highest principals' willingness to pay is found in the treatment where the efficiency difference between principals and agents' portfolios is maximized. As a matter of fact, since they are not explicitly incentivized, agents do not feel like collecting information, when no full description is provided: the quality of their decisions is quite low, even if they know that, by default, their portfolios will determine principals' payoffs. Nevertheless, it is also worth noting that principals seem to overinvest in delegation avoidance, overall. A comparative analysis of portfolios' expected returns shows that in every treatment but ED principals might get higher expected earnings from their own portfolios.[25] However, irrespective of the treatment and the lottery group, principals would systematically earn more by delegating rather than paying the premium they choose (WRT: p-value < 0.01).

[25] For the sake of completeness, we introduce a distinction among lottery sets, and we always observe a significant difference in Set B (WRT on DD: p-value< 0.05; WRT on EE: p-value< 0.1; WRT on DE: p-value< 0.05). As for Set A, principals' portfolios ensure higher expected returns only when the principal decides under Description (WRT on DD: p-value< 0.1; WRT on DE: p-value< 0.001)

2.3.4 Regression Analysis

Table 2.6 reports a regression analysis concerning the determinants of participants' behavior. Four different dependent variables are considered and, accordingly, four different estimates are reported: *Model (1)* takes as its dependent variable the expected returns of the 10-prospect subportfolios (MPLs); *Model (2)* focuses on determinants of non-dominated sub-portfolios; *Model (3)* focuses on clicking effort; *Model (4)* analyzes determinants of principals' BDM bids. In *Model (1)*, *Model (3)*, and *Model (4)* estimates are obtained via a Linear Mixed Model (LMM). In *Model (2)* a Generalized Linear Mixed Model (GLMM) Logit is adopted, given the dichotomous nature of the dependent variable.

Among explanatory variables, $Portfolio\ St.\ Dev.$ controls for the risk of each sub-portfolio. The first treatment dummy variable is $Self$: it is equal to 1 when the portfolio is built by a principal, otherwise it is 0. The other treatment dummy is *Experience*: it is equal to 1 if the portfolio is built under Experience, and equal to 0 if built under Description. The effect of the interaction between these two variables is estimated by introducing the term $Self\&Experience$.

The dummy *Set B* takes value 1 when the list from which lotteries are selected belongs to Set B, instead of Set A. Dummy variables $Prob$ 0.7 and $Prob$ 0.9 take into account the probability distribution of the ten prospects included in a portfolio (either 0.70/0.30 or 0.90/0.10). In Model (3) the variable *Rightward Prospect* is equal to 1 for clicks on the riskier button, and to 0 for clicks on the safer button; the dummy $EV+$ takes value 1 when the expected return of the clicked prospect is higher than that of the alternative. In Model (4), *Agent Des* and *Principal Des* capture choices in which the Agent and the Principal are in the Description condition, respectively. The dummy variable $Non\ Dominated$ takes value 1 in case of an efficient portfolio and 0 in case of a dominated one.

TABLE 2.6: Regression analysis

	Portfolio Return (1)	ND Portfolio (2)	Searching Effort (3)	Control Premium (4)
Portfolio Return				0.503 (3.899)
Portfolio St. Dev.	0.535 (0.013)***			−3.299 (4.342)
Non Dominated				−8.568 (18.576)
Set B	15.999 (0.206)***	−0.830 (0.153)***	−0.630 (0.675)	−16.827 (76.575)
Self	2.357 (0.487)***	1.257 (0.279)***	2.445 (0.674)***	
Experience	−0.447 (0.289)	−1.077 (0.222)***		
Prob 0.7	−1.864 (0.258)***	0.290 (0.184)	−0.302 (0.170)$^{\circ}$	
Prob 0.9	−10.332 (0.359)***	0.000 (0.184)	2.127 (0.170)***	
Self&Experience	−1.098 (0.409)**	−0.054 (0.305)		
Rightward Prospect			2.221 (0.139)***	
EV+			−1.749 (0.156)***	
Principal Des				78.870 (35.065)*
Agent Des				−26.040 (30.993)
Constant	69.367 (0.432)***	0.062 (0.234)	6.391 (0.604)***	508.551 (306.106)$^{\circ}$
Observations	936	936	9360	156
Num. groups: ID	156	156	156	78
Fitting model	*LMM*	*GLMM(Logit)*	*LMM*	*LMM*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Model (1) confirms that principals extract significantly higher returns, yet at a higher risk, from their portfolios than agents do: in fact, both the dummy variable *Self*, and the variable *Portfolio St. Dev.* are highly significant. The interaction term *Self&Experience* confirms that the negative impact of experience on returns is stronger for principals than for agents. *Model (2)* shows that principals are more likely to choose non-dominated prospects, and that, in general, the Experience condition and the *Set B* variable both lead towards a greater degree of inefficiency. *Model (3)* confirms that principals explore more than agents. Riskier prospects seem to induce more search, which, however, decreases for prospects with larger EV relative to the alternative. Finally, *Model (4)* shows that principals are ready to pay a higher payoff premium when they are in the Description condition than when they are in the Experience condition, but they do not discriminate between the conditions faced by their agents.

In the Appendix (see Table 2.9), we report the outcomes of a regression analysis which replicates the analysis of Table 2.6, but adds several further control variables for idiosyncratic features of the participants. Specifically, we control for gender, enrollment in the economics program (a dummy variable *Economics* that takes value 1 when the participant is a student of economics), locus of control (*Levenson* variable), and risk attitude (*Dospert* variable). In terms of main explanatory variables, the analysis corroborates the results reported above. In terms of the impact of the control variables, it is interesting to observe that those with higher scores in the Levenson test are more likely to choose a dominated portfolio and that females pay higher fees to retain control over their portfolio composition. This gender effect might be due to the fact that females seem to value the act of choosing more than males, especially in service purchases (Mattila, 2010). An alternative or additional explanation can be related to the emphasis females usually put on voice in interpersonal decision process (McColl-Kennedy, Daus, and Sparks, 2003).

2.4 Conclusions

We show that the mode of information acquisition produces a systematic effect not only on choice performance but also on the emergence of delegation itself. Overall, participants exhibit a worse performance under Experience than under Description: they face a more complex decision situation in which their direct willingness to gather and collate information affects the underlying degree of uncertainty.

A novel result of our research concerns agents' and principals' decision quality. The great majority of subjects deciding on behalf of someone else make dominated decisions: especially when information sampling is required, agents select prospects ensuring an inefficient combination of risk and expected returns. Since their final payoff is not linked to their decisions, agents seem unwilling to exert effort in acquiring information on prospects' probability distributions. This might be one of the causes of agents' poor performance as compared to that of principals, under Experience.[26] Nevertheless, a performance discrepancy also emerges under Description, where sampling errors can play no role at all: the quality of agents' decision-making improves when they are provided with full information, though it still falls short of that exhibited by principals faced with full information. In fact, irrespective of the process of information acquisition, principals make more ambitious and efficient decisions, even if, to a certain extent, the experience framework affects negatively also their performance. Our study mimics ubiquitous real-world situations in which the decisions of the agent have consequences for another individual but not for the agent themselves: this is not only the case of financial decisions,

[26] Only principals sample more than what observed in other studies involving decision from experience for binary lotteries (see Hau, Pleskac, and Hertwig, 2010).

but also of medical decisions, for instance. Understanding how to optimally design incentives in delegated risky choices goes beyond the scope of our study, but may represent an interesting venue for future research.

A further result concerns the effect of experience on principals' confidence in delegation: principals tend to show a preference for the decision setting that involves prospects' full description. The control premium is highly positive: it is larger when agents learn from experience, and it is orthogonal to the main characteristics of principals' portfolios (expected returns, standard deviations, and dominance). This result is in line with a questionnaire study by Botti and Iyengar (2004): they report that, when facing a decision problem, people prefer making their own decisions, instead of having their decisions either dictated by someone else or determined by a random device.

This confirms the relevance of agents' process of information acquisition: agents learning in a more uncertain environment are less likely to be trusted with delegation. Therefore, besides the inefficiency issue, which might be addressed by means of monetary incentives, agents need to understand how to attract customers. They may decrease principals' unwillingness to delegate, leveraging on their own reliability: they can make decisions based on solid knowledge, not on vague evaluations. In this framework, experience as a learning mode can help improve agents' reliability, when combined with a valid training: in fact, customers or patients' delegation decision may also depend on information such as the place where the agent has graduated or previously worked. Future research might focus on the effect of combining the two sources of knowledge.

Missed delegation relationships are detrimental both to agents and to principals, who overestimate the difference between their own performance and that of agents. In our study, principals' portfolios tend to ensure higher expected returns, but, at the same time, principals are willing to pay an excessive control premium to enact their decisions and avoid delegation: despite agents' inefficiency, they could earn more by delegating than by paying to retain control over their outcomes.

2.5 Appendix A - Experimental Instructions

This is a translated version (originally in Italian) of the instructions used for the experimental sessions. Instructions change according to the treatment. This will be indicated in the text. As for the within-subject manipulation: Treatment T1 (either first or second part) has been paired with Treatment T2 (either first or second part); Treatment T3 (either first or second part) has been paired with Treatment T4 (either first or second part).

GENERAL INSTRUCTIONS

Welcome,

Thank you for coming. You are going to take part in an experiment on economic decisions. For arriving on time, you will receive 3 Euros at the end of the experiment.
Following you will be given instructions for the experiment. Please, read them carefully. May you have any doubt, raise your hand and a member of the experimental staff will come to answer your question.
During the experiment, you are not allowed to talk to other participants. If you disturb your colleagues or use the computer for activities not strictly related to the experiment, you will be excluded from the experiment and any reward. You can trust that what happens during the experiment is in line with the following instructions.

The experiment consists of two independent parts.
You will be randomly assigned a role (either *Participant 1* or *Participant 2*), that will remain unchanged during the entire experimental session (including both the first and the second part). If you are a Participant 1, you will be asked to make decisions for you, i.e. decisions that will affect only your own payoff. On the contrary, if you a Participant 2, you will be asked to make decisions for another participant, i.e. decisions that will affect only the payoff of this participant and not your own payoff.
Every Participant 1 will be randomly assigned to one of the Participants 2, so that to each Participant 2 corresponds one (and only one) Participant 1.

Both the first and the second part of the experiment consists of two sequential decision phases for those playing the role of Participant 1, while they consist of only one phase for those with the role of Participant 2.

In the end, you all will be asked to answer a brief questionnaire, and you will be informed of your final payoff, which is determined as the sum of the payoffs you get during the first and the second part of the experiment.

Following, you will find the experimental instructions. You will be given five minutes to read them. Instructions will be then read aloud by a staff member; you will be asked to answer few simple questions on instructions comprehension.

During the experiment, ECU (Experimental Currency Units) will be used to express your earnings. At the end of the experimental session, the ECU you will have earned are converted in Euros (and rounded to the nearest ten euro-cent) in order to determine your real payoff (1000 ECU = 2 Euros).

INSTRUCTIONS: FIRST PART

At the beginning of this part of the experiment, you will be informed of your role.

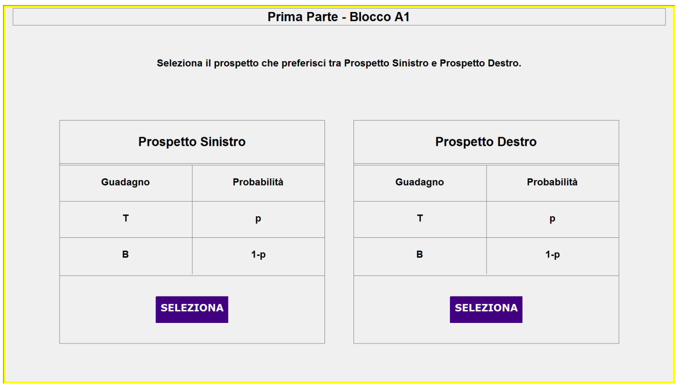
`Treatment DD`

Participant 1

- *First Decision Phase*: You will be asked to make 30 decisions that will affect your payoff. These decisions are divided into three subsets: therefore, each of them consists of 10 decisions. You will (sequentially) go through 30 couples of prospects, and you will have to choose the prospect you prefer (between Left Prospect and Right Prospect) for each of the 30 couples. In general, a prospect offers an outcome T with probability p and an outcome B with probability *1-p*. The value of T and B can vary for every prospect. Outcomes are in ECU.

  Following, you can find an example of a couple of prospects. For each couple, click on the SELECT button corresponding to the prospect you prefer.

  FIGURE 2.7: Participant 1 Decision Task - First Part - Example

  In the meanwhile, your Participant 2 will decide on the same prospects. Therefore, you both will be asked to make 30 decisions on the same list of prospects. However, Participant 1's decisions will affect only his own payoff, while Participant 2's decisions will affect the payoff of the corresponding Participant 1.

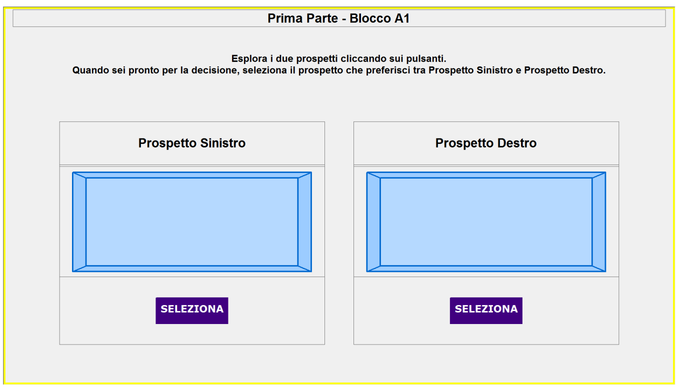
Treatment EE

Participant 1

- *First Decision Phase*: You will be asked to make 30 decisions that will affect your payoff. These decisions are divided into three subsets: therefore, each of them consists of 10 decisions. You will (sequentially) go through 30 couples of prospects, and you will have to choose the prospect you prefer (between Left Prospect and Right Prospect) for each of the 30 couples. In general, a prospect offers an outcome T with probability p and an outcome B with probability *1-p*. The value of T and B can vary for every prospect. All outcomes are in ECU.

  Consider that, in the beginning, you will not receive any information about the prospects. However, you will have the opportunity to collect the information you might need to make your decisions. For this reason, every prospect will be represented by a button: therefore, for each of the 30 decisions, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff you would have received by choosing the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose. At this point, click on the SELECT button corresponding to the prospect you prefer. Following, you can find an example of a couple of prospects.

  FIGURE 2.8: Participant 1 Decision Task - First Part - Example

  In the meanwhile, your Participant 2 will decide on the same prospects (presented as buttons). Therefore, you both will be asked to make 30 decisions on the same list of prospects. However, Participant 1's decisions will affect only his own payoff, while Participant 2's decisions will affect the payoff of the corresponding Participant 1.

`Treatment ED`

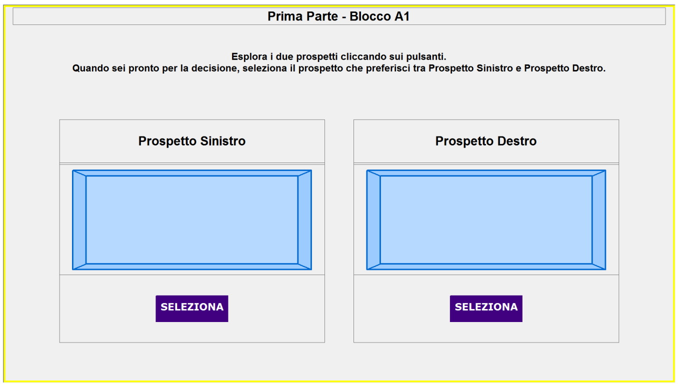
Participant 1

- *First Decision Phase*: You will be asked to make 30 decisions that will affect your payoff. These decisions are divided into three subsets: therefore, each of them consists of 10 decisions. You will (sequentially) go through 30 couples of prospects, and you will have to choose the prospect you prefer (between Left Prospect and Right Prospect) for each of the 30 couples. In general, a prospect offers an outcome T with probability p and an outcome B with probability *1-p*. The value of T and B can vary for every prospect. All outcomes are in ECU.

  Consider that, you will receive no prior information. You will have the opportunity to collect the information you might need to make your decisions. Every prospect is represented by a button: for each of the 30 decisions, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff you would have received by choosing the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough. At this point, click on the SELECT button corresponding to the prospect you prefer. Following, you can find an example of a couple of prospects.

  In the meanwhile, your Participant 2 will decide on the same prospects (yet presented with a full description of outcomes and probability distributions). Therefore, you both will be asked to make 30 decisions on the same list of prospects. However, Participant 1's decisions will affect only his own payoff, while Participant 2's decisions will affect the payoff of the corresponding Participant 1.

FIGURE 2.9: Participant 1 Decision Task - First Part - Example



`Treatment DE`

Participant 1

- *First Decision Phase*: You will be asked to make 30 decisions that will affect your payoff. These decisions are divided into three subsets: therefore, each of them consists of 10 decisions. You will (sequentially) go through 30 couples of prospects, and you will have to choose the prospect you prefer (between Left Prospect and Right Prospect) for each of the 30 couples. In general, a prospect offers an outcome T with probability p and an outcome B with probability *1-p*. The value of T and B can vary for every prospect. All outcomes are in ECU.

  Following, you can find an example of a couple of prospects. For each couple, click on the button SELECT corresponding to the prospect you prefer.

FIGURE 2.10: Participant 1 Decision Task - First Part - Example

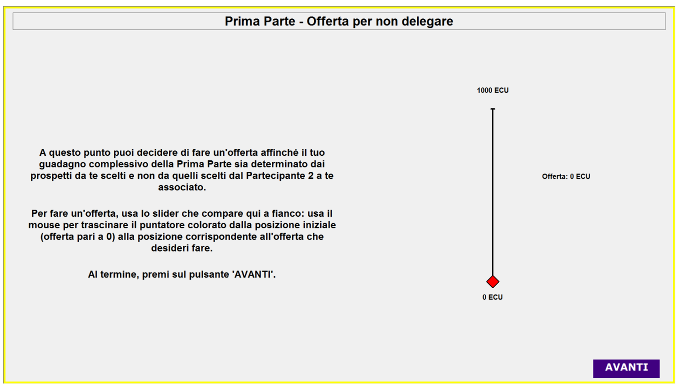
In the meanwhile, your Participant 2 will decide on the same prospects (yet presented as buttons). Therefore, you both will be asked to make 30 decisions on the same list of prospects. However, Participant 1's decisions will affect only his own payoff, while Participant 2's decisions will affect the payoff of the corresponding Participant 1.

`Common to all Treatments`

Participant 1

- *Second Decision Phase*: You will be asked to send a bid, so that your payoff (relative to the first part of the experiment) is determined by your choices, and not by those of the Participant 2 (you have been assigned to). The minimum bid you can send is equal to 0 ECU, while the maximum bid is equal to 1000 ECU. In order to state your bid, you can use a slider: drag the pointer in correspondence of the sum of ECU you are willing to pay.

FIGURE 2.11: Bid Mechanism - Example



In order to determine whether your bid is such that your payoff depends only on your decisions, the following procedure will be adopted. A number between 0 and 1000 is randomly generated by the computer so that every number can be drawn with the same probability.

- *If the randomly generated number is lower than or equal to your bid*, your bid is accepted. Your payoff (for this first part of the experiment) will be determined by playing the prospects you have previously chosen, minus the randomly generated number.
- *If the randomly generated number is higher than your bid*, your bid is rejected. Your payoff (for this first part of the experiment) will be determined by playing the prospects the Participant 2 you have been assigned to has previously chosen.

Consider that the higher is your bid, the higher is the probability that your bid is accepted, and, thus, that it's you determining your payoff. However, a too high bid might make you pay more than your willingness (if the randomly generated number is larger than your willingness to pay, but, at the same time, lower than your 'too high' bid). On the contrary, the lower is your bid, the higher is the probability that your bid is rejected, and, thus, that the Participant 2 determines your payoff. For all these reasons, the bid you are asked to send is the one representing your actual willingness to pay.

In any case, for a better comprehension of such a mechanism, the experiment will start with a simulation phase: you will have the opportunity of sending three independent (trial) bids that will not affect your final payoff.

At this point, the first part of the experiment ends. You will be informed of your payoff at the end of the experiment: if your bis is accepted, then your earnings will be determined as the sum of the payoffs of the prospects you have chosen; on the contrary, if your bid is rejected, your earnings will be determined as the sum of the payoffs of the prospects your Participant 2 has chosen.

Participant 2

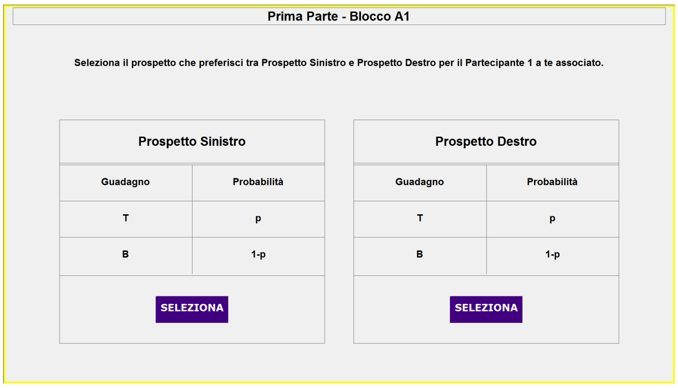
The first part of the experiment consists of a single decision phase, which is contemporary to Participant 1's first decision phase. You will be asked to evaluate 30 couples of prospects (the same of Participant 1), and, for each couple

(Left Prospect and Right Prospect), to choose a prospect for the Participant 1 you have been assigned to in the beginning.

`Treatment DD`

Following, you can find an example of a couple of prospects. For each couple, click on the SELECT button corresponding to the prospect you prefer for the Participant 1.

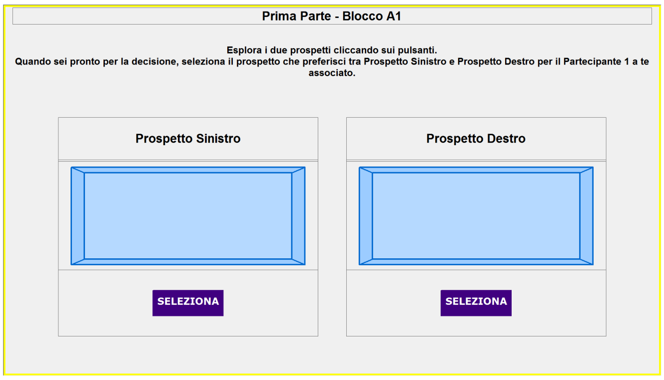
FIGURE 2.12: Participant 2 Decision Task - First Part - Example



`Treatment EE`

You will not receive any prior information. Every prospect will be represented by a button: for each decisions, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff your Participant 1 would receive with that prospect. You can click until you feel confident enough to choose. At this point, click on the SELECT button corresponding to the prospect you prefer for the Participant 1. Following, you can find an example of a couple of prospects.

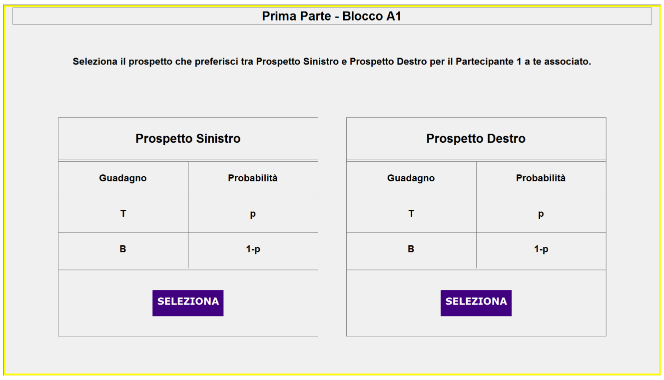
FIGURE 2.13: Participant 2 Decision Task - First Part - Example



`Treatment ED`

Decisions will be presented in a different way with respect to Participant 1's first decision phase: prospects are fully described in terms of probability and outcomes. You will not have to 'explore' the prospects in order to choose. Following, you can find an example of a couple of prospects.

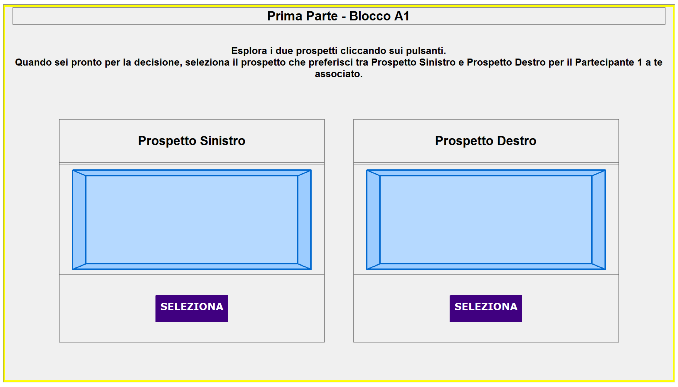
FIGURE 2.14: Participant 2 Decision Task - First Part - Example



`Treatment DE`

Consider that decisions will be presented in a different way with respect to Participant 1's first decision phase. Specifically, you will not receive any information about the prospects (probability p and outcomes T and B). Every prospect will be represented by a button: for each of the 30 decisions, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff your Participant 1 would have received if you had chosen the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose. At this point, click on the SELECT button corresponding to the prospect you prefer for the Participant 1. Following, you can find an example of a couple of prospects.

FIGURE 2.15: Participant 2 Decision Task - First Part - Example



Common to all Treatments

Once you have made all the 30 decisions, the first part of the experiment is concluded.

You will be informed of your payoff (relative to this first part) at the end of the experiment. However, your payoff is fixed, and it is not affected by the decisions you make for the Participant 1. Your choices can influence only his payoff.

INSTRUCTIONS: SECOND PART

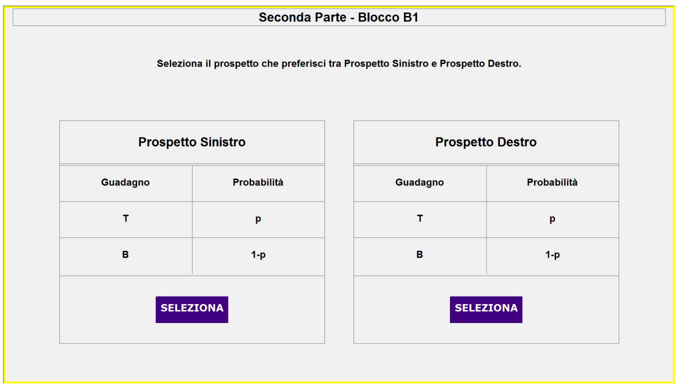
In this second part of the experiment you will have the same role as in the first part. Furthermore, also the decision phases will remain unchanged: there will be two phases (the 30 decisions for himself and the bid) for the Participant 1, and one phase (the 30 decisions for the corresponding Participant 1) for the Participant 2.

`Treatment DD`

The only difference concerns Participant 1's first decision phase and Participant 2's decision phase: in both cases, it will be asked to make 30 decisions, again in terms of sequential choices through 30 couples of prospects (different from those of the first part). However, such decision problems will be presented in a different way. More precisely, during the first part of experiment, it was asked to collect the information necessary to decide; on the contrary, during this second part, prospects are fully described (both in terms of probability p and outcomes T and B). Therefore, all the relevant information about the Left Prospect and the Right Prospect is available from the beginning.

- If you have the role of **Participant 1**, you can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer.

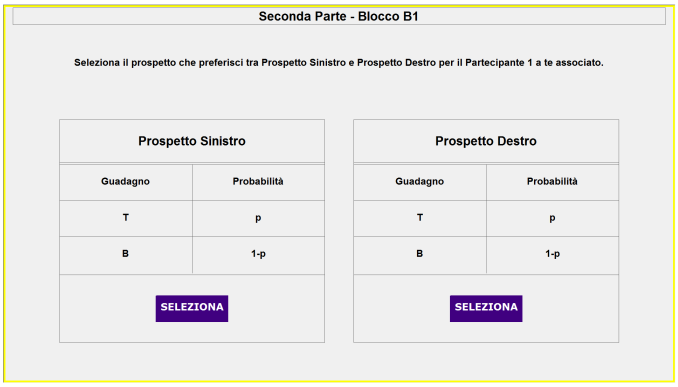
FIGURE 2.16: Participant 1 Decision Task - Second Part - Example



After your 30 decisions, your first decision phase is concluded. Then, you will move to the second phase, i.e. the one giving you the opportunity to send a bid to decide whose decisions will determine your payoff. Like in the first part, if your bid is accepted, your earnings will be defined as the sum of the payoffs of the prospects you have previously selected; on the contrary, if you bid is rejected, your earnings will be determined by the Participant 2 you have been associated to.

- If you have the role of **Participant 2**, you can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer or the Participant 1.

FIGURE 2.17: Participant 2 Decision Task - Second Part - Example



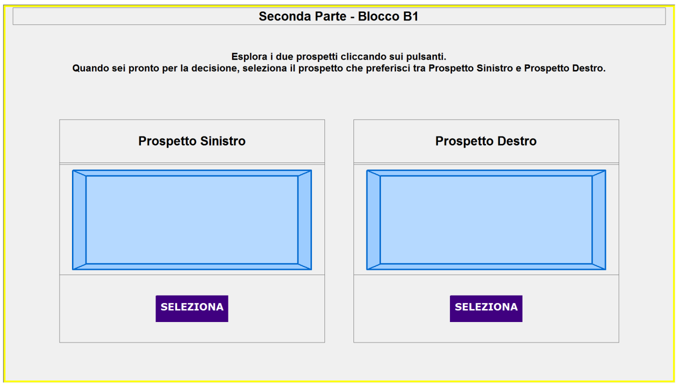
After your 30 decisions, your decision phase is concluded. Like in the first part, your payoff is fixed, and it does not depend on your decisions.

`Treatment EE`

The only difference concerns Participant 1's first decision phase and Participant 2's decision phase: in both cases, it will be asked to make 30 decisions, again in terms of sequential choices through 30 couples of prospects (different from those of the first part). However, such decision problems will be presented in a different way. More precisely, during the first part of the experiment, prospects were fully described (in terms of probability p an outcomes T and B): the relevant information is available from the beginning. On the contrary, in this second part of the experiment, you will have no prior information on the prospects; however, you will have the opportunity to collected the information necessary to decide. For this reason, each prospect will be represented by a button: for every decision, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff you would receive by choosing the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose.

- If you have the role of **Participant 1**, you can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer.

FIGURE 2.18: Participant 1 Decision Task - Second Part - Example



After your 30 decisions, your first decision phase is concluded. Then, you will move to the second phase, i.e. the one giving you the opportunity to send a bid to decide whose decisions will determine your payoff. Like in the first part, if your bid is accepted, your earnings will be defined as the sum of the payoffs of the prospects you have previously selected; on the contrary, if you bid is rejected, your earnings will be determined by the Participant 2 you have been associated to.

- If you have the role of **Participant 2**, you can find an example of a couple of prospects in the following figure. Like in the case of the Participant 1, prospects are represented by buttons. Every time you click on one of them, you will be immediately informed about the payoff your Participant 1 would have received if you had chosen the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose. For each couple, click on the SELECT button corresponding to the prospect you prefer or the Participant 1.

FIGURE 2.19: Participant 2 Decision Task - Second Part - Example

After your 30 decisions, your decision phase is concluded. Like in the first part, your payoff is fixed, and it does not depend on your decisions.

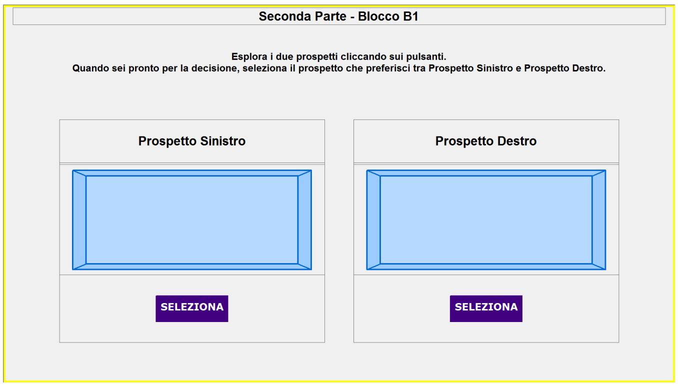
`Treatment ED`

The only difference concerns Participant 1's first decision phase and Participant 2's decision phase: in both cases, it will be asked to make 30 decisions, again in terms of sequential choices through 30 couples of prospects (different from those of the first part). However, such decision problems will be presented in a different way.

- If you have the role of **Participant 1**, during the first part of the experiment, your prospects were fully described (in terms of probability p an outcomes T and B): the relevant information was available from the beginning. On the contrary, in this second part of the experiment, you will have no prior information on the prospects; however, you will have the opportunity to collect the information necessary to decide. For this reason, each prospect will be represented by a button: for every decision, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff you would receive by choosing the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose.

  You can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer.

FIGURE 2.20: Participant 1 Decision Task - Second Part - Example



After your 30 decisions, your first decision phase is concluded. Then, you will move to the second phase, i.e. the one giving you the opportunity to send a bid to decide whose decisions will determine your payoff. Like in the first part, if your bid is accepted, your earnings will be defined as the sum of the payoffs of the prospects you have previously selected; on the contrary, if you bid is rejected, your earnings will be determined by the Participant 2 you have been associated to.

- If you have the role of **Participant 2**, during the first part of the experiment, no prior information on the prospects was available; in this second part, on the contrary, prospects will be fully described (in terms of probability p and outcomes T and B). Therefore, all the relevant information about the Left Prospect and the Right Prospect is available from the beginning.

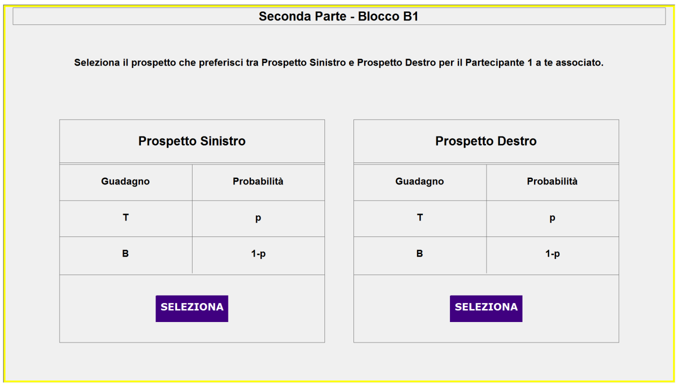
  You can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer or the Participant 1.

  After your 30 decisions, your decision phase is concluded. Like in the first part, your payoff is fixed, and it does not depend on your decisions.

`Treatment DE`

The only difference concerns Participant 1's first decision phase and Participant 2's decision phase: in both cases, it will be asked to make 30 decisions, again in terms of sequential choices through 30 couples of prospects (different from those of the first part). However, such decision problems will be presented in a different way.

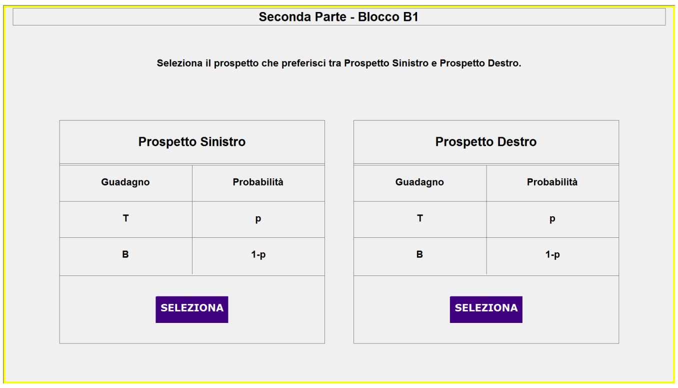
FIGURE 2.21: Participant 2 Decision Task - Second Part - Example



- If you have the role of **Participant 1**, during the first part of the experiment, no prior information on the prospects was available; in this second part, on the contrary, prospects will be fully described (in terms of probability p and outcomes T and B). Therefore, all the relevant information about the Left Prospect and the Right Prospect is available from the beginning.

  You can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer.

FIGURE 2.22: Participant 1 Decision Task - Second Part - Example



After your 30 decisions, your first decision phase is concluded. Then, you will move to the second phase, i.e. the one giving you the opportunity to send a bid to decide whose decisions will determine your payoff. Like in the first part, if your bid is accepted, your earnings will be defined as the sum of the payoffs of the prospects you have previously selected; on the contrary, if you bid is rejected, your earnings will be determined by the Participant 2 you have been associated to.

- If you have the role of **Participant 2**, during the first part of the experiment, the prospects were fully described (in terms of probability p an outcomes T and B): the relevant information was available from the beginning. On the contrary, in this second part of the experiment, you will have no prior information on the prospects; however, you will have the opportunity to collect the information necessary to decide. For this reason, each prospect will be represented by a button: for every decision, two buttons (one for the Left Prospect and one for the Right Prospect) will appear on your screen. Every time you click on one of them, you will be immediately informed about the payoff your Participant 1 would receive if you choose the corresponding prospect (according to the outcome and probability distributions associated to that specific prospect). You can continue to click until you feel confident enough to choose for the Participant 1.

  You can find an example of a couple of prospects in the following figure. For each couple, click on the SELECT button corresponding to the prospect you prefer or the Participant 1.

  FIGURE 2.23: Participant 2 Decision Task - Second Part - Example

  After your 30 decisions, your decision phase is concluded. Like in the first part, your payoff is fixed, and it does not depend on your decisions.

(This is an English translation of the questionnaires participants answered to at the end of the experiment.)

Levenson's Scale

We kindly ask you to answer the following questionnaire truthfully. We ask you to indicate how much you agree with each of the following statements by using a scale of 6 values that goes from "I strongly disagree" to "I strongly agree".

1. To a great extent my life is controlled by accidental happenings.
2. When I make plans, I am almost certain to make them work.
3. Often there is no chance of protecting my personal interests from bad luck happenings.
4. When I get what I want, it's usually because I'm lucky.
5. I have often found that what is going to happen will happen.
6. It's not always wise for me to plan too far ahead because many things turn out to be a matter of good or bad fortune.
7. When I get what I want, it's usually because I worked hard for it.
8. My life is determined by my own actions.

Dospert

We kindly ask you to answer the following questionnaire truthfully. For each of the following statements, please indicate the likelihood that you would engage in the described activity or behavior if you were to find yourself in that situation. Provide a rating from Extremely Unlikely to Extremely Likely, using the following scale: 1 = "Extremely unlikely", 2 = "Moderately unlikely", 3 ="Somewhat unlikely", 4 = "Not sure", 5 ="Somewhat likely", 6 = "Moderately likely", 7 ="Extremely likely".

1. Admitting that your tastes are different from those of a friend.
2. Betting a day's income at the horse races.

3. Investing 5% of your annual income in a very speculative stock.

4. Betting a day's income on the outcome of a sporting event.

5. Investing 10% of your annual income in a new business venture.

6. Choosing a career that you truly enjoy over a more secure one.

7. Speaking your mind about an unpopular issue in a meeting at work.

Demographics and Other Information

Please, fill in the following fields.

- Date of Birth:
- Gender:
- Field of Studies:
- Number of experiment in which you have taken part:

2.6 Appendix B - Additional Analysis

Table 2.7: MPLs' Expected Returns and Standard Deviations

mean (sd)	Agent-Des	Agent-Exp	Principal-Des	Principal-Exp
		A.5		
μ	78.605 (9.441)	76.965 (7.841)	82.737 (5.786)	77.983 (7.019)
σ	16.909 (6.639)	16.722 (5.205)	19.146 (4.17)	16.393 (5.17)
		A.7		
μ	79.542 (9.377)	76.239 (9.241)	84.038 (5.290)	78.639 (6.612)
σ	21.085 (9.402)	19.2 (8.67)	25.14 (5.037)	21.321 (6.211)
		A.9		
μ	78.773 (9.767)	75.282 (8.391)	83.379 (6.320)	79.335 (7.540)
σ	34.050 (18.037)	30.223 (16.470)	42.324 (10.354)	35.986 (14.805)
		B.5		
μ	93.323 (9.128)	94.763 (7.533)	98.303 (6.207)	97.533 (5.798)
σ	16.551 (6.239)	17.855 (6.086)	18.777 (4.953)	19.347 (5.063)
		B.7		
μ	93.291 (8.961)	995.184 (8.203)	99.875 (5.254)	98.835 (6.069)
σ	20.269 (8.304)	22.337 (9.033)	25.424 (5.849)	24.924 (7.484)
		B.9		
μ	94.858 (9.033)	93.502 (8.414)	100.365 (3.800)	97.351 (6.174)
σ	37.855 (14.312)	33.414 (20.369)	46.361 (5.767)	40.969 (13.180)

Notes: For every MPL, the average of both portfolios' expected returns (μ) and standard deviations (σ) is computed according to role (principal vs. agent) and information gathering condition (description vs. experience). Corresponding standard deviations are in parentheses.

FIGURE 2.24: Portfolios in mean/variance space

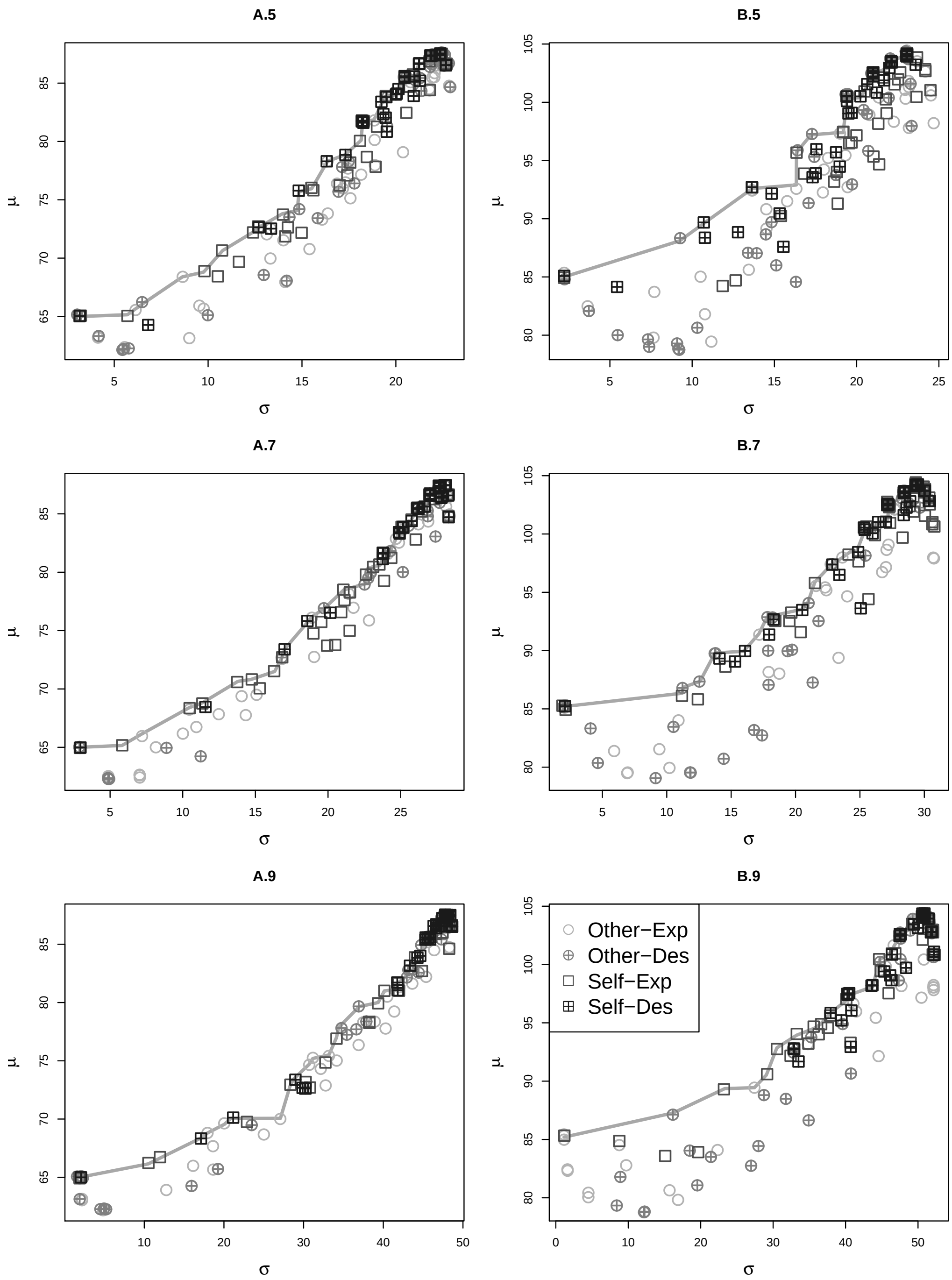


Table 2.8: Proportion of Dominance - MPL

%	Dominated Portfolios	Non-Dominated Portfolios
	A.5	
Agent-Des	**57.89**	42.11
Agent-Exp	**77.50**	22.50
Principal-Des	23.08	**76.92**
Principal-Exp	43.59	**56.41**
	A.7	
Agent-Des	34.21	**65.79**
Agent-Exp	**72.50**	27.50
Principal-Des	17.95	**82.05**
Principal-Exp	46.15	**53.85**
	A.9	
Agent-Des	44.74	**55.26**
Agent-Exp	**70.00**	30.00
Principal-Des	20.51	**79.49**
Principal-Exp	38.46	**61.54**
	B.5	
Agent-Des	**60.0**	40.0
Agent-Exp	**73.7**	26.3
Principal-Des	41.0	**59.0**
Principal-Exp	**61.5**	38.5
	B.7	
Agent-Des	**57.5**	42.5
Agent-Exp	**73.7**	26.3
Principal-Des	35.9	**64.1**
Principal-Exp	**59.0**	41.0
	B.9	
Agent-Des	**75.0**	25.0
Agent-Exp	**73.7**	26.3
Principal-Des	48.7	**51.3**
Principal-Exp	**66.7**	33.3

FIGURE 2.25: Average Clicking - Principal vs. Agent

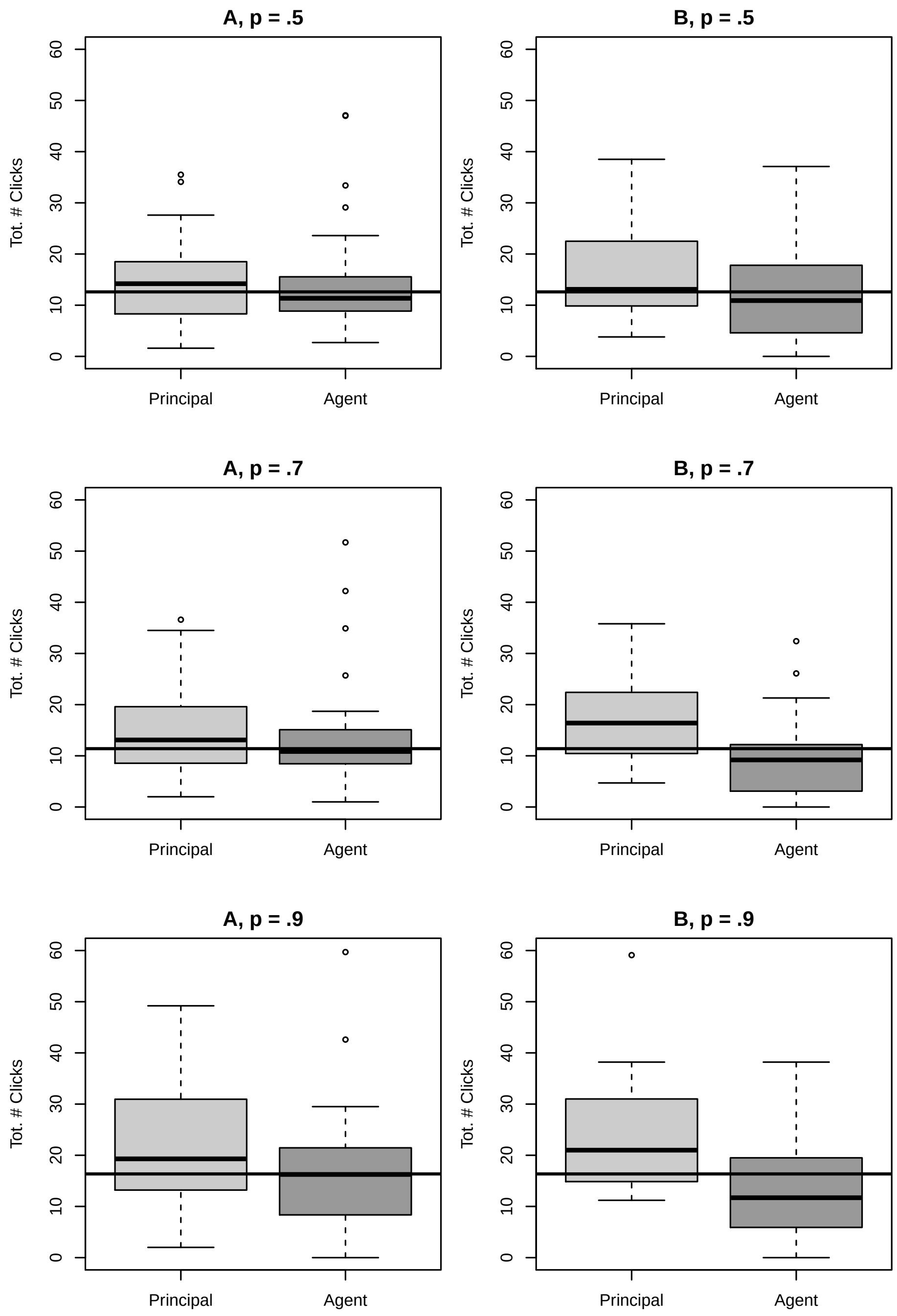


Table 2.9: Regression analysis with controls

	Portfolio Return	ND Portfolio	Searching Effort	Control Premium
Portfolio Return				0.202 (4.221)
Portfolio St. Dev.	0.526 (0.014)***			1.256 (4.941)
NonDominated				−4.726 (19.260)
Set B	16.065 (0.224)***	−0.752 (0.163)***	−1.336 (0.729)$^{\circ}$	−6.451 (19.294)
Self	1.989 (0.524)***	1.044 (0.299)***	2.409 (0.737)**	
Experience	−0.688 (0.313)*	−1.128 (0.236)***		
Prob 0.7	−1.847 (0.277)***	0.318 (0.196)	−0.267 (0.184)	
Prob 0.9	−10.207 (0.378)***	0.066 (0.196)	2.146 (0.184)***	
Self & Experience	−0.753 (0.442)$^{\circ}$	−0.078 (0.327)		
Rightward Prospect			2.263 (0.150)***	
EV+			−1.739 (0.166)***	
Principal Des				80.404 (36.198)*
Agent Des				−20.027 (31.369)
Levenson	−0.721 (0.524)	−0.701 (0.286)*	−0.033 (0.821)	92.859 (55.634)$^{\cdot}$
Dospert	−0.007 (0.262)	−0.046 (0.141)	−0.040 (0.412)	8.547 (27.350)
Female	−0.711 (0.489)	−0.126 (0.262)	−1.463 (0.773)$^{\circ}$	147.941 (55.612)**
Econ	−0.052 (0.496)	0.089 (0.265)	1.167 (0.777)	−42.437 (58.492)
Constant	72.876 (2.159)***	2.897 (1.169)*	7.092 (3.366)*	100.285 (395.725)
Observations	816	816	8160	136
Num. groups: ID	136	136	136	68
Fitting model	*LMM*	*GLMM(Logit)*	*LMM*	*LMM*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Notes: Because of a technical issue at the end of one experimental session, some participants' answers to the final questionnaire have not been properly recorded. Therefore, one of the sessions with treatments DD and EE has not been included in the following regression analysis.

2.7 Appendix C - Prospects

Table 2.10: Prospect Set A

		Leftward				Rightward			
	List	H	L	EV	SD	H	L	EV	SD
1	A.5	55	75	65	10	18	75	46.5	28.5
2	A.5	55	75	65	10	18	95	56.5	38.5
3	A.5	55	75	65	10	18	115	66.5	48.5
4	A.5	55	75	65	10	18	125	71.5	53.5
5	A.5	55	75	65	10	18	135	76.5	58.5
6	A.5	55	75	65	10	18	145	81.5	63.5
7	A.5	55	75	65	10	18	155	86.5	68.5
8	A.5	55	75	65	10	18	180	99	81
9	A.5	55	75	65	10	18	225	121.5	103.5
10	A.5	55	75	65	10	18	265	141.5	123.5
11	A.7	59	79	65	9.2	33	79	46.8	21.1
12	A.7	59	79	65	9.2	33	111	56.4	35.7
13	A.7	59	79	65	9.2	33	145	66.6	51.3
14	A.7	59	79	65	9.2	33	161	71.4	58.7
15	A.7	59	79	65	9.2	33	178	76.5	66.4
16	A.7	59	79	65	9.2	33	195	81.6	74.2
17	A.7	59	79	65	9.2	33	211	86.4	81.6
18	A.7	59	79	65	9.2	33	253	99	100.8
19	A.7	59	79	65	9.2	33	328	121.5	135.2
20	A.7	59	79	65	9.2	33	395	141.6	165.9
21	A.9	63	83	65	6	42	83	46.1	12.3
22	A.9	63	83	65	6	42	187	56.5	43.5
23	A.9	63	83	65	6	42	287	66.5	73.5
24	A.9	63	83	65	6	42	337	71.5	88.5
25	A.9	63	83	65	6	42	387	76.5	103.5
26	A.9	63	83	65	6	42	437	81.5	118.5
27	A.9	63	83	65	6	42	487	86.5	133.5
28	A.9	63	83	65	6	42	612	99	171
29	A.9	63	83	65	6	42	837	121.5	238.5
30	A.9	63	83	65	6	42	1037	141.5	298.5

Table 2.11: Prospect Set B

		Leftward				Rightward			
	List	H	L	EV	SD	H	L	EV	SD
1	B.5	78	92	85	7	26	86	56	30
2	B.5	78	92	85	7	26	106	66	40
3	B.5	78	92	85	7	26	126	76	50
4	B.5	78	92	85	7	26	136	81	55
5	B.5	78	92	85	7	26	156	91	65
6	B.5	78	92	85	7	26	166	96	70
7	B.5	78	92	85	7	26	186	106	80
8	B.5	78	92	85	7	26	206	116	90
9	B.5	78	92	85	7	26	236	131	105
10	B.5	78	92	85	7	26	296	161	135
11	B.7	81	95	85.2	6.4	42	89	56.1	21.5
12	B.7	81	95	85.2	6.4	42	122	66	36.7
13	B.7	81	95	85.2	6.4	42	155	75.9	51.8
14	B.7	81	95	85.2	6.4	42	172	81	59.6
15	B.7	81	95	85.2	6.4	42	205	90.9	74.7
16	B.7	81	95	85.2	6.4	42	222	96	82.5
17	B.7	81	95	85.2	6.4	42	255	105.9	97.6
18	B.7	81	95	85.2	6.4	42	289	116.1	113.2
19	B.7	81	95	85.2	6.4	42	339	131.1	136.1
20	B.7	81	95	85.2	6.4	42	439	161.1	181.9
21	B.9	84	96	85.2	3.6	52	92	56	12
22	B.9	84	96	85.2	3.6	52	192	66	42
23	B.9	84	96	85.2	3.6	52	292	76	72
24	B.9	84	96	85.2	3.6	52	342	81	87
25	B.9	84	96	85.2	3.6	52	442	91	117
26	B.9	84	96	85.2	3.6	52	492	96	132
27	B.9	84	96	85.2	3.6	52	592	106	162
28	B.9	84	96	85.2	3.6	52	692	116	192
29	B.9	84	96	85.2	3.6	52	842	131	237
30	B.9	84	96	85.2	3.6	52	1142	161	327

Chapter 3

Taxpayer's Behavior: From the Laboratory to Agent-Based Simulations

joint with Luigi Mittone

3.1 Tax Compliance: Theory and Evidence

As pointed out by Andreoni, Erard, and Feinstein (1998), the decision problem of tax evasion has been introduced in the economic literature just as an additional 'risky asset' to the household's portfolio. The first theoretical representations of individual taxpayers' compliance date back to the 1970s: the most influential, and probably the most criticized, rational choice models have been developed by Allingham and Sandmo (1972) and Srinivasan (1973b). These models portrait the taxpayer's decision problem as an investment choice involving a sure and a risky lottery, and adopt the formalization of Expected Utility Theory (Neumann and Morgenstern, 1944). Taxpayers are supposed to choose the extent of income declaration that maximizes their expected utility, defined according to income level, individual risk propensity, audit probability, and monetary punishment, in case of evasion detection. In this framework, they optimize the following function:

$$E[U] = (1-p)U(I-\theta D) + pU(I-\theta D-\pi(I-D)) \tag{3.1}$$

where p is the audit probability, θ is the tax rate, I is the actual income, D is the reported income, and π is the fine rate. These studies focus the effect of such parameters on evasion, and assume that the optimal proportion of evasion varies inversely with the likelihood of fiscal audits: a higher audit probability and/or a punishment proportional to the evaded tax reduce the expected value of evasion, and thus its attractiveness (Yitzhaki, 1974).

However, some issues on the aforementioned models have been raised, and their validity has been questioned, as the portfolio approach fails to address real-world complexity. Taxpayers are assumed to be able to determine the optimal proportion of tax to evade by making burdensome calculations, and having accurate information on the audit strategies adopted by the tax authority. Under these rather unrealistic conditions, such models predict that all taxpayers should evade if the audit probability and the fines most commonly used in reality were adopted. Furthermore, the label of 'tax compliance' is generally adopted to refer to a wide variety of behaviors - such as evasion of value added tax,[1] income underreporting, or tax burden reduction - which, though exhibiting remarkably different idiosyncratic characteristics, are treated with no distinction. For a recent review see Muehlbacher and Kirchler (2016).

Michael and Louls (1985) suggest that taxpayers' decisions cannot be entirely explained by the level of enforcement, as tax compliance is not only a matter of rates and penalties. Furthermore, as it is not easy to obtain precise information on the actual audit procedure that the tax authority adopt to discover tax evasion: taxpayers may not know the actual risk of being audited, and need to rely on their own estimate of such a risk, in order to make a compliance decision. Such an uncertainty about the probability of getting caught is likely to influence taxpayers' behavior.

This is supported by empirical evidence documented in many countries, coming from different sources such as random audits, surveys, laboratory and field experiments. Hence, many researchers, behavioral economists included, have tried to find models with a better fit for real taxpayers' behavior, and with a focus on potentially relevant psycho-sociological factors. This development process has often been built upon an experimental approach: just to list a few examples, the high degree of control and the greater parallelism with the natural world are among the main motivations leading researchers to turn to laboratory experiments, instead of relying only on theoretical analyses. In this sense, one can say that experiments on tax evasion are mainly motivated by economists' dissatisfaction with theoretical models: 'Rather than question the experimental method, [...] it is perhaps the theory which needs revision' (Baldry, 1987). In fact, experiments can provide a valuable support in studying people's behavior, especially when clandestine activities are involved: the

[1] In contrast, Mittone (2001), considered as a pioneer in the experimental investigation of VAT, recognizes the social nature of this kind of tax and explores it in an artificial market setting. Although VAT and income tax are strictly connected, since self-employed taxpayers evade VAT in order to reduce their tax liability on income, sellers do not decide by themselves, but need to collude with their customers. The psychological role, played by the need to find a collusive (illicit) agreement between buyer and seller, is the core issue of VAT evasion.

extent of tax evasion as a result of a specific interaction between micro and macro factors cannot be directly measured in an entirely natural and uncontrolled setting. It is not easy to collect evidence on tax compliance, and, even if this was possible, the specific conditions determining tax decisions would not be easily kept under control. In contrast, a detailed investigation of individuals' behavior is allowed by the laboratory approach, which can be considered as the proper system to understand specific real-world phenomena.

To this purpose, many extensions based on experimental findings have been proposed in order to integrate theoretical models, and make them closer to the intended domain of application: audits are costly for any audited person; the tax authority is distinct from the remainder of the government (Melumad and Mookherjee, 1989); tax collection is delegated (Sanchez and Sobel, 1993); moral and social dynamics, in terms of shame, moral rules, fairness to the tax code and its application (e.g., Erard and Feinstein, 1994; Spicer and Becker, 1980; Benjamini and Maital, 1985; Baldry, 1986; Gordon, 1989; Myles and Naylor, 1996; Torgler, 2002; Eisenhauer, 2006; Eisenhauer, 2008; Casal and Mittone, 2016), and evaluation of government expenditure and service provision (Cowell and Gordon, 1988) are included; the impact of the decision framework is tested. In this respect, Mittone (2002) finds that the introduction of an environmental structure closer to the one outside the laboratory fosters tax compliance thanks to the creation of social ties among participants. Finally, the investigation of taxpayers' views of the audit probability has shown the inadequacy of Expected Utility Theory for tax evasion (e.g., Friedland, 1982; Spicer and Thomas, 1982; Alm, Jackson, and McKee, 1992a; Alm, McClelland, and Schulze, 1992; Hessing et al., 1992; Sheffrin and Triest, 1992; Scholz and Pinney, 1995): people usually exhibit cognitive difficulties in estimating probability relationships and computing expected values (Einhorn and Hogarth, 1985; Casey and Scholz, 1991a; Casey and Scholz, 1991b), and the common uncertainty about the probability of being audited makes taxpayers' decisions more difficult than those made under the full information characterizing the laboratory environment. As traditional economic models rely on the unrealistic assumption that taxpayers have accurate information on auditing startegies, actual taxpayers' behavior cannot be predicted. Therefore, experiments with imprecise information appear to be more realistic.

A further step toward a greater realism in tax research is due to the introduction of Bounded Rationality (Simon, 1955; Simon, 1956). As suggested by Alm (1999), standard theoretical models, grounded on the simplifying assumptions of taxpayers' full rationality and homogeneity, should be revised.

These models rely on the adoption of a unique representative agent, and disregard the interaction among different types, while human behavior exhibits not only evident anomalies, but also a remarkable heterogeneity (Alm, McClelland, and Schulze, 1992): for instance, some individuals may overweight the occurrence of fiscal audits, or comply because they value what they are financing. Furthermore, laboratory experiments help prove that human beings are not always able to perform complex computations and to choose the utility-maximizing action. They are not making an investment decision in isolation, but are affected by many different 'emotional' factors and non-economic considerations that make their decision process rather complex to model.

However, all these aspects, which have been defined and first investigated at the micro level, as allowed by laboratory experiments and tax theoretical models, may have unexpected consequences and give striking results at an aggregate level. For example, the vast majority of macro-empirical research reports a strong deterrent effect of tax audits on evasion. In contrast, Gemmell and Ratto (2012) empirically explore compliance response to fiscal audit at an individual level, and observe contrasting results, due to many factors, such as the opportunity to underreport, and past audit experience. This implies that, in order to obtain relevant policy suggestions, neither of the two dimensions has to be disregarded. In this respect, as reviewed by Alm (2010), a growing number of researchers have adopted behavioral techniques, which rely on both human-based experimental economics and agent-based modeling, in order to address this micro-macro issue, and to gain new insights on taxpayers' behavior, which could not be observed otherwise. Micro-level experimental findings have widely shown that human agents are not rational, as assumed in theoretical models; on the contrary, they are guided by emotions, psychological and moral constraints, which might be mimicked in a computational simulation, as agents are calibrated according to human-based experimental evidence. In such a way, interesting and useful considerations can derive from agent-based modeling: on the one hand, it allows the implementation of a rather realistic system of individuals, with the intent of uncovering and testing specific cognitive aspects of taxpayers' decision process; on the other hand, the societal evolution, as due to an interaction among heterogeneous agents, can be studied from a macro perspective.

Hence, we point out that a synergic adoption of a human- and a computer-based approach gives the opportunity of gaining a deeper understanding of empirical phenomena or behavioral patterns, by scientifically studying a *valid* representation of these in the laboratory. Thanks to the combination of these two approaches, compliance decisions are studied both at an individual and a collective level: the exploration of the overall behavior of the society requires taking into account social interactions among heterogeneous and boundedly rational human beings. Agent-based simulations might provide a valuable support, since they can rely on realistic assumptions - i.e. behavioral regularities previously observed in the laboratory - and allow the implementation of complex settings in which both micro and macro factors interact and affect agents' behavior, as it usually happens outside the walls of the experimental laboratory.

3.2 Research on Tax Compliance: A Methodological Analysis

According to the previously presented evolution of research on tax compliance, an apparent challenge between economic theoretical models and the experimental approach seems to emerge. On the one hand, experimenters claim that anomalies observed in the laboratory are an important proof of the failure of theory in describing and predicting taxpayers' behavior in an accurate way. On the other hand, however, theorists reply that their models are intended to address phenomena taking place in the real-world, and not in the artificial environment of the laboratory. Alm, Jackson, and McKee (1992a) point out that '[...] experimental results can contribute significantly to policy debates, as long as some conditions are met: the payoffs, and the experimental setting must capture the essential properties of the naturally occurring setting that is the object of investigation. Laboratory methods may offer the only opportunity to investigate the behavioral responses to policy changes'.

In such a framework, the necessity of grounding experiments' external validity is evident: researchers claim that results are not always generalizable, i.e. applicable to the real world, because the environment reproduced in the laboratory is too simplistic and does not take into account many relevant variables. For instance, Webley (1991) argues that '[experimental] results may reflect a person's understanding of economics rather than the behavior that would be displayed in the real situation". The experimental setting might be perceived as too artificial and far from the environment outside the laboratory, if it is not a perfect replica of the real-world. The experimental system needs an external validity hypothesis, which maps laboratory elements onto elements of the phenomenon observed in the field. Only if this hypothesis holds, researchers can draw valuable inferences on individuals' decision process from the laboratory and move to the world outside (Guala, 2002).

The present chapter enters such a debate and explicitly focuses on the problem of tax experiments' external validity, adding to the small literature available on the topic. On the contrary, it disregards the internal validity issue, which has already received much attention in the experimental literature. Specifically, the novelty of this methodological review resides in the proposal of a synergic approach, involving both human-based observations and agent-based simulations as a valuable tool aimed to solve the problem of tax experiments' external validity.

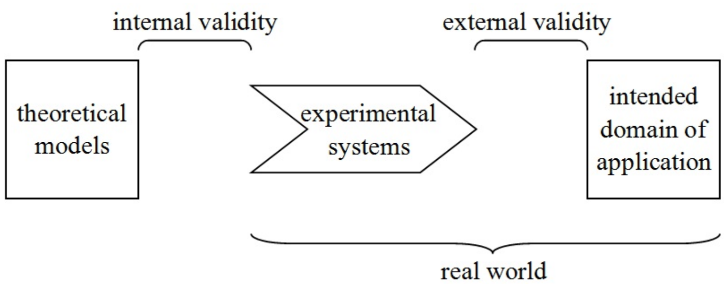
In a very recent contribution, Muehlbacher and Kirchler (2016) address this methodological issue, providing an interesting review of both experimental and empirical research on tax compliance. The authors point out that little is known about the external validity of tax experiments, and identify a number of criticisms: in addition to a rather general critique on artificiality, participants' self-selection, experimenter effect, social desirability, and social blaming, a more detailed methodological review on tax research is offered. For instance, as already well documented in previous studies, income-reporting decisions in a tax setting systematically differ from those in an abstract setting (Baldry, 1986; Alm, McClelland, and Schulze, 1992; Mittone, 2006; Choo, Fonseca, and Myles, 2016); the introduction of a redistribution mechanism strongly affects taxpayers' decisions (Alm, Jackson, and McKee, 1992b; Alm, McClelland, and Schulze, 1992; Mittone, 2006); students might not be representative because they have no experience in paying taxes (Webley, 1991). Furthermore, compliance depends on the way in which subjects' income is provided (Boylan and Sprinkle, 2001; Boylan, 2010; Durham, Manly, and Ritsema, 2014). Finally, in reality, there is a temporal distance between compliance decisions and audits, which might have a significant effect on actual compliance decisions, and make experiments disregarding this issue less reliable (Kogler, Mittone, and Kirchler, 2016). Based on this, Muehlbacher and Kirchler (2016) suggest that experimental investigations in the laboratory should induce the same psychological mechanisms taxpayers adopt outside, and take into account possible interactions between treatment factors and setting characteristics. If such requirements are met, experimental findings can be applied also outside the laboratory, and thus provide useful insights for policy interventions.

According to Alm, Sanchez, and De Juan (1995), 'a government compliance strategy based only on detection and punishment may well be a reasonable starting point but not a good ending point. Instead, what is needed is a multi-faceted approach (...) Put differently, explaining tax compliance requires recognizing the myriad factors that motivate individual behavior, factors that go much beyond the standard economics-of-crime approach to include theories of behavior suggested by psychologists, sociologists, and other social scientists. Until this effort is made, it seems unlikely that we will come much closer to unraveling the puzzle of tax compliance.' Following the same approach, Guala and Mittone (2005) get into this debate on experiments' role and external validity, suggesting that experiments might help theoretical models to get closer to real-world phenomena, and thus answer specific questions about causal relationships.

From this viewpoint, Guala and Mittone (2005) claim that experiments serve as *epistemic mediators* between theoretical models and empirical economic phenomena. In fact, theory and experiments are not considered as two distinct entities: they both require initial hypotheses and inference; they are two useful and complementary structures to study and subsequently understand economic behavior. Figure 3.1 shows this relationship as presented by Guala and Mittone (2005): They identify a gap between theoretical models and the intended domain of application, and experimental systems occupy the middle ground between the two. Nevertheless, both experiments and targeted economic phenomena belong to the same 'real world': in fact, according to the authors, compared to theoretical models, experimental systems are closer to the target, since they actually allow the collection of observations of real people's behavior under specific conditions, though in an environment that has been artificially manipulated by the experimenter.

Figure 3.1 also refers to the way in which the gap - or better the gap between theory and experiments, and the one between these and the target - can be closed. On the one hand, *internal validity* - in terms of testing different hypotheses in isolation by controlling for confounding variables, and ruling out undesired effects - bridges the gap between theoretical models and experiments. On the other hand, *external validity* - in terms of laboratory identification of mechanisms that characterize also the targeted phenomena - is intended to bridge the gap between experimental systems and the specific domain of application. While the former has received much attention in the economic literature, little can be found on the analysis of external validity (Muehlbacher and Kirchler, 2016). The problem of internal validity can be 'easily' solved by adopting a number of techniques allowing the identification of causal relationships. However, even a high degree of internal validity does not ensure that the external validity requirement is met. Experiments provide a higher degree of concreteness with respect to theoretical models, by including features that could be reasonable for externally valid inferences. However, they are still artificially isolated from the world outside the walls of the laboratory, in which a wide variety of micro and macro factors - tax audit plans, risk preference, reasoning biases, moral constraints, social norms, social comparison, interaction and imitation, fairness, trust, just to name a few - interact in determining actual taxpayers' behavior. Experiments try to implement these factors, yet under the constraint of balancing between internal and external validity: an excessively complicated experimental setting impairs the identification of clear causal effects, and makes experimental results harder to interpret (Cowell, 1991). Therefore, most laboratory experiments are not able to perfectly replicate the specific targeted phenomena, and feed back into the theoretical literature. Experiments can help at an intermediate stage, as they cannot bridge the gap between the target and the theoretical model: the highly controlled experimental setting is aimed to determine which theory better explains a certain pattern of data, but this explanation might be not valid outside the laboratory (Guala, 1998; Guala, 1999; Guala, 2003).

FIGURE 3.1: Experimental systems as mediators between theoretical models and economic phenomena



Nevertheless, as suggested by Guala and Mittone (2005), experiments might also be intended to discover new real and robust empirical phenomena, not necessarily explained by existing theories to be tested in the laboratory. These phenomena might include generic psychological effects, biases, heuristics to be applied to specific empirical situations. In such a case, experiments do not need to perfectly reproduce the target, but may contribute to the creation of a *library of phenomena* (Guala and Mittone, 2005): they simply discover new facts useful from a policy perspective.

In addition, Guala and Mittone (2005) propose interesting examples of robust biases involved in probabilistic reasoning and the effects of uncertainty that have been identified and extensively studied in the laboratory and in the field, also as strictly related to research on tax compliance. Both Sheffrin and Triest (1992) and Scholz and Pinney (1995) perform an econometric analysis of the influence of taxpayers' perceived probability of detection on compliance decisions. The former analysis finds that individuals who perceive a higher audit probability expect significantly less evasion in the population, and those not trusting others or the government engage in more evasion. Nevertheless, such an analysis solely relies on survey data, and therefore, as suggested by Andreoni, Erard, and Feinstein (1998), the results could be biased by the coherent image individuals tend to convey in surveys. In contrast, Scholz and Pinney (1995) also collect tax-return data, and their analysis is intended to investigate the extent of people's guilt and moral obligation, by testing the duty heuristic hypothesis: if taxpayers have no accurate information on the probability of detection, they can rely on heuristics to derive subjective estimates of the risk and to make their compliance choice. The authors observe a significant positive relationship between subjective probability and duty, which in fact leads to an overestimation of the risk of getting caught, and therefore to a higher degree of compliance. Such an effect is even strengthened by people's tax knowledge and previous contacts with the authority. This evidence is supported also by Hessing et al. (1992): although, according to Andreoni, Erard, and Feinstein (1998), their results seem to partially contradict those reported by Scholz and Pinney (1995), it emerges that the duty is fostered by mere contacts between taxpayers and the tax authority, while it is impaired by previous audits and fines. In fact, traditional enforcement activities built on coercive power seem to negatively affects taxpayers' sense of duty (Kirchler, Hoelzl, and Wahl, 2008); tax agencies prefer to adopt an horizontal monitoring approach, by treating taxpayers as customers to whom they can provide useful services.

Friedland (1982), Spicer and Thomas (1982) and Alm, Jackson, and McKee (1992a) manipulate the quality and the accuracy of information on fines and probabilities in the laboratory: they observe that a higher degree of informational ambiguity enhances compliance. Nevertheless, as theoretically proved by Snow and Warren (2005), such an effect strictly depends on individual ambiguity aversion.

From a similar viewpoint, Bernasconi (1998) suggests that the portfolio approach needs to be integrated with subjects' probability weighting. The non-linear weight function proposed according to Rank Dependent Utility models (Quiggin, 1982), and Prospect Theory (Kahneman and Tversky, 1979), may describe the higher degree of compliance actually observed, compared to the theoretically predicted low level. In this respect, Prospect Theory provides new approaches to modeling tax evasion decisions (Schepanski and Shearer, 1995; Dhami and Al-Nowaihi, 2007; Ping and Tao, 2007; Trotin, 2010; Piolatto and Rablen, 2014; Piolatto and Trotin, 2016), by taking into account probability weighting, and reference-dependence (Copeland and Cuccia, 2002; Bernasconi and Zanardi, 2004; Watrin and Ullmann, 2008).

Also Erard and Feinstein (1994) underline the significant impact of probability weighting on taxpayers' decisions: in order to provide useful and reliable behavioral insights, fiscal models have to take into account the difference between actual audit probabilities, and estimates. In support of the occurrence of this probability weighting process, Spicer and Hero (1985) build a repeated measurement setting, and find that the extent of under-reporting diminishes as the number of previous undergone audits increases. This evidence has been explained by the *availability heuristic*: people tend to rely on immediate examples they recall when evaluating a decision problem (Tversky and Kahneman, 1973). An alternative explanation is the *target effect*, according to which people assume that a fiscal investigation is likely to be followed by another one (Hashimzade, Myles, and Tran-Nam, 2013).

In contrast, Mittone (1997) reports on an experiment investigating the difference between probability subjective estimation and weighting: tax-payers exhibit overestimation, when simply asked to judge the probability of being audited, and underweighting, when asked to actually make a decision. Specifically, according to their estimated probability, compliance is expected to ensure a higher expected value than evasion does; however, in the compliance decision, evasion is the predominant choice.

More detailed analyses of the dynamics underlying taxpayers' decisions in a repeated-measurement framework, mimicking a 'taxpaying life span', are provided by Mittone (2006) and Kastlunger et al. (2009). In contrast to Bayesian updating, such that audited taxpayers have higher estimates of audit probability than non-audited taxpayers, and are more deterred from evasion, these authors observe that the occurrence of an audit seems to make taxpayers more prone to evade. This result is commonly referred to as the *bomb crater effect*: the probability of observing compliance decreases if a taxpayer has just undergone a fiscal audit. According to Guala and Mittone (2005), this phenomenon observed in the laboratory has to be tested under a variety of conditions in order to verify whether it exhibits robustness and external validity. As for the former property, Kastlunger et al. (2009) find similar results and report that the decrease in compliance after an audit is very rarely due to loss-repair tendencies: the decrease in compliance seems not to depend on whether the taxpayer is fined in the previous round or found to be compliant. As for the latter, it might not be easy to observe the bomb crater effect outside the laboratory: in many countries, variability in declarations increases the probability of being investigated. Nevertheless, such an effect might emerge under specific conditions. Studies about the impact of audits on subsequent compliance have shown that the decline in compliance after an audit can also be observed in real taxpaying situations (DeBacker et al., 2015), and not only with respect to income tax. Bergman and Nevarez (2006) analyze VAT data from individual tax return information in Argentina and Chile, and identify the effect; however, authors also argue that taxpayers who evade more tend to be less deterred by audits. The bomb crater effect in VAT evasion is also confirmed by the experimental research project 'Tax morale among self-employed and their customers: A psychological comparison of value added tax versus income tax' I have been conducting in joint collaboration with J. Olsen, C. Kogler, L. Mittone and E. Kirchler. We actually observe that subjects who have just undergone a fiscal audit are more prone to evade - i.e. to offer their customers a VAT-esclusive price - irrespective of whether they are found to comply.

The so-called *echo effect* is another laboratory phenomenon Guala and Mittone (2005) deal with. Mittone (2006) studies the effect of different patterns of audits over time, and finds that frequent audits experienced early in 'tax life' may lead to higher compliance at later stages. Guala and Mittone (2005) report that this phenomenon is robust to changes in the experimental setting, and suggest that this laboratory evidence is supported by a number of real life examples: for instance, fare evasion on Italian public transport is increased by the experience of infrequent controls. Therefore, it seems reasonable to assume that taxpayers evaluate or weight the audit probability according to their experience: repeated audits may lead to a decrease in evasion even in the long run because of chance misperception. Taxpayers learn that the likelihood of audits is higher than the objective probability when these are rather frequent in the beginning; therefore, they rely on this sample to form their probability evaluations, and stick to a high compliance level even when the frequency of investigations diminishes.

In summary, our analysis starts recalling the approach by Guala and Mittone (2005) who identify the mediator role of economic experiments in the study of empirical phenomena. Experiments rely on hypotheses and allow the investigation of specific, framed and concrete settings: individuals' decisions are real, though in an artificial environment. We recognize the undeniable relevance of experimental systems in supporting theoretical models and providing better insights on empirical regularities, which otherwise could not be studied and understood so clearly outside the laboratory.[2] For this reason, experiments call for internal validity, while an a priori external validity is not necessary: as previously pointed out, experimental investigations might contribute to the identification of robust economic and psychological phenomena that can be borrowed and applied to specific cases inside or outside the laboratory. At the same time, however, it is also true that, in order to increase their reliability, experimental findings might need to be further tested before valid inferences are drawn.

In this respect, we provide a novel contribution to the literature on experimental methodology in tax research, by extending the framework presented by Guala and Mittone (2005) and claiming that agent-based simulations offer a valuable support. In fact, both theoretical economic models and related experiments are mainly defined in a microeconomic setting and they address empirical issues with a high degree of specificity. In addition to this, experiments cannot control for all cognitive drivers involved in the decision process of tax compliance, but only for those specifically targeted and isolated by the experimental design. In this framework, a computational approach to the study of tax evasion tests not only the robustness of experimental findings, but also their external validity. On the one hand, agent-based simulations may provide valuable insights of cognitive nature, which an experimenter would not be able to get simply observing the behavior of a limited sample of human subjects in the laboratory. Human-based experiments contribute to the library of phenomena; computer-based simulations aim to validate laboratory findings, and help understand complex cognitive processes involving psychological biases and heuristics. On the other hand, simulations allow the combination of micro- and macro-level factors actually interacting outside the laboratory and determining people's compliance.

[2]This is exemplified by the research presented in Chapter 4: the laboratory experiment aims to mimic a highly specific empirical setting, and relies on previous and robust experimental results taken from the more generic literature of decision-making under uncertainty.

3.3 From Human-Subject to Computational-Agent Experiments

From the previous analysis, it is evident that a pure theoretical approach may offer an 'unrealistic [or better, incomplete] picture of human decision-making', which is neither based on nor confirmed by empirical evidence (Selten, 2001). Therefore, it requires to be mediated by an experimental approach, in order to effectively target the empirical domain of interest. This implies the adoption of heterogeneous and less strict assumptions on individual behavior. Nevertheless, in spite of helping theoretical models target specific phenomena, some experiments might still lack external validity. On the one hand, decisions observed in the laboratory are real and the choice setting is specifically intended to address the issue of interest; on the other hand, human samples usually are rather small, and the setting might turn out to be too simple and prevent valid inferences to be transferred outside the laboratory. According to the *materiality thesis* by Guala (2002), experiments may not display a formal similarity to the complex framework of the target system, though being able to replicate almost the same causal processes taking place in the real world outside the laboratory. Therefore, relying on the assumption that human beings are basically the same inside and outside the laboratory, it is possible to identify a correspondence at a 'material' level between the experimental and the target system, but not necessarily at a 'formal' and 'abstract' level, which might hinder experiments' external validity.

In this framework, Agent-based Computational Economics (ACE) may significantly contribute to the development of more realistic decision-making models; it helps bridge the gap between economic models and the intended domain of application. Similar to theoretical and experimental approaches, simulations require a formal definition of behavioral types. As a matter of fact, ACE is not intended to disregard theoretical considerations, as relationships describing human behavior need to be known in advance for the calibration of agents. Nevertheless, simulations can rely on behavioral assumptions (and experimental observations) so that different agent types are defined, the standard neoclassical economics idea of a homogeneous representative agent is overcome, and a realistic replication of the world is provided. In this sense, both the theoretical and the experimental analysis are enriched by the introduction and the robustness check of heterogeneous behavioral patterns, which might be designed according to previously collected empirical and experimental evidence.

Nevertheless, in contrast to laboratory experiments, according to the ontological analysis by Guala (2002), simulations rely on a process of abstraction: the external validity requirement might be hardly met at a material level, as the correspondence between the simulating and the target system is of a more 'formal' kind. ACE agents are virtual entities endowed with specific attributes, purposes, and behaviors; they interact with a rather complex landscape - it consists of institutions, enforcement rules, social networks, etc. - which, in general, resembles the real-world but cannot be replicated in human experiments; they receive an input, and, based on this, select an action allowing them to reach their pre-defined motive goals, such as wealth, happiness, or honesty.

Nowadays, in the field of economic behavior, the spectrum of possible experimental methodologies is quite broad and ranges from 100% human-subject to 100% computational agent experiments. These two extremes were first thought to be either in opposition or completely unrelated: until a few years ago, the majority of ACE researchers did not consider human-subject experiments as a valuable and real source of information and results in order to build and calibrate simulation models, as reviewed by Duffy (2006). Similarly, the great majority of experimentalists tend to exclusively rely on human-based tests or explorative investigations, without trying to increase the potential and the extent of their experimental results by means of computational simulations.

However, these two methodologies tend to converge: half way, different techniques, such as a mixture of human and computer agents interacting with each other, human-calibrated computer agents, and computer agents with real world data streaming, are gaining relevance. Researchers admit that the laboratory with human subjects is a rather artificial context: time is compressed, subjects are asked to make unnatural repeated decisions, so that lifetime span can be mimicked, and the landscape is fully controlled and manipulated by the experimenter. Experimental design factors, such as round numerosity, are strictly related to the specific aim of the investigation: even a few rounds are enough to study some simple learning processes, while a higher number of repetitions is necessary if more complex behavioral dynamics are investigated. Nevertheless, an excessive increase in the number of rounds can often harm the results' reliability, as participants get bored. Therefore, from this viewpoint, well-designed experiments allow researchers to carefully study and deeply understand simple dynamics and individual behaviors. In a complementary way, simulations permit to disregard the boredom issue and analyze more complex and dynamic behavioral processes over an extended period of time and among heterogeneous agents: in order to see the emergence and the evolution of behaviors over time, and investigate cognitive processes, ACE researchers implement artificial agents that make decisions and react to consequences and signals. As this approach is based on heterogeneous and predominantly boundedly rational agents acting within a dynamic environment, it extends the idea of the representative agent that does not evolve, is fully rational and endowed with an unlimited computational power.

In doing this, simulation models may rely on data from human-subject experiments. The agent-based methodology can be used to understand results from human-based studies, since it allows the exploration of the decision process in a more complex economic environment, by replacing humans with agents. The potential of experimental results can be increased by means of these computational tests: it is possible to explore the psychological mechanisms giving rise to phenomena whose robustness and external validity can be checked. Simulations relate the micro-level (i.e., agent-level) behavior to macro-level (i.e., system-level) dynamics, represent multiple scales of analysis in a natural way, and investigate adaptation and learning. Agents are built on experimental evidence, and behave according to actually observed heuristics; in addition, the implementation of an interaction among different agents over time provides insights on macro evolution, which could not be investigated in a simpler human-based experiment. The 'formal' similarity ensured by simulations is combined with the 'material' one provided by laboratory experiments in a complementary way. The potential of both methodologies is exploited in order to meet the external validity requirement (Guala, 2002). Therefore, from this viewpoint, not only simulations contribute to the external validity hypothesis of experimental systems, but, in turn, experiments increase ACE studies' validity, which is considered as one of the key aspects to judge the performance of a computational model (Taber and Timpone, 1996). In fact, simulation results can be tested in the laboratory in order to better grasp human behavior in computational settings, and observe whether and why computer and human behaviors differ. Collected data can feed the software model, and contribute to ameliorate agent-based predictions of real world economic behaviors, and ground them on a material basis, rather than a merely formal one.

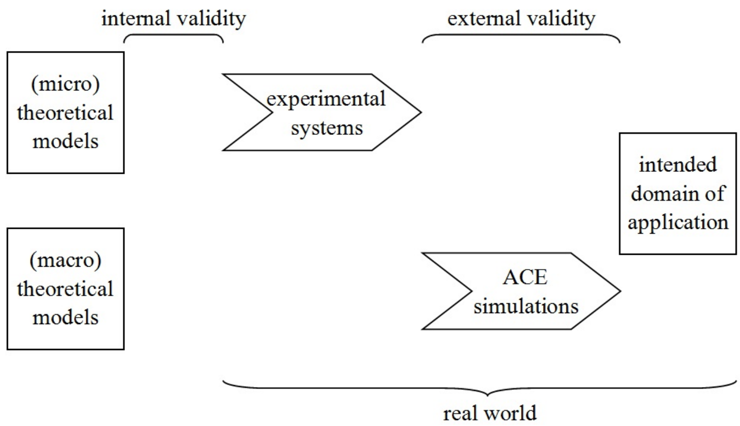
Based on this synergic approach, agents' behavioral traits are no longer defined only according to simplifying theoretical assumptions, but to observations actually taken from the real world: behavioral regularities discovered in economics and psychology experiments (e.g., Andreoni, Erard, and Feinstein, 1998; Mittone, 2002; Mittone, 2006; Kirchler, 2007) can be used to calibrate and/or test simulation models, which, in turn, can help check and explain experimental results. Therefore, both approaches gain in external validity: the high number of degrees of freedom in agent-based models can be managed with human calibration, and human-based experiments, which are not always able to perfectly manipulate subjects' behavior and control their cognitive processes, can find a further confirmation in simulations.

In light of the above, the complementarity of simulations and laboratory experiment also emerges as a support to external validity: the 'formal' similarity between simulations and real-world phenomena can be combined with the 'material' similarity characterizing the relationship between the experimental and the target system. On the one hand, the simulations' need for a relevant background knowledge can be met by means of an experiment-based calibration: evidence on human decision processes are collected in the laboratory and used to feed simulated agents, so that they can resemble real decision-makers also at a more material level.[3] On the other hand, the mere materiality of laboratory experiments is enriched by simulations' formal correspondence to reality: people's behavior is first observed in a rather simple and artificial laboratory setting; then, it is further investigated and tested in a more realistic environment, in which human-calibrated agents interact. Therefore, it seems possible to conclude that none of the two methodologies has epistemic privilege over the other (Parke, 2014): 'material' and 'formal' correspondence should be used in a complementary manner, so that each methodology can take advantage from the other while addressing the common external validity issue.

A graphical analysis of this relationship between experiments and ACE models is provided in Figure 3.2: the framework adopted by Guala and Mittone (2005) is extended in order to include agent-based simulations as a support for experiments in studying empirical economic phenomena and bridging the external validity gap. The figure shows that ACE simulations rely on both macro and micro theoretical models: in fact, they allow the investigation of the evolution over time of network systems involving heterogeneous individuals, and these heterogeneity is built upon micro experimental evidence. In this sense, also simulations belong to the real world, as experimental observations are used for agents' calibration, and complex and more complete settings can

[3]According to Winsberg (2009), the main difference between simulations and experiments depends on the prior 'knowledge that is invoked to argue for the external validity of the research'.

FIGURE 3.2: ACE simulations supporting experiments



be implemented, to the purpose of making experimental results more likely to be externally valid. Agent-based models rely on preference assumptions but they exhibit a high degree of complexity with respect to human-based experiments, as they mimic societies made of heterogeneous individuals. Not only human-calibrated agents can be endowed with diverse attributes, such as income level, risk propensity, compliance preferences, norm adherence, heuristics, biases, etc., but also various policy parameters, and the effect of these on the interaction among agents can be taken into account.[4] This allows both the investigation of taxpayers' cognitive process, and the combination of micro-level evidence and macro dynamics among heterogeneous agents in a unique decision setting resembling the economic environment of interest. The analysis of tax compliance dynamics in a pretty realistic, though complex, system may lead to discover new and efficient policy options (Garrido and Mittone, 2013; Pickhardt and Seibold, 2014), which could take into account the variety of reactions emerging in a population of heterogeneous taxpayers.

[4]Despite the quite recent development of agent-based modeling in the literature of tax evasion, two distinct strands can be already identified. Models belong to the economic domain if the interaction process is due to a change in parameter values in the taxpayer's utility function (e.g., Mittone and Patelli, 2000; Davis, Hecht, and Perkins, 2003; Bloomquist, 2004a; Bloomquist, 2004b; Antunes et al., 2007; Bloomquist, 2007; Korobow, Johnson, and Axtell, 2007; Bloomquist, 2006; Bloomquist, 2011b; Bloomquist, 2011a; Hokamp and Pickhardt, 2010; Méder, Simonovits, and Vincze, 2012). In contrast, models with an interaction process driven by statistical mechanics fall in the domain of econophysics/sociophysics (e.g., Lima and Zaklan, 2008; Zaklan, Lima, and Westerhoff, 2008; Zaklan, Westerhoff, and Stauffer, 2009; Hokamp and Pickhardt, 2010; Lima, 2010; Lima, 2012b; Lima, 2012a; Seibold and Pickhardt, 2013; Hokamp and Seibold, 2014a; Hokamp and Seibold, 2014b; Pickhardt and Seibold, 2014; Bazart et al., 2016). Agents are not endowed with a utility function; their behavior is described as a stochastic process affected by the changing balance between individual's autonomy and influence from neighbors' behavior. This chapter focuses on the economic domain.

3.4 An Agent-based Approach to Taxpayers' Behavior

With a closer focus on tax experiments, this section proposes two separate ACE approaches intended to pursue the aforementioned goal of filling the external validity gap: both of them are aimed to tackle the limitations of full rationality and behavioral homogeneity, which impair the external validity of theoretical and experimental claims.

Firstly, agent-based models may analyze the interaction among *types* and study the subsequent emerging macro dynamics; this is mainly based on the implementation of recurring behavioral styles in the population of taxpayers previously identified in laboratory experiments (Mittone and Patelli, 2000; Davis, Hecht, and Perkins, 2003; Antunes et al., 2007; Hokamp and Pickhardt, 2010; Hokamp, 2014). Due to their scope, these models are usually characterized by a modest degree of granularity: they try to tackle the unrealistic theoretical assumption of a lack of heterogeneity in taxpayers' behavior, yet without always addressing the bounded rationality issue. They are not intended to explore individuals' cognitive dynamics; therefore, behavioral types are specified as rather simple agents. This interest in the identification of groups of taxpayers dates back at least to the 1990s. Building on Cowell (1991), Hessing et al. (1992) identify three behavioral types according to willingness to comply, and underline the importance of behavioral heterogeneity to evaluate the extent of efficiency and effectiveness of different policy instruments: some auditing strategies might have the negative impact of crowding out honesty, and thus reducing individual willingness to comply; in contrast, an efficient strategy might fight tax evasion by sustaining honesty and compliance. In this respect, different from human-based experiments, agent-based simulations allow the implementation and manipulation of population heterogeneity in a highly controlled manner, so that this, and its interaction with other variables, can be treated and analyzed as a determinant of policies' efficacy. In a synergic view, such simulation results can be subsequently tested on human subjects.

Secondly, simulations may also look into *micro behavioral patterns* that go beyond the macro type specification. Therefore, they are characterized by a higher granularity, since agents are more complex in their attributes. In fact, in this case, human behavior is first investigated at an individual level in the laboratory, and then reproduced by means of artificial agents (Bloomquist, 2006; Garrido and Mittone, 2008; Méder, Simonovits, and Vincze, 2012): simulations help uncover and understand human cognitive processes and psychological drivers, which cannot be fully investigated in a purely human setting. Therefore, this kind of analysis is well suited to the implementation of boundedly rational decision-makers, who are intended to mimic human subjects, and choose according to a restricted set of information.

The following sections provide an exemplification of the methodological validity of combining human- and agent-based techniques in the study of tax phenomena. To this purpose, the aforementioned distinction between mainly macro or micro computational analysis is adopted; in addition, some attempts of reconciling such a distinction are analyzed (e.g., Korobow, Johnson, and Axtell, 2007; Garrido and Mittone, 2013; Mittone and Jesi, 2016). The analysis of these simulation examples is aimed to offer a guidance in the implementation of the synergic approach, involving both human- and agent-based models, with the intent of filling the external validity gap of economic experiments. Providing an extensive review of research with human calibrated models is, instead, beyond the scope of the present chapter.

3.4.1 The Macroeconomic Approach

Mittone and Patelli (2000) carry out a dynamic simulation in order to model a fiscal environment in which different types of taxpayers interact and, according to their degree of compliance, a public good is provided[5]. The two authors investigate taxpayers' psychological and moral motives by using human calibrated simulation models. The idea of studying specific taxpayers' behavioral traits is developed in the seminal work by Mittone (2002): he categorizes behavioral regularities, and identifies classes of subjects reacting in a similar way to certain *economic* and *moral factors*.[6]

Mittone (2002) verifies whether subjects' behaviors can be captured and classified in homogeneous categories, by performing a cluster analysis.[7] He

[5]The two authors specify that, given their focus, what they refer to as 'agent' or 'taxpayer' (taken as synonyms in this context) is a 'taxpaying behavior', defined as a form of reaction to the introduction of a tax.

[6]The former are exemplified by income level, tax rate, audit probability, and fees, which enter taxpayers' utility function. Moral factors are investigated by manipulating the decision context in a between-subject fashion: the effects produced by either the introduction of a tax yield redistribution, which depends on all taxpayers' compliance decisions, or the lack of any reference to the fiscal environment are tested. This not only allows the researcher to study the role of moral constraints, but also to check the robustness of the emerging categories over different settings.

[7]For this kind of analysis, the author adopts the average linkage between groups method (also called UPGMA - Unweighted Pair-Group Method using arithmetic Averages), and uses standardized variables. According to this method, the distance between two groups is the

finds four main clusters:[8] the great majority of subjects do not exhibit a stable behavior, in line with the intuition that previous experience affects taxpayers' decisions. Behavioral clusters are almost identical across experimental conditions. Results confirm the difficulty of modeling and explaining the actual dynamics of taxpayers' behavior by simply relying on the traditional expected utility approach: refinements based on empirical and experimental observations are necessary, in order to understand the interaction between behavioral heterogeneity and enforcement policies. As pointed out by Mittone (2002), contrarily to rational predictions, participants seem not to be comfortable with repeated choices under risk, and alternate opposite choices, probably because the ongoing interaction with the environment leads them to weight probabilities, and not to stick with a predetermined pure strategy. However, he also reports that tax yield redistribution triggers honesty. This change in the composition of the population due to the institutional setting might have serious policy implications: these experimental results seem to suggest that the policy maker should implement fiscal plans designed according to the institutional setting, and subsequently exploit the composition of taxpayers' population to foster honesty imitation, as proposed by Hessing et al. (1992). To this purpose, a valid support is offered by ACE simulations, which test the efficiency and the efficacy of different enforcement strategies on a large population consisting of a realistic variety of human-calibrated types. This computational approach adds to the experimental one, since it manipulates the composition of the population, and thus controls for the macro effects deriving from the interaction among different behavioral types under given institutional and fiscal settings.

Building on this, Mittone and Patelli (2000) study the true nature of tax compliance, by focusing not only on the effect of tax authority enforcement, but also on social interaction and moral concerns. They focus on the coexistence of three different behavioral styles which require agents' interaction in order to evolve, and adopt a computational approach that might contribute to the validity of experimental findings. The work by Mittone and Patelli (2000)

average of the distances between all pairs of individuals (i.e. by taking one individual for each of the two clusters).

[8] The four styles are: (i) the (pure) 'absolute stability' of subjects always paying the tax due; (ii) the 'relative stability' of subjects always evading, yet at different extents; (iii) the 'oscillatory behavior' between full compliance and partial evasion; (iv) the 'mixed behavior' of subjects adopting the oscillatory behavior in the first half of the session, and fully evading in the second half. The large majority of subjects exhibited a stable behavior: it was rare to observe a drastic change in their 'style'.

can be considered as a good example of the synergic approach of ACE simulations and human-based experiments aimed to provide greater realism and concreteness for the subsequent application of results to specific policy targets. In fact, Mittone (2002) identifies in the laboratory some behavioral regularities; Mittone and Patelli (2000) not only test the robustness of laboratory findings, but also study the macro evolution of a dynamic and heterogeneous population facing different enforcement systems. Behavioral types identified with the experimental micro approach are used to calibrate agents, whose behavior is analyzed from a macroeconomic viewpoint: type imitation and population evolution are the main scopes of this ACE investigation. These types include the *honest taxpayer*, the *imitative taxpayer*, and the *perfect free-rider*.[9] Agents share a decision algorithm, leaded by utility maximization, while each type has a unique utility function that specifies its behavior.[10] This allows the implementation of heterogeneous agents, and aims to extend microeconomic models towards the investigation of behavioral evolution and evasion activity in a population of interacting taxpayers, who have different preference structures.

In such a system, at regular intervals, a genetic algorithm can be activated in order to update the composition of the population, without modifying the overall number of agents. The two authors set an initial scenario, and observe how a given population composition evolves over time: taxpayers initially belong to one of the three categories, but then they can decide to switch to another type, according to the degree of success of their style in pursuing the goal of utility maximization.

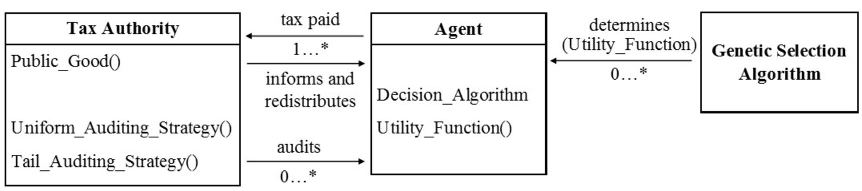
This optimization strictly depends on tax-payment decision and the risk of being investigated. In fact, in each round, a fixed number of agents are audited according to either a uniform auditing (all agents have the same probability of being investigated) or a low-tail auditing strategy (agents who report the lowest amount of tax have a higher probability of being audited). This

[9]This classification resembles the one identified and investigated by Fischbacher, Gächter, and Fehr (2001) and Burlando and Guala (2005): unconditional cooperators, conditional cooperators, and free-riders.

[10]The utility function is built on the work by Myles and Naylor (1996), in which tax compliance is assumed to be affected by social customs and group conformity. The *honest agent* derives additional utility, proportional to the percentage of honest taxpayers in the population, form behaving in accordance to the social norm of compliance. The *imitative agent*'s utility function depends on the amount of tax he should pay, and on the average amount of tax paid by the population. Finally, the *freerider agent* derives a positive utility from adopting an opportunistic behavior. Moreover, they all get a non-negative utility from the public sector: the model hypothesizes the existence of a single public good, which is produced every round thanks to the tax yield previously collected.

diversification is intended to investigate the effect that different auditing policies produce on taxpayers' behavior, also depending on the specific degree of population heterogeneity. Figure 3.3 shows the functioning of the simulated economy: it is evident that agents' decisions are determined by both social interaction, and the enforcement activity of the tax authority.

FIGURE 3.3: System structure diagram designed after Mittone and Patelli (2000)



In summary, Mittone and Patelli (2000) study the macro experimental interaction among behavioral types identified by means of a microeconomic approach; they test the efficacy of different audit strategies in fighting tax compliance when population heterogeneity is not the result of abstract assumptions, but of real-world observations. Such a controlled exploration of the interaction of different population compositions with the environment would not be easily implementable in a purely human-based setting. This justifies the adoption of an agent-based approach, which allows the observation of macro behavioral dynamics, but needs human calibration for the implementation of realistic taxpayers.

Results show that a uniform auditing strategy is more effective than a low-tail one in fostering compliance; imitating the honest behavior is a winning strategy when low-tail auditing is implemented. Finally, genetic selection favors honesty, as frequently observed also inside and outside the laboratory when moral concerns on contribution are involved in the decision process. When taxpayers are aware that they will actually benefit from their fiscal contribution, they appear to be more prone to comply.

Hence, if combined together with theoretical models and laboratory experiments, agent-based simulations can help understand and explain behavioral processes underlying tax payment decisions. The novelty of this study resides in the implementation of a simulation model investigating the relationship between enforcement activity and social interaction among different behavioral styles, that have emerged as regularities in previous human-based experiments. Taxpayers are heterogeneous and their behavior is described by utility functions; furthermore, they can switch to a different type according to the 'satisfaction' they are able to derive from the behavior of their own category. Nevertheless, agents' decision-making process still relies on optimization, and depends only on the information they receive from the system. In this sense, they are myopic, since they are not designed to take into account either intertemporal or strategic expectation on the evolution of the environment. For this reason, a further development in this direction might include a distinction between naive and sophisticated agents, where the latter should be modeled in order to mimic agents capable of making efficient predictions about their future behavior and that of their mates.

3.4.2 The Microeconomic Approach

Besides this macro approach, mainly focused on the analysis of the effect of behavioral heterogeneity, a parallel line of research is based on the assumption of boundedly rational agents has gained relevance, too: the micro dimension of individual history serves as a base for ACE simulations aimed to understand and explain decision-makers' cognitive process. Under this view, decisions are expected to vary according to individuals' *state*, which is determined by external environment and past experience, and is translated in a 'local' set of information the agent may use to decide. For instance, in the fiscal context, evasion might be more likely when an individual has been audited either during the previous round (bomb crater effect) or at the beginning of his fiscal life (echo effect). Human-based experiments show whether different experiences lead to states characterized by diverse levels of willingness to evade, and thus, more in general, whether subjects modify their behavior according to their current condition. Thanks to agent-based simulations, it is possible to systematically analyze and explain human behavior, in order to check the robustness and the external validity experimental phenomena on a larger population. For instance, the standard theoretical approach could be replaced by a setting closer to the Aspiration Adaptation Theory by Selten (1998): agents have a limited set of decision dimensions, and they can select even opposite actions, depending on their specific current state, which affects probability evaluation and weighting.

This micro approach is well exemplified by Garrido and Mittone (2008), who use the theory of finite automata (Rubinstein, 1986; Romera, 2000) to interpret Italian and Chilean experimental data on tax compliance. They report that the behavior of the great majority of subjects can be explained by either unconditional honesty or the bomb crater effect, which is part of the library of phenomena (Guala and Mittone, 2005; Mittone, 2006; Kastlunger et al., 2009).

Recalling the original notion of boundedly rational approach, Garrido and Mittone (2008) consider individuals as limited in their computational power: each taxpayer can rely on a restricted set of information, in order to decide whether to comply or fully evade. The two authors assume that the probability of evading depends on the current state of the taxpayer (referred to as 'locally determined decision maker'): this state may change according to external events, such as the occurrence of a fiscal audit. Every artificial agent consists in a finite state automaton (Moore, 1956; Sipser, 2006), whose binary stochastic output (compliance vs. evasion) does not depend only on the current state (for instance audited in the previous period), but also on the probability of evasion associated to that specific state. Garrido and Mittone (2008) collect human-subject data: experimental results show the bomb crater effect, which however turns out to be less evident at an aggregate level. For this reason, the experiment is followed by an agent-based simulation aimed to identify the specific micro determinants of taxpayers' behavior, i.e. to identify the automaton with the highest success ratio in predicting human subjects' decisions.

According to our view, this application of an agent-based model helps understand how simulations can support the experimental approach in the field of tax research. Human-based experiments provide more or less clear insights on taxpayers' behavior, and an agent-based system enriches our understanding of human behavioral regularities, by testing many cognitive drivers and inner motives which are supposed to be involved in the decision process. This synergic approach might resemble theory testing in the laboratory: as a matter of fact, like experiments help identify which theory best explains human behaviors observed in the laboratory, simulations allow the identification of the main cognitive drivers explaining human behavior and thus check its validity outside the laboratory.

Specifically, in their work, Garrido and Mittone (2008) propose a set of seven hypotheses,[11] which might explain experimental findings by testing the robustness of the bomb crater against the loss-repair effect.[12] Each hypothesis

[11] The decision of evading depends on whether the subject is (H_0) audited in the previous period; (H_1) audited in the previous period, and caught; (H_2) audited during the previous two periods; (H_3) audited in the previous two periods, and caught; (H_4) audited in the previous three periods; (H_5) audited in the previous four periods; (H_6) audited in the previous three periods, and caught.

[12] Both effects imply an immediate decrease in compliance after an audit. However, the former is due to chance misperception, and depends only on the occurrence of an audit,

is translated into an automaton, whose states map the characteristics of the hypothesis itself. This tests different behavioral motives, and helps identify the best one in explaining patterns observed in the laboratory. From this perspective, it is again evident how ACE simulations can provide a valuable support also at a micro level: they help increase the potential of experimental evidence and fully understand the psychological and cognitive drivers characterizing the individual decision process.

The hypothesis that gives the most detailed description and prediction of subjects' behavior is the one involving the bomb crater effect; the only other relevant automaton is the one describing unconditionally honest agents, i.e. those who fully comply, irrespective of their current state. These results confirm, on the one hand, the robustness of the bomb crater effect as a common behavioral trait, and, on the other hand, the existence of an honest type (Mittone and Patelli, 2000).

In addition, Garrido and Mittone (2008) test three further hypotheses, in order to control for the effect of different audit sequences.[13] In line with the expectations of robustness and external validity of the echo effect (Guala and Mittone, 2005), computational results confirm that human subjects' behavior can be explained by means of a rather simple hypothesis: repeatedly auditing subjects at the beginning of their fiscal life has a positive impact on compliance over a certain time period, because of a wrong probability evaluation people form when relying on sampled experience (Guala and Mittone, 2005; Mittone, 2006; Kastlunger et al., 2009).

Hence, two main behavioral patterns are identified: about 70.3% of the entire experimental pool consists of subjects who never evade and subjects who evade strategically according to the bomb crater effect. Honest subjects exhibit an evading probability close to 0, irrespective of their state; in contrast, in case of strategic evaders, the likelihood of evasion is low only in the 'not audited' and in the initial state. The remaining 29.7% does not exhibit a clear behavioral pattern, since, in every state, they evade with a probability close to 0.5. Therefore, the adoption of ACE modeling leads to conclude that even a simple behavioral hypothesis, which might be modeled as a heuristic,

and not on actual evasion detection. In contrast, the loss-repair effect emerges only when a taxpayer is found to be an evader.

[13]They assume that taxpayers' evasion might depend (H_7) positively on being audited in the previous period, and negatively on experiencing an audit in the first five periods of the experiment. This hypothesis is extended in H_8, which also considers the effect of being caught during the latest audit. Finally, they test whether (H_9) subject's decision depends on being audited during the previous two periods, and on experiencing an investigation during the first five periods of the experimental session.

can explain a large proportion of subjects' decisions. Nevertheless, such a comprehension of human behavior can be achieved thanks to agent-based investigations, as the mere observation of human subjects might not be sufficient to draw valid conclusions.

3.4.3 Micro-level Dynamics for Macro-level Interactions among Behavioral Types

This section deals with ACE models that combine micro behavioral aspects and macro dynamics, with the intent of providing a better understanding of both experimental evidence and economic phenomena taking place outside the laboratory. Therefore, this kind of comprehensive analysis can be of great relevance for policy implications, by relying on the implementation of human-calibrated agents.

The first example is the work by Garrido and Mittone (2013), which analyzes how the efficiency and the efficacy of an enforcement strategy - defined in terms of audit frequency and targeting - can be considered as a function of the population composition. However, different from the macro analysis by Mittone and Patelli (2000), behavioral types are defined according to income distribution and specific traits that characterize individual cognitive process (Garrido and Mittone, 2008). Taxpayers are endowed with a decision function, and, in each round, they choose whether to evade. Honest taxpayers tend to comply in any case, irrespective of their current state; strategic evaders behave according to the bomb crater effect. Right after all taxpayers make their decisions, the policy-maker applies an optimizing selection rule that targets a subset of agents to audit: on the one hand, collected tax increases revenues; on the other hand, audits are costly and not always successful.[14]

Garrido and Mittone (2013) conclude that the optimal audit scheme must take into account income distribution, the possibility of identifying behavioral patterns with micro foundations, and the specific fiscal history of individuals. Micro-level behavioral regularities emerging in laboratory experiments turn out to be fundamental in designing an auditing strategy: being aware of some cognitive biases can help predict people's behavior; agent-based simulations built on these biases are useful to plan a coherent and efficient fiscal policy. As income inequality increases, the optimal plan targets the richest taxpayers, and frequently repeats two consecutive audits as a strategy against the bomb crater effect. In contrast, as income distribution becomes more uniform, the optimal

[14]The degree of efficiency takes its maximum value 1 only when each audit catches an evader.

plan suggests to spread audits throughout the entire population. Every time an agent is investigated, his last four declarations are verified: if the tax authority audits each agents every four periods, also strategic evaders are caught.

Nevertheless, despite the relevant contribution of this study in understanding how the policy-maker can address the issue of tax evasion when realistically dealing with a heterogeneous population, results are partially due both to the rationality assumption on the tax authority and to the main characteristics of taxpayers' choice function (no intensive decision is allowed, and no actual learning is implemented). This simplifying decision process might lead to partially misleading behaviors: in fact, in their laboratory experiment, authors allow for intensive decisions, and observe that rich individuals often prefer to evade a small amount of tax with the aim of reducing the probability of being targeted. In contrast, the simulation by Garrido and Mittone (2013) disregards this important aspect, and the optimal plan might target the richest taxpayers so that expected revenues of the tax authority are maximized.

The second example of a computational analysis combining the macro and the micro approach is the one by Mittone and Jesi (2016). By extending the agent-based analysis by Garrido and Mittone (2013) and Mittone and Patelli (2000), they build a complex adaptive system in which a variety of behavioral types coexist. However, different from Mittone and Patelli (2000), these types are based on the definition of simple heuristics, and not of utility functions to optimize. In addition, with the intent of overcoming the limitations of the model by Garrido and Mittone (2013), they also allow for intensive decisions, learning, different risk perceptions, and for probability weighting as a common feature of individuals' decision process. Specifically, Mittone and Jesi (2016) investigate the functioning and the evolution of a system where boundedly rational agents cope with a public good which might be consumed and created by the agents themselves. The authors build a self-reproducing economy as a setting for the study of the emergence of a responsible behavior in managing a renewable resource. They study the necessity of an exogenous mechanism of auditing in order to achieve a sustainable set-up.

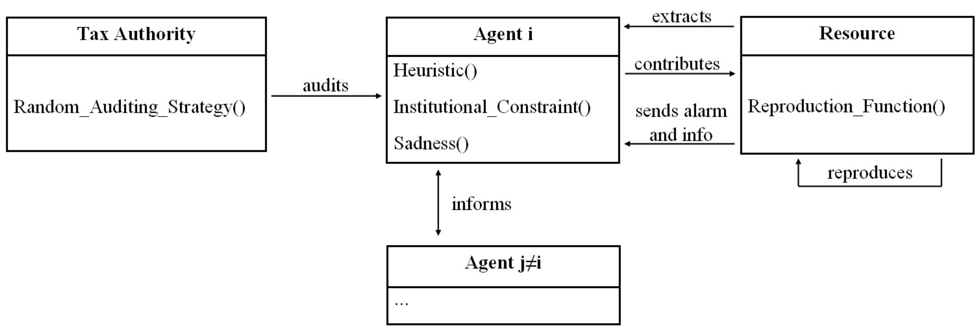
In every period, agents extract their private endowment from the good; then, they contribute by paying their tax due. These actions are carried out according to a limited set of heuristics (basically either imitative behaviors or habits) and to the employment type of the agent (employee vs. self-employed):[15] heuristics suggest the amount to take in order to perform a

[15] Agents can be categorized into two different cross-sectional sets according to institutional constraint introduced: agents subject to the high constraint (employees) cannot evade more

satisfying extraction, but the agent can try to extract more resources. In the beginning, each agent is randomly assigned a type and one of the five available heuristics, then, in order to keep the system dynamic, new agents are injected into the economy, and individuals can switch from one heuristic to another according to the achieved satisfaction level: the higher the level of sadness (i.e. the lower the degree of satisfaction), the higher the probability that an agent opts for switching to another heuristic. This sadness is mainly determined by the level of the extracted endowment, and the proportion of agents actually contributing to the public good. In addition, irrespective of the individual heuristic adopted, agents share the bomb crater effect as a common micro-founded psychological trait: as widely observed in human-based experiments, after an investigation occurs, the audited agent evades, underestimating the probability of a repeated audit. Such a characterization makes agents closer to human beings: they are not supposed to be rational, but rather emotional and biased in their decision process. For this reason, results can be of great interest and relevance for externally valid policy suggestions.

At the end of every period, the tax authority performs random audits and the good reproduces itself so that it cannot extinguish. Before the reproduction takes place, the good triggers a signal if the critical status in terms of quantity is reached. Agents react to this alarm according to their sensitivity level, i.e. to their propensity to take risk, and their adopted heuristic. See Figure 3.4 for a comprehensive representation of the overall system structure.

FIGURE 3.4: System structure diagram designed after Mittone and Jesi (2016)



Thanks to the implementation of this complex framework with a micro-founded behavioral heterogeneity, Mittone and Jesi (2016) identify the extent of behavioral heterogeneity that is able to trigger a responsible behavior, and

than 10% of the tax due, while those subject to the low constraint (self-employed) are free to evade up to 50% of the tax due.

thus find interesting results from a normative point of view. In fact, as already suggested by Hessing et al. (1992), they claim that selfishness, and thus evasion, can be effectively counterbalanced if other behavioral types are more attractive for taxpayers. In their model, an efficient fiscal policy can tremendously decrease tax evasion not only by means of audit deterrence, but also by sustaining the advantage that a taxpayer can get adopting an honest behavior. From this perspective, an efficient policy should exploit behavioral heterogeneity and induce taxpayers to the imitation of honest agents by making compliance more attractive for both employees and self-employed workers.

Hence, also in this case, the validity of the synergic approach of human-based and agent-based experiments is undeniable: human evidence serves as a basis to build behavioral types, and simulations allow the manipulation of population heterogeneity as a treatment variable, with the intent of leveraging the full potential and overcoming the limits of human-based experiments. This results in a more complete and deeper analysis of tax payment decisions: useful policy suggestions can be derived, as it is possible to implement a rather realistic system, in which different fiscal strategies are tested on a dynamic and heterogeneous population of interacting agents.

3.5 Conclusions

Since the appearance of the first theoretical models in the early 1970s, the study of tax compliance has moved a long way towards the development of new models, taking into account psychological regularities and anomalies of decision-making. The increasing success of the application of behavioral economics has shown the importance of relying on empirical and experimental data in order to integrate theoretical analyses, and overcome the traditional limit of representative agent. In fact, recent evidence from laboratory experiments and surveys underlines the impact of non-economic considerations in determining individuals' behavioral heterogeneity in real-world compliance, and the relevance of understanding taxpayers' behavior and the underlying cognitive process, in order to provide useful normative policies, able to sustain compliance and deter evasion.

From our perspective, much interest and effort need to be devoted to the combination of experimental techniques and agent-based models, in order to investigate the interaction between taxpayers' cognitive process and the surrounding environment. This would not only contribute to the external validity of experimental findings by testing the robustness of human subjects' behavior in systems with an increased degree of complexity, but it would also allow the integration of a micro-level perspective with macro-level considerations: dynamics at an aggregate level can be studied starting from micro-level observations.

This chapter explains how simulations can increase tax experiments' external validity in two different ways. On the one hand, agent-based models consist in the implementation of a set of human-based behavioral types and extend experimental analyses by manipulating the composition of agents' population. This provides a greater adherence to the environment outside the laboratory, and tests the effects of a variety of policies on a heterogeneous population. In fact, agent-based models may define different macro behavioral types interacting with diverse policy solutions adopted by the tax authority, and this heterogeneity can be based on the identification of micro-level behavioral dynamics emerging from psychology, economics laboratory experiments, and empirical studies. On the other hand, experiments' external validity can be increased by identifying the main cognitive drivers that explain phenomena observed in the laboratory. Human-based experiments contribute to the library of phenomena, by simply searching for facts and regularities, while agent-based simulations analyze and test these phenomena, so that they can be applied to specific cases to a normative purpose.

Overall, this kind of innovative approach adds to the ongoing discussion about the inclusion of behavioral realism into theoretical studies in the literature on tax evasion. Therefore, it supports a greater parallelism with the natural world, yet without denying the importance of model development: the synergic combination of theoretical analyses and human-calibrated simulations may help shed new light on the issue of tax evasion, since it focuses on the specific problem of new policy implementations in a rigorous way and in a realistic environment, before an actual application in the field.

Chapter 4

Commitment to Tax Compliance: Timing Effect on Willingness to Evade

joint with Luigi Mittone

4.1 Introduction

Many experimental studies on decision-making demonstrate the existence of a significant difference between planning and ongoing decisions. When asked to plan their actions, people often overweight events with small probabilities while in ongoing (i.e. real time and, in general, repeated) decisions, they tend to underweight these events and thus behave as if they ignored them (Yechiam, Barron, and Erev, 2005; Schurr, Rodensky, and Erev, 2014b; Camilleri and Newell, 2013). The so-called planning-ongoing gap, which can be considered a translation of the renowned description-experience gap (Barron and Erev, 2003; Hertwig et al., 2004; Rakow and Newell, 2010), emerges in many diverse situations e.g. drivers passing on a two-lane road (Harris, 1988) or very rare computer backups. Following this stream of research, Schurr, Rodensky, and Erev (2014a) and Erev et al. (2010) study problems related to not obeying safety rules in the workplace. They observe that workers are aware of the likelihood of rare events (i.e. accidents), and thus of the importance of rules but they actually tend not to obey them as if they underestimated the probability of accidents. Erev et al. (2010) test a gentle enforcement program aimed to help people behave according to safety rules. The program requires that workers remind each other to obey rules, especially when a violation occurs. Authors report a significant increase in compliance with safety rules thanks to this program.

We argue that violation of safety rules has a structural similarity to the problem of tax evasion: when taxpayers are asked to make a number of sequential decisions regarding their tax declaration, their behavior may change according to the Planning-Ongoing Gap. Not only their behavior, but also their attitudes towards risk of evasion may depend on the timing of decisions: as in the case of workers evaluating relevant risks in their workplace, the Planning-Ongoing Gap can be translated into a discrepancy between taxpayers' plans about income declaration and their actual behavior.[1] Many taxpayers, such as retailers or taxi drivers, deal with frequent tax payment decisions. Every time they sell a good or a service, they can decide whether to give proof of payment (either an invoice or a bill) to their customers and thus they automatically decide whether to declare the transaction or not. When they evade, they have to consider the risk of a fiscal investigation. Hence, according to the Planning-Ongoing Gap in the context of tax evasion, it is reasonable to expect that when asked to evaluate in advance the risk of evasion (i.e. at the beginning of the business year, for instance), taxpayers show a preference for compliance. Based on this assumption, our work investigates the existence of the Planning-Ongoing Gap in taxpayers' behavior in order to understand whether compliance systems can be introduced. A possible empirical example of this is the widespread adoption of POS devices for taxi fare payments in the US where taxicabs have a customer-friendly system that requires neither a signed receipt nor a minimum payment, and allows passengers to swipe their card and tip with reasonably high preset amounts. Thanks to compliance with credit card payments, the collection of revenue is higher: without POS, some customers would use taxicabs less frequently and tips would not be so generous. On the contrary, cash payments leave much more room for evasion but nevertheless do not 'ensure' frequent rides and tips. According to the data collected by the Taxi and Limousine Commission of New York, even if cab drivers may refuse to accept credit cards in favor of cash, a widespread use of POS is actually observed. Over time drivers seem to prefer to adopt this device since it allows them to earn more when complied with. Clearer conclusions about the functioning of such a system in tax-payment decisions and its relationship with the Planning-Ongoing Gap could be drawn by means of experimental implementation. In addition, such a research may provide normative recommendations for tax administration and regulation of ridesharing, as related to taxes and profitability for Uber's business model (Oei and Ring, 2016).

[1] Mittone (1997) provides evidence of taxpayers' underweighting in ongoing decisions.

We conduct our research in an experimental setting with repeated measurements by mimicking the taxicab scenario: subjects make a number of sequential decisions regarding their tax declaration. The application of similar settings (e.g., Allingham and Sandmo, 1972; Srinivasan, 1973a; Guala and Mittone, 2005; Mittone, 2006; Kastlunger et al., 2009) aims to replicate the decision dynamics during a taxpayer's life span. Differently, the sequential repetition of trials in the present study is intended to mimic the condition of those taxpayers (e.g. retailers or taxi drivers) that face these decisions every time they sell a good or a service during a fiscal year.

In order to examine the Planning-Ongoing Gap and to investigate the existence of a possible resolution, we experimentally test the availability of a compliance mechanism similar to that of the POS device used by taxi drivers. The inconsistency between planning and ongoing decisions is investigated by allowing taxpayers to commit to automatically pay half of the due tax (plastic payments) and then decide whether to pay the remaining part.[2] Thus, we not only compare Planning and Ongoing decision-making but also explore whether an enforcement system can be built on this gap in order to foster compliance in the long-term.

4.2 Methodology

We investigate people's behavior in a computerized tax-evasion experiment within a repeated choice framework. Such a dynamic setting was originally adopted to investigate the effect of audits on taxpayers' expectation: tax compliance depends not only on audit probability and punishment magnitude but also on time lag among audits (e.g., Spicer and Hero, 1985; Webley, 1987; Antonides and Robben, 1995; Mittone, 2006).

In the present research, we adopt the same setting but without dealing directly with the frequency and the pattern of audits during repeated trials. Nevertheless, this is useful for the investigation of the Planning-Ongoing Gap, given the focus on probability weighting in planning and purely experience-driven behavior in ongoing decisions.

4.2.1 Experimental Design and Task

At the beginning of every session we elicit subjects' risk preferences through the Bomb Risk Elicitation Task (Crosetto and Filippin, 2013)[3] where results

[2] A detailed explanation of the commitment mechanism can be found in Section 4.2.1.

[3] This method was selected because it allows us to measure risk preferences without explicitly asking participants to choose among lotteries, which is the task participants are given

are given to participants only at the end of the experiment.[4] In this task, every subject is shown100 boxes: he knows that 99 boxes contain 50 ECU each, while the remaining one contains a bomb. He is asked to collect as many boxes as he likes. Boxes are then opened: if the box with the bomb has not been selected, a subject's earnings depend on the number of collected boxes (50 ECU per box) but, if the bomb is among his boxes, this explodes causing null earnings (see Appendix B for details).

Sessions last for 60 rounds.[5] In each round the participant is informed of his exogenous income Y, and of the amount of tax he is asked to pay tY, where $t = 0.3$ is the tax rate. The sequence of incomes participants are assigned varies every ten rounds: the ten levels of income are $\{991, 1006, 989, 1005, 990, 1013\}$. These are determined as the algebraic sum of 1000 and a random error uniformly distributed between -15 and 15. This allows for a slight variation among the levels of income and tax which, however, does not determine an actual change of income stakes. The same sequence is kept stable across subjects since errors are not expected to have an impact on participants compliance decisions. This was simply intended to prevent subjects from getting escessively bored and distracted.

As shown in Table 4.1, we identify three between-subject treatments. In this setting, we investigate the effect of the availability of a compliance system based on the Planning-Ongoing Gap applied to the context of tax evasion. We examine whether a change in the timing of decisions and in the characteristics of the compliance enforcement system affects evasion.

TABLE 4.1: Treatments Table

Treatment	**Commitment Timing**	**Commitment Characteristics**
$(self)ControlTreatment$	10-round planning	Reduced evasion attractiveness
$PlanningTreatment$	10-round planning	Reduced tax liability Reduced evasion attractiveness
$OngoingTreatment$	round-by-round	Reduced tax liability Reduced evasion attractiveness

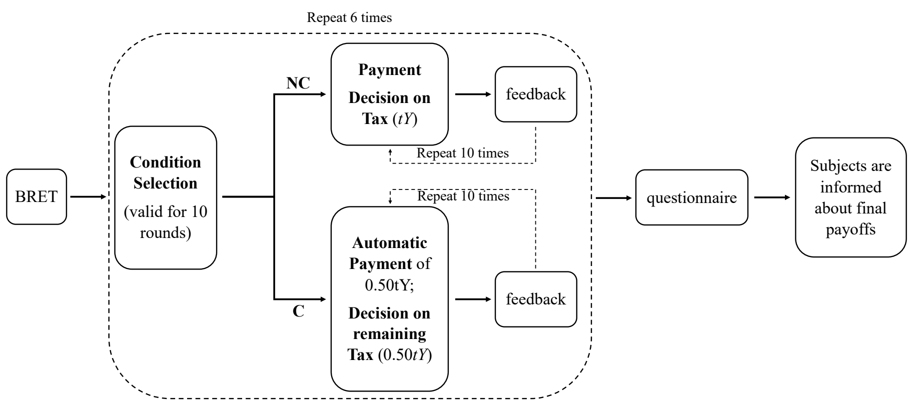
during the main part of the experiment. We adopted the structure of the static version and added a computerized visual support. Nevertheless, differently from Crosetto and Filippin (2013), subjects are not asked to state the number of boxes they desire to collect as they can actually choose and select the boxes they want.

[4]Subjects do not receive immediate feedback, so that their decisions about tax are not influenced by BRET outcome. For a similar reason, we did not administer this task at the end of the experiment as subjects decisions could have been affected by the feedbacks received during the 60 rounds.

[5]Even if no actual learning process is involved, we decided to set quite a high number of rounds to capture possible variations over time, such as the duration of commitment attractiveness.

In the *Control treatment*, the participant is asked every 10 rounds (i.e. in round 1, 11, 21, 31, 41, and 51) to select the condition under which he prefers to play: he can choose either the No Commitment (NC) or the Commitment (C) condition. The chosen condition is implemented for the following 10 rounds until the next selection. This means that if he opts for the NC condition for ten rounds, he can decide whether to declare his income. In the case of declaration, he simply pays tY and his round payoff is $Y(1-t)$. Where no declaration exists, he can be audited[6] and fined with probability $p = 0.20$. If he is not audited, his round payoff is equal to his entire income Y while, if he undergoes a fiscal investigation, his payoff is $Y - f$ where $f = 4tY$ is the fine.[7] If, instead, the participant selects the Commitment condition, he accepts to automatically pay half of his due tax at the beginning of each of the following 10 rounds. Based on this, he can then decide, round by round, whether to declare the rest of his income and pay the corresponding tax $0.50(tY)$. Thus as in the No Commitment condition, in the case of declaration, a participant's round payoff is $Y(1-t)$. If he does not pay the remaining half of his tax, he can be audited and his round payoff is equal to $Y(1-0.50t)$ with probability $1-p$ and to $Y-f$ with probability p. At the end of every round, subjects are informed of their payoff.

FIGURE 4.1: Decision structure (self) Control treatment



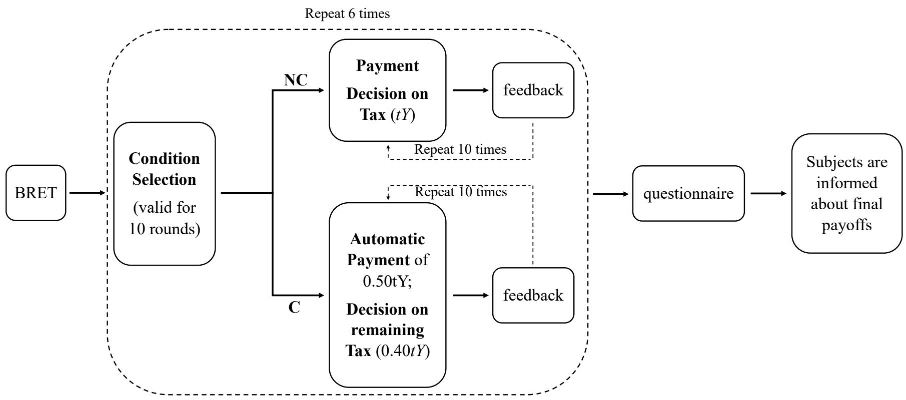
Similarly, in the *Planning treatment*, the participant chooses the condition (either No Commitment or Commitment) every ten rounds. However, while the No Commitment condition remains unchanged, the Commitment condition now includes a reduction of 10% on tax liability in return for advance tax

[6]According to the classification adopted by Hertwig et al. (2004) regarding probability distributions, evasion is defined as the option including the rare event (audit).

[7]The entity of the fine has been determined such that $Y - f < 0$. By introducing this mixed prospect, we do not focus only on either the (pure) gain or the (pure) loss domain.

payment.[8] If the participant opts for the automatic payment of half of his due tax ($0.50tY$) for the following ten periods, then in every round he decides whether to pay the discounted residual amount of $0.40tY$. As in the (self) Control treatment, the tax authority can audit the participant. Therefore, if he does not comply, his payoff is equal either to $Y(1-0.50t)$ with probability $1-p$ or to $Y-f$ with probability p.

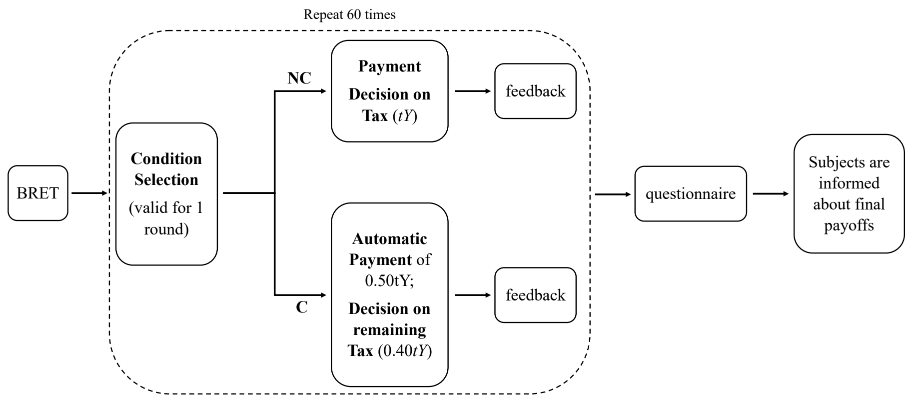
FIGURE 4.2: Decision structure Planning treatment



Finally, in the *Ongoing treatment*, the subject makes every round decisions regarding both tax payment and condition selection. At the beginning of each period, he first chooses a condition (again, between No Commitment and Commitment), which lasts only for that round; then, according to his selection, he decides whether to pay his due tax. The condition and the payoff structure are those adopted for the Planning treatment: the Commitment allows for a discount of 10% on tax liability in return for certain and automatic payment of half due tax (i.e. $0.50tY$). The only actual difference regards the timing and the duration of the condition choice: in the Ongoing treatment participants choose the condition every period while in the Planning treatment they choose the condition every ten rounds.

[8]Parameters are determined in order to avoid a near ceiling effect (evasion is too attractive to most people). We tested this in preliminary sessions by implementing different values for tax rate (40% vs. 30%) and discount rate (5% vs. 10%).

FIGURE 4.3: Decision structure Ongoing treatment



In order to reduce possible differences across treatments, subjects' decisions are divided in two sequential screens.[9] The first (i.e. the condition-selection screen) contains a brief description of the two conditions (these are also presented in detail in the instructions). In addition, subjects receive all the information they need to make their decisions: they are informed about the income level, the rate and amount of due tax under each condition, the fee, and the payoff they get in case of evasion and lack of audit (also with a distinction between the two conditions). The second screen (i.e. the one including the actual tax-payment decision) recalls the same information as the previous screen but with a focus on the chosen condition.[10] If the subject selects the NC condition, then he is shown a description (in terms of payoffs and probability distributions) of the two 'lotteries', i.e. the certain one of full tax compliance, and the risky one of full tax evasion. If the subject chooses the C condition, he is also informed of how his due tax has been split: the amount, and not only the proportion, that has been automatically paid, and the amount that is still to be paid. Regardless of the treatment, the subject is informed of his current payoff at the end of every round: he knows whether he has been audited and how much he has earned. The final payoff is determined as the sum of all the round payoffs.

[9]Screenshots can be found in Appendix B.

[10]Perhaps because of this information repetition, the experimental sessions took almost the same time independently of the treatment. It seems that in the Ongoing treatment subjects spend more time in front of the first screen than in front of the second; in the Planning treatment condition selection takes time, and then subjects spend some time to make their tax-payment decision (more time with respect to the Ongoing treatment).

4.3 Participants and Procedure

The experiment was conducted at the Cognitive and Experimental Economics Laboratory (CEEL) in the University of Trento. We conducted six sessions (two for each treatment). Participants were recruited among undergraduate students who enrolled for CEEL. The total number of recruited subjects was 100, and the average age was 21.77 ($s.d. = 2.457$): 98 students (47 females, and 51 males) took part in the experiment, while the other 2 were eventual replacements. Most of them (49%) were students of Economics, 24% of Law, 9% of Engineering, 7% of Humanities, 6% of Social Sciences, and 5% of Mathematics and Hard Sciences. None of the participants was informed in advance about the purpose of the experiment and subjects were allowed to participate only once.

In the laboratory, participants were randomly assigned to a computer; they received the general instructions[11] and those for the first part of the experiment, i.e. the risk preference elicitation task. Subjects had five minutes to read them individually. Then, instructions were read aloud by one member of the experimental staff and participants were asked to answer some comprehension questions. Once the first part of the experiment was completed, subjects received instructions for the second part. Again, they had five minutes to read them individually and then instructions were read aloud. After a brief comprehension test,[12] the second part of the experiment started. Before leaving, subjects were asked to fill out a brief demographic questionnaire.

The experiment was designed and administered using the Borland Delphi programming language. Each session lasted about 50 minutes. Subjects received 3 Euros as show-up fee, plus a sum of experimental currency units that varied according to their performance during the experiment and that was converted in Euros (1000 ECU = 0.25 Euros) at the end of the session. The average payment was equal to 14 Euros, with a maximum of 16 Euros and a minimum of 11 Euros, show-up fee included.

[11] A translated and complete version of the instructions is available in Appendix A.

[12] Comprehension of the instructions was controlled before the beginning of each of the two parts of the experimental sessions. The questions are reported at the end of the instructions in Appendix A where we tested the correct comprehension of the payoffs in the different strategies. Appendix B includes the original experimental screens: subjects are not asked to make computations when making their decisions since they receive all the relevant information they may need to compare the different options. As a result, we conclude that subjects behavior cannot be due to a wrong understanding of how the experiment works and payoffs are computed. Even if subjects in the Ongoing treatment can experience the option space more easily and rapidly, since no actual learning is required, it is possible to exclude that behavioral differences across treatments are only due to diverse learning possibilities.

4.4 Behavioral Predictions

According to the majority of experimental studies on tax evasion, the assumption of risk neutrality can be adopted for a comparison with behavioral predictions based on the Planning-Ongoing Gap. This experimental study is based on a dynamic setting where subjects make very similar decisions for a high number of rounds. Therefore, since the repeated decision problem does not change noticeably[13] during the experimental session, predictions can be based on a single round.

For the sake of simplicity, we start from the *Control treatment* where no actual difference exists between the two conditions in terms of compliance payoff. Regardless of the selected condition, the expected value is $EV(NC - Compliance) = EV(C - Compliance) = 700$. As for evasion, payoffs vary according to the condition: $EV(NCEvasion) = 760$ and $EV(CEvasion) = 640$.

In the *Ongoing treatment* the participant first selects the condition (No Commitment vs. Commitment). Under the No Commitment condition the subject may fully comply and get $EV(NC - Compliance) = 700$ or fully evade and get $EV(NC - Evasion) = 760$. Similarly, under the Commitment condition, he may comply and get $EV(C - Compliance) = 730$ or evade and get $EV(C - Evasion) = 640$. The highest expected value is associated with evasion under the No Commitment condition. Therefore, in each round, a risk-neutral utility maximizer would adopt this strategy.

Both in the Control and in the Planning treatment, risk-neutrality predicts evasion under the No Commitment condition in each round.

Even if the assumption of (Constant Absolute) Risk Aversion were taken into account, we would expect subjects to choose a mixture of evasion and compliance lotteries which does not depend on the treatment applied.

On the other hand, the Planning-Ongoing Gap predicts that many taxpayers, when asked to decide between immediate full compliance and immediate full evasion, underweight the probability of being audited, while they overweight this event when planning their future behavior. In the ongoing decision setting, we expect subjects to choose the condition under which evasion is more attractive (i.e. they do not adopt the POS device). When planning is involved, however, we expect them to focus on the benefits of paying tax and to choose the condition under which compliance is more profitable (i.e. they

[13] For the sake of simplicity, in the present numerical analysis an income of 1000 ECU is used. Predictions do not vary when considering the other income levels.

adopt the POS device that ensures them higher compliance revenues).[14] In addition to this, we also expect that fewer taxpayers select the commitment if this has the only purpose of making evasion less attractive (i.e. they do not opt for a POS device that does not give higher revenues). The Planning-Ongoing Gap predicts that subjects, when asked to evaluate their tax-payment problem in advance, focus on evasion drawbacks, overestimating audit occurrence. Our compliance mechanism is built on the fact that subjects are expected to choose the condition that allows them to earn more by paying tax; in contrast, they are expected to select less frequently the condition that simply helps them manage self-control problems by reducing evasion temptation.[15] Here the (self) Control treatment plays a fundamental role as it disentangles the effects of a commitment involving only advance tax-payment and the effects of a commitment which also offers tax reduction in return. Making evasion less attractive is sufficient to induce only sophisticated participants to adopt the commitment mechanism. Contrarily, the commitment system with a tax discount may involve managing self-control problems, but it does not actually require participants to be sophisticated and thus aware of possible decision inconsistencies. In our setting, the system relies on the fact that the majority of participants adopts it because they think that evasion is too risky and not necessarily because they know they will change their mind in ongoing decisions.

A discrepancy between the Planning and the Ongoing treatment would imply a difference in the evaluation based on decision timing. When subjects plan, they recognize the importance of compliance while this does not happen in ongoing decisions.

Consequently, we formulate the following research hypotheses:

[14]Erev and colleagues have recently conducted a similar investigation on the role of gentle rule of enforcement on driving speed. Here the Planning-Ongoing Gap shows that the majority of drivers that usually exhibit reckless behavior agree to install an IVDR ('In Vehicle Data Recorder' or 'Green Box') that punishes them when speed is excessive.

[15]This prediction is mainly based on results presented by Schurr, Rodensky, and Erev (2014b). In their experiments participants sequentially face 100 lotteries (this means the repetition of 100 trials): in the planning condition, they are asked to plan in advance how many of the 100 lotteries they would play for money; in the ongoing condition, at each trial participants are asked whether they want to play the current lottery for money or without money. Therefore, in this setting, planning affects all the trials and subjects have no possibility of (even just partially) changing their plans. To build a decision setting closer to reality, in our experiment planning simply reduces subjects' decision room for some periods: decisions taken in advance are expected to affect participants' behavior, but they do not prevent either tax-compliance or tax-evasion. After choosing the condition, subjects are still free to decide about their income declaration: indeed, it is difficult to find real situations in which taxpayers can commit in advance to tax-compliance for a certain number of periods Despite the difference between our experiment and that proposed by Schurr, Rodensky, and Erev (2014b), it is plausible to assume that the Planning-Ongoing Gap plays a role in affecting taxpayers' behavior (even in the case of a 'partial' commitment).

H1a - *Ceteris paribus, a higher proportion of taxpayers adopt the compliance mechanism in planning than under ongoing decisions.*

H1b - *Under planning, the reduction of evasion temptation is not sufficient to induce the majority of taxpayers to adopt the commitment.*

H2 - *Under planning, the adoption of the commitment mechanism helps taxpayers to stick with their compliance plans.*

H3 - *Overall, compliance is higher when people are asked to plan their future decisions and to adopt a long-lasting compliance system.*

Finally, we also expect that the widespread tax compliance due to the long-lasting mechanism leads to higher tax revenues (in spite of the discount on tax liability).

H4 - *Despite the higher commitment rate (hence the higher exploitation of the tax discount), a larger amount of tax is collected in the Planning than in the Ongoing treatment.*

4.5 Results

Section 4.5.1 introduces a statistical descriptive analysis of participants' decisions where our main hypotheses on the relevance of the Planning-Ongoing Gap in taxpayers' behavior are supported by our findings. Section 4.5.2 presents the application of a binomial regression model (with mixed effects taking into account the presence of repeated measurements).

4.5.1 Descriptive Statistics and Main Results

Subjects do not show relevant differences in terms of risk preferences across treatments. These are elicited by means of the Bomb Risk Elicitation Task (Crosetto and Filippin, 2013). We checked for the presence of outliers in our samples and observed that subjects collected on average 42 boxes (*s.d.* = 14.481) in the Planning treatment and 48 boxes in both the Ongoing treatment (*s.d.* = 11.459) and in the Control treatment (*s.d.* = 11.115). A slight difference in average risk preferences exists (as in standard deviations), but it does not seem plausible to assume that all our results are explained only by subjects' risk attitudes because we also observed a low correlation (0.30) between treatments and risk preferences. Following the classification by Crosetto and Filippin (2013), the cumulative distribution of choices shows 50% of risk averse subjects (less than 50 boxes collected), 15% of risk neutral subjects (50 boxes collected), and 35% of risk seekers (more than 50 boxes collected).[16]

From a first comparison between the Planning and the Ongoing treatment, a difference in subjects' commitment rate emerges: participants opt for the Commitment condition more frequently in the former. The average commitment rate[17] is equal to 0.56 ($s.d. = 0.317$) in the Planning treatment and to 0.40 ($s.d. = 0.335$) in the Ongoing treatment. This difference is statistically significant across subjects (WRT: $p - value = 0.029$).[18]

Figure 4.4 shows the average commitment rate and displays a clear, immediate comparison between the two treatments. In the first 50 periods the difference is highly significant, while in the last 10 periods we observe a decrease in the distance. However, this fact may be explained mainly by a decline in the commitment rate in the Planning treatment rather than an increase in the Ongoing treatment, which tends to be quite steady over the entire experimental sessions.

Since this is not the specific purpose of the present study, future research can address this interesting issue and investigate how to make the attractiveness of the commitment more stable over time and avoid subjects' boredome. Similar evidence did emerge in some pilot experimental sessions we ran over 100 (instead of 60) rounds where we observed a similar decrease but at a later stage. For this reason, and thanks to subjects' instruction comprehension checks (see Footnote 12), we are able to exclude slow exploration and learning as determinants of the initial high commitment rate in the Planning treatment.

Result 1a - *Taxpayers who are offered the option of a system (with tax discount) having effect just on their immediate compliance decision are less willing to adopt this system compared to those who are offered the same option but with a longer effect.*

Figure 4.4 also allows for comparison between the Planning and the Control treatment where the average commitment rate is 0.16 ($s.d. = 0.151$). As expected, participants adopt the system more frequently in the Planning treatment (WRT: $p - value < 0.001$): a commitment that involves only an advance tax payment is not as attractive as one also offering a discount on tax liability

[16] Results are in line with Crosetto and Filippin (2013): their cumulative distribution of choices reports 52.1% of risk averse subjects, 14.7% risk neutral subjects and 33.2% risk seekers.

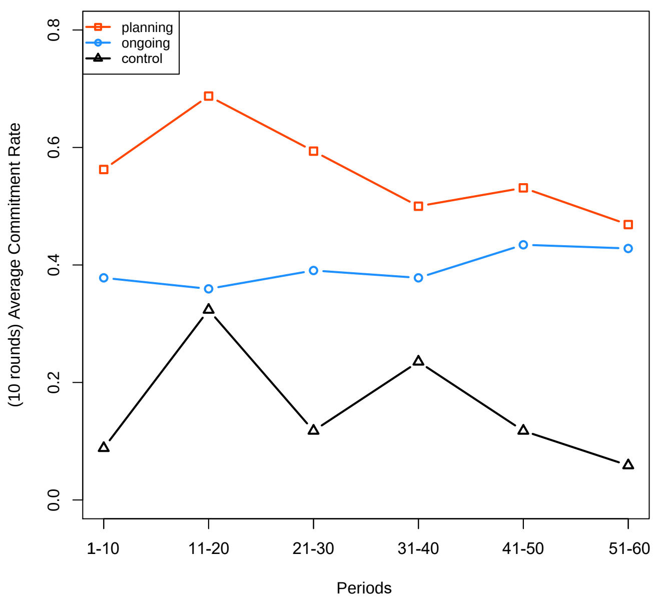
[17] The average commitment rate is computed as the mean of the individual average rates.

[18] All tests are two-sided, if not specified. WRT stands for Wilcoxon Rank Sum Test.

even if the decrease in evasion attractiveness is the same. Given the specific structure of the commitment of the Planning treatment, it seems possible to conclude that taxpayers who are asked to evaluate compliance decisions in advance are less willing to evade. It follows that they recognize the importance of compliance and tend to choose the condition under which the safe option is more profitable. On the contrary, the commitment is not widely adopted if it does not produce any difference other than a reduction in evasion temptation.

Result 1b - *The number of taxpayers opting for the commitment mechanism is systematically higher when a discount on tax liability is involved. Only a few of them are explicitly driven by the need to manage self-control problems.*

FIGURE 4.4: Average Commitment Rate over Time



Apart from commitment adoption, what is relevant for the investigation of taxpayers' behavior is how actual compliance is affected by planning. Figure 4.5 shows how compliance rate varies according to the selected condition: by analyzing individual average behavior over the 60 rounds in the Planning treatment, taxpayers choosing the Commitment (C) condition systematically comply more (WRT: $p - value < 0.001$) than those choosing the No Commitment (NC) condition.

For the sake of completeness, Table 4.2 and Figure 4.5 report all treatments. However, while the analysis of the Planning and the Control treatment

Table 4.2: Compliance Relative Frequency

		Mean	Median	St. Dev.
(self) Control Treatment	No commitment	0.402	0.4	0.075
	Commitment	0.665	0.667	0.245
Planning Treatment	No commitment	0.383	0.360	0.142
	Commitment	0.757	0.740	0.102
Ongoing Treatment	No commitment	0.215	0.216	0.081
	Commitment	0.949	1	0.062

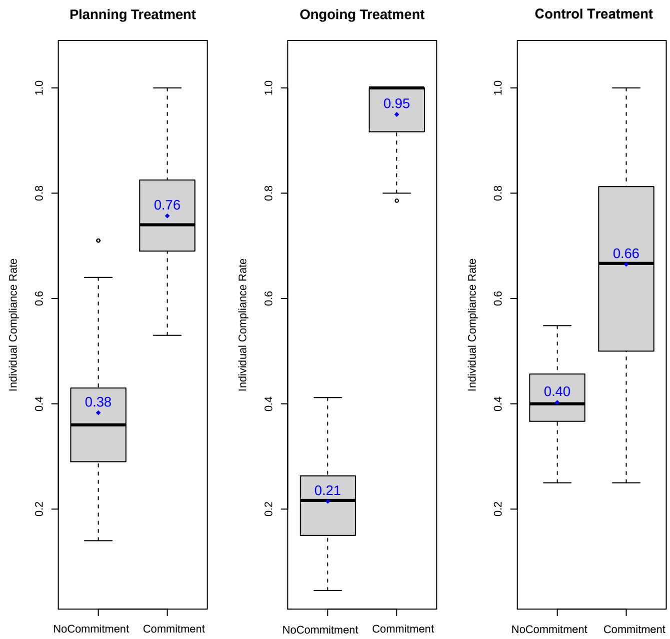
is useful to understand whether the adoption of the commitment is actually correlated with a higher compliance rate, the same investigation for the Ongoing treatment is relevant only to verify subjects' comprehension of the tax-payment mechanism. In the Ongoing treatment, it is reasonable to assume that the taxpayers' commitment adoption strictly depends on their immediate intention to pay taxes in that specific period. Indeed, no long-term evaluation is required. Thus, the proportion of people complying under Condition C is higher in the Ongoing than in the Planning treatment. But this fact needs to be interpreted together with evidence on the commitment rate where the relative frequency of commitment selection in the Ongoing treatment is systematically lower than in the Planning treatment.

Table 4.2 also shows a significant difference in individual compliance rates across conditions in the Control treatment (WRT: $p - value < 0.001$). This may indicate that some sophisticated taxpayers use the commitment to reduce evasion temptation but the compliance rate under the commitment is higher in the Planning treatment. Nevertheless, as depicted in Figure 4.4, the relative frequency of commitment adoption is very low (or even negligible in the end).

Result 2 - *Taxpayers who adopt the long-lasting commitment actually comply more than those who do not adopt it.*

This can be interpreted as a confirmation of Result 1a: subjects who seem to be willing to pay tax when evaluating and planning their decisions in advance actually comply. Although the commitment requires the automatic payment of half tax and leaves room for *ongoing* evasion, the great majority of taxpayers stuck with their compliance plan even if they could be tempted to evade.

FIGURE 4.5: Compliance Rate across Conditions



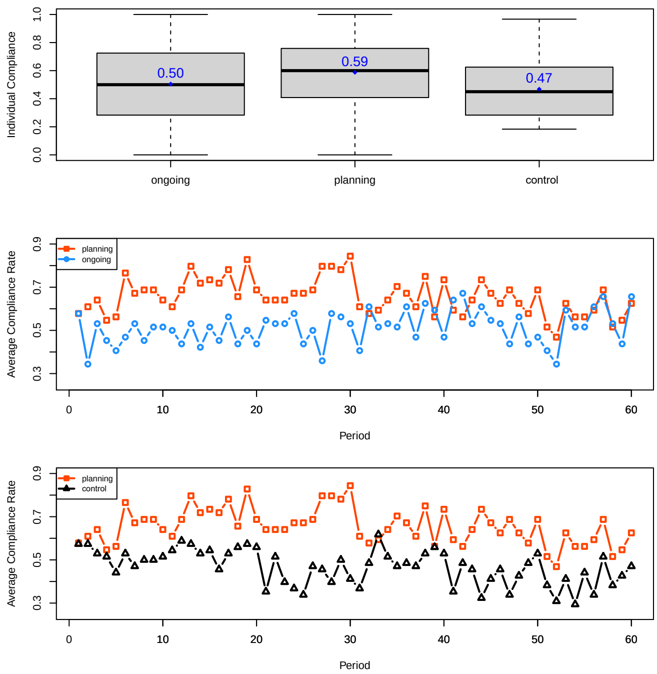
Considering what has been observed so far, it is reasonable to assume that taxpayers comply more on average in the Planning than in the Ongoing treatment. A direct comparison of compliance rates across treatments, with no distinction regarding the condition, can help verify this. Figure 4.6 shows that, on average, compliance is equal to 0.588 ($s.d. = 0.247$) in the Planning treatment and to 0.50 ($s.d. = 0.248$) in the Ongoing treatment. As expected, individual average compliance rates are higher in the Planning treatment (one-tailed WRT: $p - value < 0.1$).

Result 3 - *Regardless of the commitment adoption, the compliance rate is higher when taxpayers are asked to plan their future behavior and offered a long-lasting commitment mechanism.*

Despite the decrease in commitment adoption observed at the end of the Planning treatment, the difference between these two treatments is significant with respect to both condition selection and actual compliance. Such a distinction is necessary since in the present experiment subjects can pay tax in both conditions: on the one hand, the commitment adoption simply represents a choice room reduction (people can evade less but they are still free to evade); on the other hand, not opting for the commitment does not prevent from compliance.

Similarly, when considering the Control treatment ($mean = 0.441$; $s.d. = 0.223$), compliance in the Planning treatment is significantly higher (WRT: $p - value < 0.01$). This is in line with the low commitment rate observed in the Control treatment.

FIGURE 4.6: Average Compliance across treatments

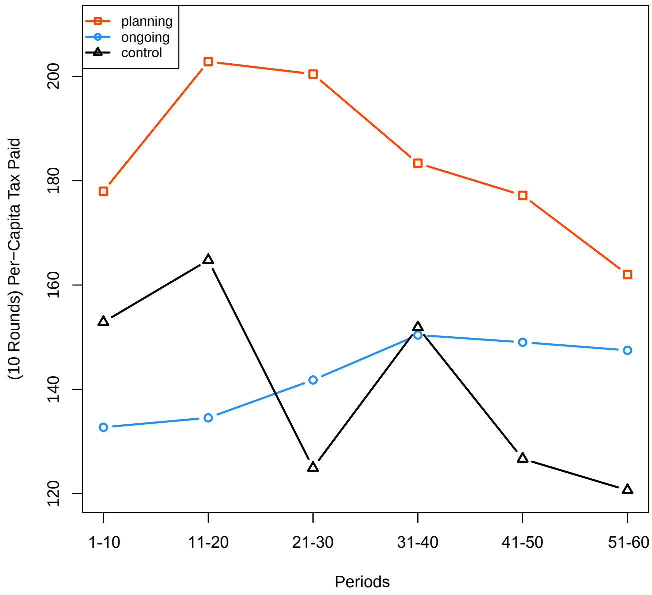


Result 3 does not necessarily imply that the amount of collected tax is higher in the Planning than in the other two treatments. The analysis of the compliance rate is not sufficient to determine which treatment is characterized by the highest amount of paid tax. The majority of subjects who comply in the Planning treatment adopt the Commitment, which has the positive effect of fostering compliance, but also the negative effect (from the point of view of the government) of offering a discount to taxpayers. Figure 4.7 shows that in Planning much more tax has been systematically collected (WRT: $p - value < 0.001$):[19] the average of (per-capita) collected tax is 163.36 in the Planning treatment but 139.635 in the Ongoing treatment. This proves how important the automatic payment of the commitment mechanism is: although the difference in individual average compliance across treatments seems to be quite small (Figure 4.5), the amount of collected tax is higher in the Planning treatment. However, a decline in tax compliance in the Planning treatment is also seen here.

[19]We performed the test on the average levels of tax paid over the session. We found a significant difference both between Planning and Ongoing and between Planning and Control.

Finally, the average is equal to 140.305 in the Control treatment where the introduction of the discount on tax liability increases tax revenues.

FIGURE 4.7: Average (per capita) collected tax over time



Result 4 - *Not only do taxpayers comply more in the Planning than in the Ongoing treatment, and they also pay a higher amount of tax.*

In spite of the widespread exploitation of the discount in the Planning treatment, we observe an overall increase in collected tax: the discount is more than compensated by the high compliance in the Planning treatment.

4.5.2 Regression Analysis

Table 4.3 describes determinants of taxpayers' behavior by drawing a comparison between the Planning and the Ongoing treatment while Table 4.4 focuses on the Control and the Planning treatment. A Logit Model is adopted because the dependent variable is compliance (1 = full compliance; 0 = full evasion) and we introduce Mixed Effects to check for repeated decisions.

Model (1a) analyzes the main treatment effects on taxpayers' behavior. That subjects comply more in the Planning than in the Ongoing treatment and that commitment and compliance are positively correlated is confirmed. Then, in line with previous results, the Commitment-Planning interaction term has a negative impact on compliance: the very few subjects committing in the Ongoing treatment almost always comply.

Table 4.3: Regression analysis - Planning vs. Ongoing.

	Compliance		
	(1a)	(1b)	(1c)
Commitment	4.869 (0.244)***	4.907 (0.254)***	4.917 (0.254)***
PlanningTreatment	1.479 (0.367)***	1.478 (0.377)***	1.360 (0.376)***
BRET	–	–	−0.016 (0.010)
Male	–	–	0.283 (0.385)
Economics	–	–	−0.704 (0.414)*
$Audit_{-1}$	–	−0.699 (0.128)***	−0.698 (0.128)***
$Audit_{-2}$	–	−0.689 (0.127)***	−0.687 (0.127)***
Commitment&Planning	−3.725 (0.277)***	−3.701 (0.284)***	−3.726 (0.288)***
$Audit_{-1}\&Audit_{-2}$	–	1.302 (0.284)***	1.302 (0.284)***
Constant	−1.591 (0.262)***	−1.375 (0.272)***	−0.454 (0.525)
Log Likelihood	−1677.6	−1584.0	−1579.7
Observations	3840	3712	3712
Num. groups: ID	64	64	64

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

As for the comparison between the (self) Control and the Planning treatment (see Table 4.4), *Model (2a)* analyzes the role of the discount: when the adoption of the commitment makes compliance more attractive (and not only evasion less profitable), the probability of paying tax is significantly higher.

Table 4.4: Regression analysis - Planning vs. (self) Control.

	Compliance		
	(2a)	(2b)	(2c)
Discount	0.775 (0.323)**	0.825 (0.342)**	0.592 (0.081)*
BRET	–	–	−0.019 (0.009)**
Male	–	–	0.193 (0.339)
Economics	–	–	−0.408 (0.344)
$Audit_{-1}$	–	−1.154 (0.110)***	−1.154 (0.110)***
$Audit_{-2}$	–	−0.726 (0.106)***	−0.726 (0.106)***
$Audit_{-1}\&Audit_{-2}$	–	1.066 (0.235)***	1.069 (0.235)***
Constant	−0.251 (0.224)	0.038 (0.239)	1.156 (0.557)
Log Likelihood	−2333.4	−2175.5	−2172.1
Observations	3960	3828	3828
Num. groups: ID	66	66	66

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Model (1b) and *Model (2b)* introduce fiscal investigations as possible determinants of compliance. According to the literature of tax experiments with a repeated measure design, an increase in tax evasion is expected immediately after an audit.[20] This is the so-called *bomb crater effect* (Mittone, 2006; Kastlunger et al., 2009; DeBacker et al., 2015; Mittone, Panebianco, and Santoro, 2016) and it is confirmed by our results: the probability of observing compliance decreases if a subject has been audited in the previous period (i.e. if the dummy variable $Audit_{t-1}$ is equal to 1). In addition, these models also

[20]In the Ongoing treatment, audits are expected to produce a similar effect also on the commitment rate: as shown in Table 4.2, almost any subject willing to evade decides not to adopt the commitment.

report how compliance in period t is affected not only by an audit in period t-1, but also in t-2. Results confirm an increase in tax evasion if one audit is recent: the coefficient of both $Audit_{t-1}$ and $Audit_{t-2}$ are significantly negative.[21] However, the interaction term has a positive impact on compliance: this implies that if subjects are audited both in period t-1 and in period t-2, then the probability compliance is not necessarily reduced as in the case of a single investigation, either in period t-1 or in period t-2. Therefore, subjects who are audited only once during the previous two rounds tend to underestimate the probability of observing an immediate new audit whereas those who have experienced two sequential investigations recognize the likelihood of repeated audits.[22] This result (as well as the bomb crater effect) is in line with the 'Black Swan Effect' suggested by Taleb (2007). Subjects do not perceive each round as completely independent from the others and for this reason they try to infer the likelihood of audit sequences by relying on limited and recent information. When a sequence of small feedback-based decisions is involved, and subjects perceive audits as non-frequent (or even rare), then the average tax-evasion rate is quite high especially immediately after a fiscal investigation. Subjects underweight the probability of experiencing many sequential audits. However, when the audit is not perceived as rare (e.g. because it has been experienced in the previous two rounds), then average tax evasion is reduced with respect to the other case. During the experimental session, subjects make decisions by experiencing (and thus by learning) patterns of outcomes. This fact supports our hypothesis that participants decisions may be affected by probability weighting.

Finally, *Model (1c)* and *Model (2c)* add control variables: gender is never significant; the dummy variable representing the enrollment in Economics courses shows that these students tend to comply less but only in *Model (1c)*. For risk preferences,[23] risk seeking is negatively correlated with compliance in *Model (2c)*.

4.6 Conclusions

We experimentally explore how taxpayers' decisions are affected by evaluating decision problems in advance and by planning actions. Results confirm the

[21] Although the regression without the interaction term $Audit_{t-1}\&Audit_{t-2}$ is not reported for the sake of exposition, it confirms the bomb-crater effect.

[22] With this respect, it might be useful to recall the echo effect, identified by Guala and Mittone (2005), and studied (in terms of the effect produced by different audit schemes) by Kastlunger et al. (2009).

[23] The variable BRET is determined as the number of boxes a subject collected: the higher the value of such variable, the higher the degree of risk-seeking.

existence of a Planning-Ongoing Gap in taxpayers' behavior and the need of a compliance mechanism: ex-ante evaluation induces the majority of people to adopt a long-lasting commitment (when compliance is more profitable) which, in turn, fosters tax compliance. In the case of ongoing decision-making, however, we find a less frequent adoption of the (short-term) commitment and a significantly lower rate of compliance which suggests that taxpayers perceived evasion as less risky.

Taxpayers (in their ex-ante evaluation) show preferences for compliance perceiving evasion as too risky but not exclusively as they expect possible inconsistencies between their planned and ongoing decisions (sophistication). Unsurprisingly the majority of subjects choose the commitment because it ensures a more profitable compliance. Only a few of them seem to predict that they will be tempted to evade in their ongoing decisions. When planning, they are frightened by evasion and thus opt for the condition that ensures the most beneficial compliance. Only a small proportion of our pool chose the commitment to reduce temptation. In addition, the compliance rate under commitment is lower in the Control treatment (where sophisticated subjects are supposed to commit) than in the Planning treatment.

On the basis of this experimental and empirical analysis, it would be interesting to investigate both the effect of a different timing on decisions and the effect of diverse incentives for the adoption of the commitment mechanism. In our experiment, the implemented incentive is the introduction of a discount on due tax; nevertheless, also the implementation of a reduction in the audit probability in return for certain declaration could be investigated. Our results could provide the basis for field explorations as investigations on the Planning-Ongoing Gap could shed new light on and help reshape current fiscal policies.

4.7 Appendix A - Experimental instructions

[This is a translated version (originally in Italian) of the instructions used for the experimental sessions. Instructions change according to the treatment. This will be indicated in the text.

As indicated in the general instructions, both the first and the second part of the experiment start only once every subject has correctly answered a brief comprehension test. We report the questions for each part at the end of the instructions.]

GENERAL INSTRUCTIONS

Welcome.

Thank you for coming. You are going to take part in an experiment on economic decisions. For arriving on time, you will receive 3 Euros at the end of the experiment.

Now you will be given instructions for the experiment. Please read them carefully. If you have any doubts, raise your hand and a member of the experimental staff will come and answer your question.

During the experiment, you are not allowed to talk to the other participants. If you disturb your colleagues or use the computer for activities not strictly related to the experiment, you will be excluded from the experiment and any reward. You can trust that what happens during the experiment is in line with the following instructions.

All participants will have the same role and they will not interact with each other.

The experiment consists of two parts.

You will be given two minutes to read the instructions about the first part. They will be then read aloud by a staff member; you will be asked to answer a few simple questions to check the instructions have been understood.

The first part of the experiment starts now.

At the end of this part, you will have five minutes to read the instructions on the second part of the experiment. Once again, the instructions will be read aloud and you will be asked to answer few simple questions on comprehension.

Then, the second part starts.

At the end, a brief questionnaire is administered and you will be informed of your reward. During the experiment, ECU (Experimental Currency Units) will be used as your earnings. At the end of the experimental session, the

ECUs you have earned are converted into Euros to determine your real payoff (1000 ECU = 0.25 Euros).

INSTRUCTIONS: FIRST PART

100 boxes appear on your screen. In one of them, there is a bomb; each of the other 99 boxes contains 50 ECU. You do not know where the bomb is but you know that it might be in any of the 100 boxes with the same probability.

For this part of the experiment, you are asked to select all the boxes you like. You will earn 50 ECU for each box (without the bomb) you collect. To select a box, you can simply click on it. Selection does not imply immediate opening: you will discover the actual content (ECU or bomb) of your boxes only at the end of the experiment (that is, after the second part). If you select the box containing the bomb, then everything you have collected is destroyed and you will earn 0 ECU for this first part of the experiment.

After collecting all the boxes you like, select the STOP button. This completes the first part of the experiment.

[Here we include the comprehension test administered before the first experimental part starts. The order is randomized.]

1. How many ECU are there inside a box (without bomb)?
   - 20 ECU
   - 10 ECU
   - 50 ECU
2. How many boxes have a bomb inside?
   - Every box contains a bomb
   - No box contains a bomb
   - Only one box contains a bomb
3. Imagine you have collected a certain number of boxes. At the end of the experiment, you discover that one of these boxes contains the bomb. How much do you earn?
   - It depends on the number of collected boxes (50 ECU per box)
   - 0 ECU
4. Imagine you have collected a certain number of boxes. At the end of the experiment, you discover that none of these boxes contains the bomb. How much do you earn?

- It depends on the number of collected boxes (50 ECU per box)
- 0 ECU

INSTRUCTIONS: SECOND PART

`Common to all treatments`

This part of the experiment will last for 60 periods. You are asked to make sequential decisions and after each period you will be given feedback. In every period, you can earn a certain amount of ECU. Your per-period earnings do not depend on the ECU you get in other periods and your final payoff is determined as the sum of all the ECU you have earned during the first and the second part of the experiment. You will be informed of such a payoff at the end of the experimental session.

Imagine you are a self-employed worker. For 60 consecutive periods you will be asked to declare your income (for taxation purpose). At the beginning of every period, you will be informed about the amount of your income and the corresponding amount of tax. The tax rate is equal to 30% and it does not change during the experimental session but your income will vary slightly every 10 rounds. In every period, you will be asked to make a decision regarding your declaration. You can decide whether to declare your income (i.e. whether to pay your tax) or not.

`Control and Planning Treatment`

Every 10 periods (i.e. in periods 1, 11, 21, 31, 41, and 51) you will have the opportunity to select the condition under which you will subsequently decide about your income declaration in the following 10 periods. Therefore, the condition you select will last for 10 periods: you will first choose the condition (at the beginning of a series of 10 periods) and then, in each period, you will have to decide whether to declare all your income or none of it.

`Ongoing Treatment`

In each period, you will be asked to make two decisions. At the beginning of every period, you will have the opportunity to select (first decision) the condition under which you will subsequently decide about your income declaration (second decision) in that period.

`Common to all treatments`

Here is a detailed description of the two conditions. This will be help you to select the one you prefer.

Condition N - If you choose this condition, then you are free to decide whether to declare your income. However, remember that in every period you can be audited by the Tax Authority with a probability of 20%. If you are found guilty of no declaration, you will be fined. If you declare your income, your final amount of ECU for that period is determined as your after-tax income. However, if you do not declare your income, the final amount of ECU for that period will depend not only on your decision but also on the possibility of being audited and fined. Hence, your round payoff will either be equal to your income in case of no audit or equal to your income minus the fine.

Condition C - If you choose this condition, then you are free to decide whether to declare your income. However, if you choose this condition, half of your tax is automatically paid at the beginning of the period before you can decide about your income declaration. In every round, you can then only decide whether to pay the remaining part of your tax.

`Control Treatment`

In this condition, you can also be audited and fined with a probability of 20%. If you declare your income, your final amount of ECU for that period is determined as your after-tax income.

`Planning and Ongoing Treatment`

Thanks to this automatic payment, the Tax Authority offers you a discount on the remaining half of tax. Such a reduction is computed as 10% of the total amount of tax due. In this condition, you can also be audited and fined with a probability of 20%. If you declare your income, your final amount of ECU for that period is determined as your after (discounted) tax income.

`Common to all treatments`

If you do not declare your income and you are not audited, the final amount of ECU will be equal to your income minus the part of tax that has been automatically paid at the beginning of the period. Otherwise, if you are audited, your payoff will be equal to your income minus the fine (paid tax is deducted from the fine).

`Control and Planning Treatments`

To summarize: every 10 periods you are asked to select the condition you prefer for the following 10 rounds. Then in each of the 10 periods, you will have to decide whether to declare your income under the condition you have previously selected. From the beginning condition-selection round, you will be given all the information you might need to make your choice.

At the end of every period, you will receive feedback.

`Ongoing Treatment`

To summarize: in every period you will be asked first to select the condition you prefer and then to decide whether to declare your income under the condition you have previously selected for that period. From the beginning of each round (i.e. in the condition-selection screen), you will be given all the information you might need to make your choice.

At the end of every period, you will receive feedback.

[Here we include the comprehension test administered before the first experimental part starts. The order is randomized. Answers vary according to the treatment. Subjects are given questions 5-7 and either question 8 or question 9, and either question 10 or question 11 (in a randomized way).]

1. How many times do you choose between Condition N and Condition C?
    - At the beginning of every period
    - Only once, at the beginning of this second part of the experiment
    - 6 times
2. What is the frequency of income change?
    - Income change occurs every period

- Income change occurs every ten periods

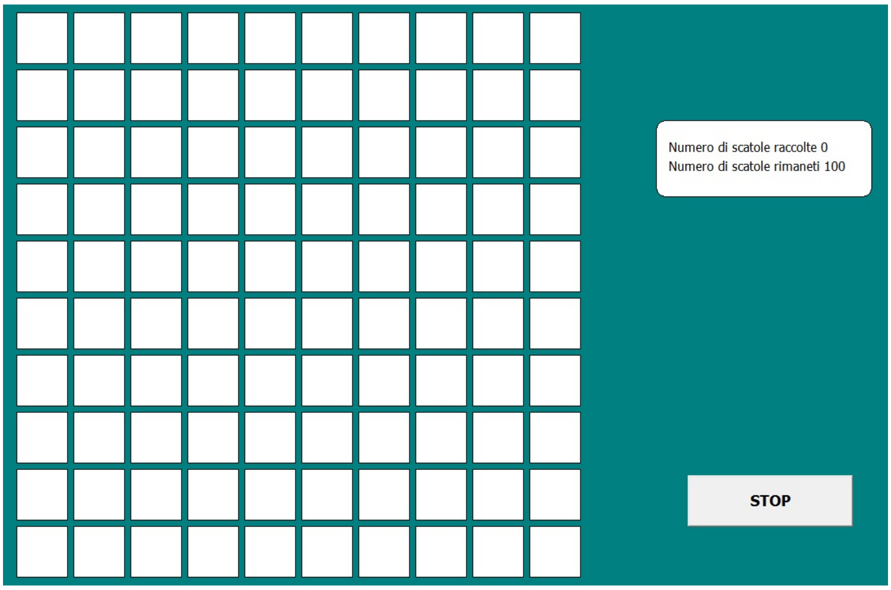
3. Under both conditions, what is the probability of being investigated by the fiscal authority in a given period?
    - It depends on what has happened in the previous periods
    - 20%
4. Imagine your income is 2,000 and you choose Condition C (50% of due tax is paid). How many ECU do you get if then you also declare the rest?
    - 1,400 ECU
    - 1,460 ECU
5. Imagine your income is 2,000 and you choose Condition C (50% of due tax is paid). How many ECU do you get if you then do not declare the rest and you are not audited?
    - 1,700 ECU
    - 2,000 ECU
6. Imagine your income is 2,000 and you choose Condition N (no tax paid in advance). How many ECU do you get if then you declare your income?
    - 1,400 ECU
    - 1,460 ECU
7. Imagine your income is 2,000 and you choose Condition N (no tax paid in advance). How many ECU do you get if you then do not declare the rest and you are not audited?
    - 1,700 ECU
    - 2,000 ECU

4.8 Appendix B - Breakdown of the experiment

FIRST PART - RISK PREFERENCES ELICITATION

For the BRET task, participants are shown 100 boxes and they are told that each box (apart from the one with the bomb) contains 50 UMS. They can collect all the boxes they like (without immediately discovering what is inside each box) and then move on to the second part of the experiment. At the end, they are told whether the bomb is among the boxes they have collected: if so, they get no reward for the BRET because the bomb destroys everything. If the bomb is not present, they are given an amount of UMS, which depends on the number of boxes they have previously collected.

FIGURE 4.8: Bomb Risk Elicitation Task - Screenshot

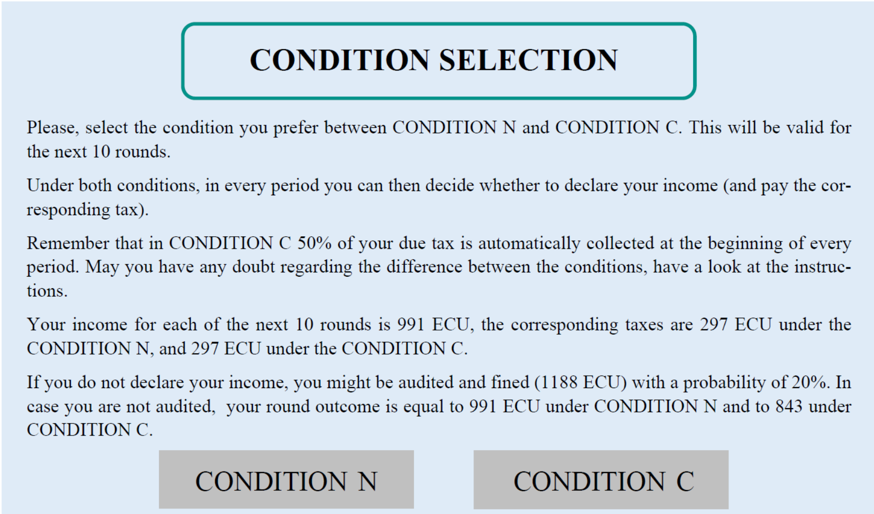


SECOND PART - CONDITION SELECTION

In the following screens, subjects receive any relevant information they might need to make a choice. For each of the two conditions, they know the amount of due tax (discount included for Condition C if applicable), the fee, and final earnings in the case of evasion and no audit. No complex computations are required as subjects can draw a direct comparison between the two conditions.

At the beginning of rounds 1,11,21,31,41, and 51 in the Control treatment, the following condition-selection screen is displayed.

Figure 4.9: Condition Selection Screen - Control Treatment


CONDITION SELECTION

Please, select the condition you prefer between CONDITION N and CONDITION C. This will be valid for the next 10 rounds.

Under both conditions, in every period you can then decide whether to declare your income (and pay the corresponding tax).

Remember that in CONDITION C 50% of your due tax is automatically collected at the beginning of every period. May you have any doubt regarding the difference between the conditions, have a look at the instructions.

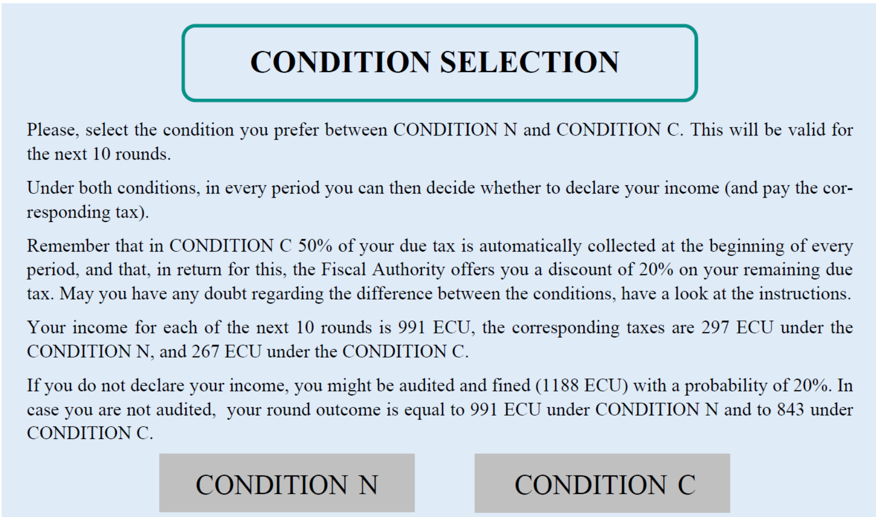
Your income for each of the next 10 rounds is 991 ECU, the corresponding taxes are 297 ECU under the CONDITION N, and 297 ECU under the CONDITION C.

If you do not declare your income, you might be audited and fined (1188 ECU) with a probability of 20%. In case you are not audited, your round outcome is equal to 991 ECU under CONDITION N and to 843 under CONDITION C.

CONDITION N CONDITION C

At the beginning of rounds 1,11,21,31,41, and 51 in the Planning treatment, the following condition-selection screen is displayed.

Figure 4.10: Condition Selection Screen - Planning Treatment


CONDITION SELECTION

Please, select the condition you prefer between CONDITION N and CONDITION C. This will be valid for the next 10 rounds.

Under both conditions, in every period you can then decide whether to declare your income (and pay the corresponding tax).

Remember that in CONDITION C 50% of your due tax is automatically collected at the beginning of every period, and that, in return for this, the Fiscal Authority offers you a discount of 20% on your remaining due tax. May you have any doubt regarding the difference between the conditions, have a look at the instructions.

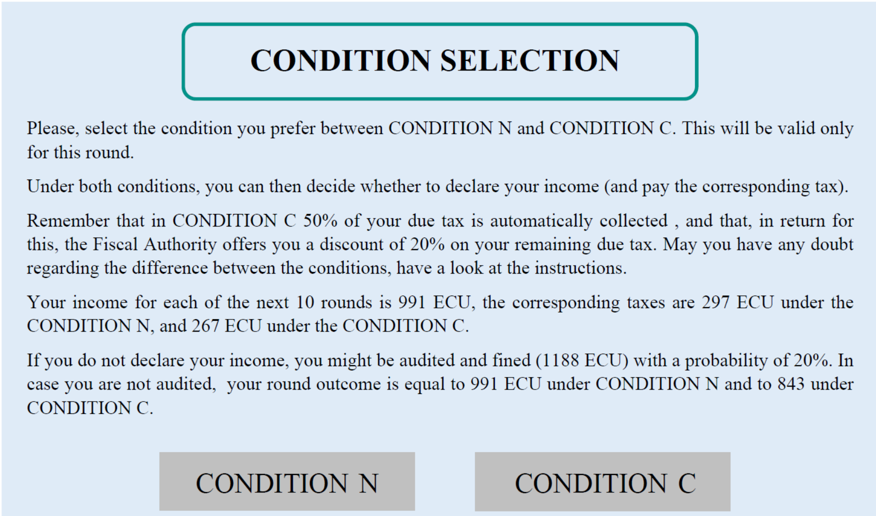
Your income for each of the next 10 rounds is 991 ECU, the corresponding taxes are 297 ECU under the CONDITION N, and 267 ECU under the CONDITION C.

If you do not declare your income, you might be audited and fined (1188 ECU) with a probability of 20%. In case you are not audited, your round outcome is equal to 991 ECU under CONDITION N and to 843 under CONDITION C.

CONDITION N CONDITION C

At the beginning of each period in the Ongoing treatment, the following condition-selection screen is displayed.

Figure 4.11: Condition Selection Screen - Ongoing Treatment

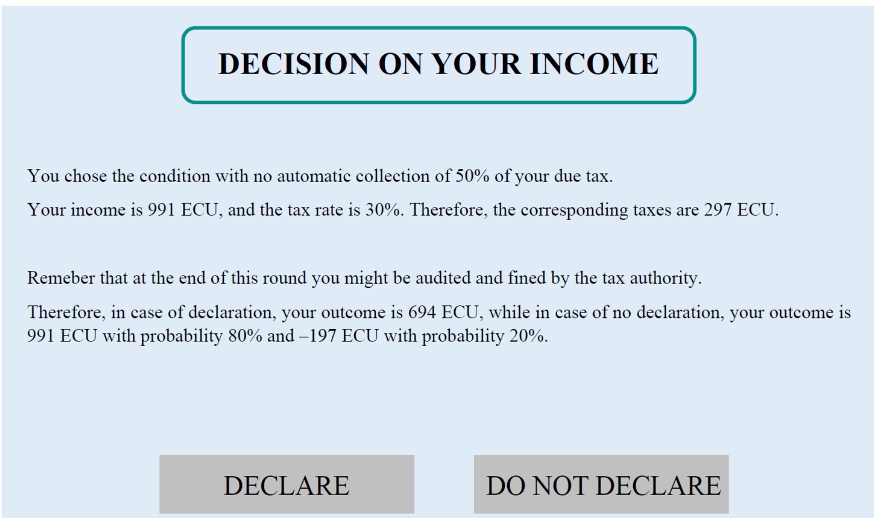


THIRD PART - TAX PAYMENT DECISION

At this stage, subjects know the amount of tax (discount included for Condition C if applicable), the after tax income, and final earnings in the case of evasion. No complex computations are required to decide whether to send the income declaration.

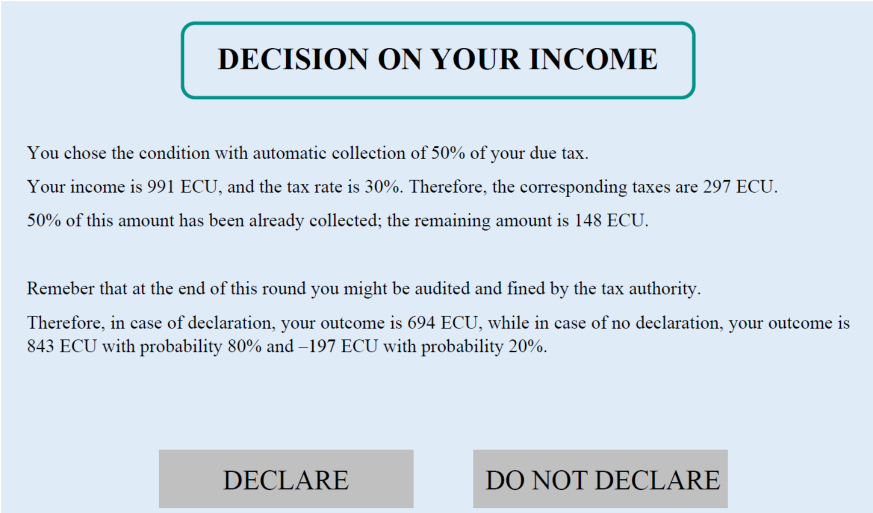
If the No Commitment (N) condition is selected, the following screen is displayed.

Figure 4.12: No Commitment - Tax Payment Screen



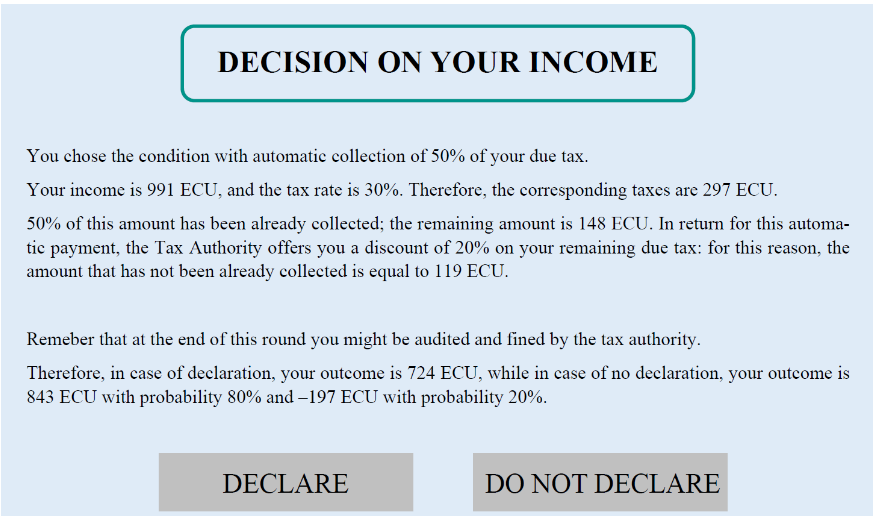
If the Commitment (C) condition is selected in the Control treatment, the following screen is displayed.

FIGURE 4.13: Commitment in the Control Treatment - Tax Payment Screen



If the Commitment (C) condition is selected either in the Planning or in the Ongoing treatment, the following screen is displayed.

FIGURE 4.14: Commitment in the Planning and in the Ongoing Treatment - Tax Payment Screen



Concluding Remarks

This thesis mainly adds to the literature on Decision from Experience (DfE), defined in *Chapter 1*. By means of an experimental approach, it investigates the effect of experience not considered as history but as a source of information acquisition on human behavior in the field of decision-making under risk. In line with previous evidence, this research shows that reliance on experience affects people's probability estimation and weighting (description-experience gap). This leads to decisions that differ from those induced by a full knowledge of true outcome distributions. As reported in *Chapter 2*, in a simplified system of delegated choices (either medical or financial, for instance), the effect of the learning modality on the extent of risk-taking and efficiency varies according to the role of the decision-maker (principal vs. agent). In addition, the source of knowledge has a systematic impact on principals' expectations on agents' decisions, and therefore on their willingness to delegate.

Some related phenomena are described in *Chapter 3*, which provides not only a methodological analysis of external validity of laboratory experiments, but also an overview of behavioral anomalies such as the bomb crater effect and the echo effect identified in the field of tax research, and strictly linked to the effect of experience on probability estimation and weighting. *Chapter 4* focuses on an experimental investigation of the planning-ongoing gap in the context of tax evasion, and tests the introduction of a gentle enforcement system that relies on people's probability weighting in order to sustain compliance.

This last section summarizes the main findings of the thesis, provides an overview of the possible implications, and suggests potential extensions for further research.

Overview: Main Findings and Future Research

Chpater 2 is a novel contribution to the literature on decision-making under risk: it provides an analysis of the impact that different information acquiring modalities have on risk-taking and decision efficiency in the context of delegation. The literature on decision-making under risk investigates how people choose in different information acquiring situations (description vs. experience), and how the extent of risk-taking varies when a three-party agency dilemma is involved - i.e., when decision are made by either principals (recipients) or agents. Nevertheless, the two strands have never been combined: in this sense, the present paper can be considered as a first step towards briding this gap. In addition, it also investigates the problem from delegating agents' viewpoint, by eliciting people's willingness to delegate risky decisions.

Specifically, we replicate in the laboratory two learning modes. Under description decision-makers receive full information on the main features of the given problem; therefore, this can be regarded as a situation in which people have previously undertaken a formal training. According to the other modality, decision-makers can acquire information from personally experiencing over time similar problems, without having a solid background of knowledge. In such a framework, the main goal of the research is the investigation of how risk-taking and delegation decisions are affected by the way in which principals and agents gain their expertise on the problem. We find that, irrespective of the learning modality, principals' decisions are more ambitious and efficient; this is more evident under description, as learning from experience impairs both principals' and agents' decision efficiency. As a matter of fact, information acquired from experience is more likely to be incomplete and therefore inaccurate and misleading. Such an incompleteness can be balanced only by exerting a substantial effort in information gathering: the longer experience is, the closer the decision problem gets to the one presented under description. In this respect, we observe that only principals tend to adapt their effort extent to the complexity of the problem; in contrast, agents' lack of effort contributes to determine the poor quality of their portfolios.

Hence, it is evident that *Chapter 2* does not focus on the motives for delegation, such as the potential difference in expertise between agents and principals. We provide no information on agents' expertise and competences, and agents receive no monetary incetives for their decision efficiency. In line with this, in every treatment, we observe a highly positive willingness to pay not to delagate: principals seem to predict agents' lack of effort and poor performance; apparently, they do not expect agents to have a higher expertise. In turn, agents might predict principals' unwillingness to delegate, and make less efficient decisions. To control for this, a future extension of the present research may include the manipulation of agents' competences: this would allow the investigation of the effect produced by differences in expertise on delegation decisions. In addition, we can explicitly elicit principals' beliefs on agents' effort and performance, instead of simply inferring them by relying on principals' bids to avoid delegation. Simmetrically, we can elicit agents' beliefs on principals' preference for delegation.

Given the focus of our investigation, we also experimentally rule out the delegation motif related to the trade-off between the monetary cost of delegation and the time consumption that self decision-making requires. In fact, irrespective of principals' delegation preference, all participants are asked to build the portfolios. In general, even if principals have the chance of minimizing the time they devote to the building process through blind prospect selection, they bear some psychological costs, and, at the end of each part of the experiment, are willing to pay to retain their portfolios. In addition, as in such a structure principals not only build their own portfolios, but also have the chance to pay not to delegate, agents might feel less responsible for their choices, as they predict principals' preference for delegation avoidance. In this respect, an interesting future treatment might be characterized by the removal of principals' opportunity to decide about delegation. This would induce agents to feel more responsible, as their choices actually affect principals' final monetary well-being.

The role of experience as a learning modality is further investigated in *Chpater 3* and *Chapter 4*, which however address the problem of probability-related choice anomalies in the field of tax compliance. In fact, a closer look at such a field provides an interesting analysis of specific aspects of taxpayers' decision process involving both objective and subjective probabilities. In this framework, *Chapter 3* has a twin role. On the one hand, Section 3.1 and 3.2 contain a literature review on tax compliance, with a specific focus on the the effect of probability estimation and weighting on compliance decisions. On the other hand, a novel methodological analysis on the use of tax laboratory experiments is provided.

Hence, we introduce the rational choice model developed by Allingham and Sandmo (1972); then, we include experimental evidence on the effect of uncertainty and experience on taxpayers' decisions (e.g., Friedland, 1982; Spicer and Thomas, 1982; Alm, Jackson, and McKee, 1992a; Alm, McClelland, and Schulze, 1992; Hessing et al., 1992; Sheffrin and Triest, 1992; Scholz and Pinney, 1995; Mittone, 2006; Kastlunger et al., 2009). In this respect, we refer back to Guala and Mittone (2005): in their analysis of tax compliance, they provide examples of common biases emerging in probabilistic reasoning. Laboratory findings suggest the necessity of taking into account subjects' probability estimation and weighting when modeling taxpayers' decision process

(Mittone, 1997). To the purpose of gaining a better comprehension of people's tax choices, laboratory experiments can play a fundamental role. According to Guala and Mittone (2005), experiments can be considered as *mediators* between theoretical models and the intended domain of application, as they allow experimenters to fully control and manipulate an artificial environment in which real people's behavior is observed. Nevertheless, since experiments are usually required to tackle and balance the internal and external validity issues, they might not be able to completely bridge the gap between the target and the theoretical model. For this reason, as suggested by Guala and Mittone (2005), experiments might be intended to discover new empirical phenomena, such as psychological effects, biases, or heuristics. In doing so, experiments may contribute to the creation of a *library of phenomena*, and thus discover new facts useful from a policy viewpoint. Among these pheomena, the *bomb crater effect* an the *echo effect* can be named: laboratory findings ahow that, in contrast to Bayesian updating, people's probabilistic reasoning is biased by previous experience. In spite of knowing the objective probability of fiscal investigations, taxpayers seem to evaluate or weight such a probability according to their experience, and rely on this in order to make their compliance decisions.

Building on this, *Chapter 3* provides a novel extension of the mediation approach proposed by Guala and Mittone (2005): agent-based simulations are presented as a useful tool, intended to help experiments bridge the gap between theoretical models and and the real domain of application. Precisely, human-calibration of artificial agents is suggested as one of the possible realizations of the synergic use of laboratory experiments and computer simulations. Experimental investigations might contribute to the identification of specific and isolated economic and psychological phenomena, but they cannot control for all the cognitive drivers involved in the tax decision process. For this reason, as in the case of probabilistic biases, laboratory evidence might lack external validity, and need to be further tested for policy potentially relevant implications. This can be achieved by means of agent-based simulations, which can provide valuable insights of cognitive nature, by validating laboratory findings, and helping understand complex cognitive processes. In addition to this, simulations allow the combination of micro- and macro-level factors actually interacting outside the laboratory and determining people's compliance: cognitive drivers can be investigated in a population of interacting heterogeneous taxpayers.

Nevertheless, the synergic approach proposed in *Chapter 3* can be further developed, so that the realism and the potential of the mediator role of experiments between theoretical models and the intended domain of application is increased. Agents' features can be manipulated in order to take into account the effect of metadata: population heterogeneity can be built on taxpayers' past experience, age, and culture. To this purpose, not only collected empirical and experimental evidence can be combined and used for a better calibration of agents, but also the laboratory simultaneous interaction between human subjects and artificial agents can be implemented. This would allow researchers to draw interesting and externally valid inferences on people's behavior, which is first experimentally investigated and observed outside the walls of the laboratory, and subsequently tested in a rather realistic context by means of a more complex computer-based simulation.

Such a combined approach might be applied also to the experimental findings of the laboratory investigation described in *Chapter 4*. As a matter of fact, this chapter deals with the effect of probability weighting on compliance decisions, and aims to test and suggest the adoption of an enforcement system built upon emerging human reasoning biases. Agent-based simulations may help understand the potential of our findings, and identify different decision contexts to which they can be applied. *Chapter 4* starts from the robust experimental evidence on the existence of a substantial gap between how people make decisions from description versus experience. On the one hand, Prospect Theory is in general adopted with non-trivial choice problems on monetary gambles that explicitly describe outcomes and associated probabilities. Experimental investigations reveal that people's decisions are driven by a probability weighting function, according to which small probabilities are generally overweighted. On the other hand, Decision from Experience is characterized by repeated decisions, and lack of prior information on payoff and probability distributions. Decision makers have to rely on their experience, that is on the partial information they collect during their iterated choices or trials. Because of the structural features of DfE, its experimental application shows that low probabilities are underweighted with respect to objective probabilities.

Similarly, experimental studies have identified a difference between planning and ongoing decisions: when asked to plan their actions, people often overweight events with small probabilities; while in ongoing - i.e. real time, and, in general, repeated - decisions, they tend to underweight these events, and, thus, to behave as if they ignored them. The planning-ongoing gap has been documented in a variety of decision contexts, for instance related to compliance with safety rules in the workplace. Although workers are aware of the likelihood of accidents, and thus of the importance of rules compliance, they actually tend not to obey safety rules, as if they underestimated the probability of accidents. In such a context, the understanding of the planning-ongoing gap can be used to enhance safety: it is possible to build an enformcement system on the fact that workers, when asked to plan and think about the relevant risks in their workplace, state they are willing to behave safely.

Building on this, in *Chapter 4* we argue that such a problem of rules violation has a structural similarity to the problem of tax evasion. Many taxpayers, such as retailers or taxi drivers, face very frequent tax payment decisions: in their ongoing decisions, i.e., everytime they sell a good or a service, they can decide whether to declare the transaction and give a proof of payment to their customers. Everytime they opt for not printing the receipt, they incurr the risk of being investigated and found non compliant. Nevertheless, according to the planning-ongoing gap, it is reasonable to expect that taxpayers are less willing to evade, for instance when asked to evaluate the value of compliance at the beginning of the business year. Like in the case of workers evaluating relevant risks in their workplace, we experimentally test the availability of a system of partial commitment, and observe a discrepancy between taxpayers' plans about income declaration and their actual behavior. Taxpayers are requested to make sequential income declaration decisions; to do this, they are offered the possibility to commit to automatically pay half of the period due taxes in return for a discount on the taxes computed on the remaining part of income they can decide whether to declare. In the Planning Treatment, the commitment lasts for 10 rounds; in the Ongoing Treatment, the commitment needs to be adopted every round in order to be valid. In line with previous experimental results, we find that policy tools aimed to sustain compliance should take into account the planning-ongoing gap. In fact, the long lasting effect of the commitment in the Planning Treatment invloves an ex-ante evaluation of the risky decision problem: this helps foster compliance, by inducing the majority of participants to overweight the risk associated to evasion and, interestingly, stick with their compliance plans.

In the light of this, as suggested in *Chapter 3*, both macroeconomic behavioral styles, and microeconomic behavioral regularities emerging in this laboratory experiment (such as the bomb crater effect) can be used to calibrate computational agents. This would allow us not only to comprehend psychological drivers determining the gap in taxpayers' behavior, but also to check the robustness of our policy results. Specifically, thanks to an agent-based simulation, it would be possible to test and manipulate a variety of factors, and observe long-run effects on compliance. In fact, as shown in *Chapter 4*, the attractiveness of the planned commitment tends to decrease over time (see Figure 4.4); in this respect, a simulation-based investigation can help understand how to make the commitment adoption more stable and therefore sustain compliance also in the long run. Furthermore, the implementation of computational models also allow the study of the effect of a different timing on decisions, in terms of a different planning time span, and the effect of a diverse incentive in return for the adoption of the commitment mechanism. In our experiment, we only considered the introduction of a discount on due tax; nevertheless, also a reduction in the audit probability can be implemented. Finally, we can investigate to what extent our laboratory results can be applied to a variety of contexts - such as the one of sector studies - in order to effectively foster compliance.

Bibliography

Agranov, Marina, Alberto Bisin, and Andrew Schotter (2014). "An experimental study of the impact of competition for Other People's Money: the portfolio manager market". In: *Experimental Economics* 17.4, pp. 564–585.

Allingham, Michael G. and Agnar Sandmo (1972). "Income Tax Evasion: A theoretical Analysis". In: *Journal of Public Economics* 1, pp. 323–338.

Alm, James (1999). "Tax compliance and administration". In: *Public Administration and Public Policy* 72, pp. 741–768.

— (2010). "Testing behavioral public economics theories in the laboratory". In: *National Tax Journal* 63.4, pp. 635–658.

Alm, James, Betty Jackson, and Michael McKee (1992a). "Deterrence and beyond: Toward a kinder, gentler IRS". In: *Why people pay taxes* 1, pp. 311–329.

Alm, James, Betty R Jackson, and Michael McKee (1992b). "Estimating the determinants of taxpayer compliance with experimental data". In: *National Tax Journal* 45.1, pp. 107–114.

Alm, James, Gary H McClelland, and William D Schulze (1992). "Why do people pay taxes?" In: *Journal of public Economics* 48.1, pp. 21–38.

Alm, James, Isabel Sanchez, and Ana De Juan (1995). "Economic and noneconomic factors in tax compliance". In: *Kyklos* 48.1, pp. 1–18.

Andersson, Ola, Håkan J Holm, Jean-Robert Tyran, and Erik Wengström (2013). "Risking other peoples money: Experimental evidence on bonus schemes, competition, and altruism". In: *Research Institute of Industrial Economics (IFN) Working Paper* 989.

— (2014). "Deciding for others reduces loss aversion". In: *Management Science* 62.1, pp. 29–36.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998). "Tax compliance". In: *Journal of Economic Literature* 36.2, pp. 818–860.

Antonides, Gerrit and Henry SJ Robben (1995). "True positives and false alarms in the detection of tax evasion". In: *Journal of Economic Psychology* 16.4, pp. 617–640.

Antunes, Luis, Joao Balsa, Ana Respício, and Helder Coelho (2007). "Tactical exploration of tax compliance decisions in multi-agent based simulation". In: *Multi-Agent-Based Simulation VII*. Springer, pp. 80–95.

Arora, Neeraj K and Colleen A McHorney (2000). "Patient preferences for medical decision making: who really wants to participate?" In: *Medical care* 38.3, pp. 335–341.

Baldry, Jonathan C (1986). "Tax evasion is not a gamble: A report on two experiments". In: *Economics Letters* 22.4, pp. 333–335.

— (1987). "Income tax evasion and the tax schedule: Some experimental results". In: *Public Finance= Finances publiques* 42.3, pp. 357–83.

Barron, Greg and Ido Erev (2003). "Small feedback-based decisions and their limited correspondence to description-based decisions". In: *Journal of Behavioral Decision Making* 16.3, pp. 215–233.

Barron, Greg and Eldad Yechiam (2009). "The coexistence of overestimation and underweighting of rare events and the contingent recency effect". In: *Judgment and Decision Making* 4.6, p. 447.

Bazart, Cécile, Aurélie Bonein, Sascha Hokamp, and Götz Seibold (2016). "Behavioural Economics and Tax Evasion–Calibrating an Agent-based Econophysics Model with Experimental Tax Compliance Data". In: *forthcoming Journal of Tax Administration,*

Becker, Gary S. (1976). *The Economic Approach to Human Behavior*. University of Chicago press.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1964). "Measuring utility by a single-response sequential method". In: *Behavioral science* 9.3, pp. 226–232.

Benjamin, Daniel and David V Budescu (2015). "Advice from Experience: Communicating Incomplete Information Incompletely". In: *Journal of Behavioral Decision Making* 28.1, pp. 36–49.

Benjamini, Yael and Shlomo Maital (1985). "Optimal Tax Evasion & Optimal Tax Evasion Policy Behavioral Aspects". In: *The economics of the shadow economy*. Springer, pp. 245–264.

Bergman, Marcelo and Armando Nevarez (2006). "Do audits enhance compliance? An empirical assessment of VAT enforcement". In: *National tax journal* 59.4, pp. 817–832.

Bernasconi, Michele (1998). "Tax evasion and orders of risk aversion". In: *Journal of Public Economics* 67.1, pp. 123–134.

Bernasconi, Michele and Alberto Zanardi (2004). "Tax evasion, tax rates, and reference dependence". In: *FinanzArchiv: Public Finance Analysis* 60.3, pp. 422–445.

Blais, Ann-Renée and Elke U Weber (2006). "A domain-specific risk-taking (DOSPERT) scale for adult populations". In: *Judgment and Decision Making* 1.1.

Bloomquist, Kim (2011a). "Multi-agent Simulation of Taxpayer Reporting Compliance". In: *International Conference on Taxation Analysis and Research, London, UK.*

— (2011b). "Tax compliance as an evolutionary coordination game: an agent-based approach". In: *Public Finance Review* 39.1, pp. 25–49.

Bloomquist, Kim M (2004a). "Modeling taxpayers response to compliance improvement alternatives". In: *Annual Conference of the North American association for computational social and organizational sciences, Pittsburgh, PA*.

— (2004b). "Multi-agent based simulation of the deterrent effects of taxpayer audits". In: *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*. Vol. 97, pp. 159–173.

— (2006). "A comparison of agent-based models of income tax evasion". In: *Social Science Computer Review* 24.4, pp. 411–425.

— (2007). "Taxpayer compliance simulation: A multi-agent based approach". In: *Social Simulation: Technologies, Advances and New Discoveries: Technologies, Advances and New Discoveries*, p. 13.

Bolton, Gary E, Axel Ockenfels, and Julia Stauf (2015). "Social responsibility promotes conservative risk behavior". In: *European Economic Review* 74, pp. 109–127.

Botti, Simona and Sheena S. Iyengar (2004). "The psychological pleasure and pain of choosing: when people prefer choosing at the cost of subsequent outcome satisfaction." In: *Journal of personality and social psychology* 87.3, p. 312.

Botti, Simona and Ann L McGill (2006). "When choosing is not deciding: The effect of perceived responsibility on satisfaction". In: *Journal of Consumer Research* 33.2, pp. 211–219.

— (2011). "The locus of choice: Personal causality and satisfaction with hedonic and utilitarian decisions". In: *Journal of Consumer Research* 37.6, pp. 1065–1078.

Boylan, Scott J (2010). "Prior audits and taxpayer compliance: Experimental evidence on the effect of earned versus endowed income". In: *Journal of the American Taxation Association* 32.2, pp. 73–88.

Boylan, Scott J and Geoffrey B Sprinkle (2001). "Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income". In: *Journal of the American Taxation Association* 23.1, pp. 75–90.

Burlando, Roberto M and Francesco Guala (2005). "Heterogeneous agents in public goods experiments". In: *Experimental Economics* 8.1, pp. 35–54.

Camerer, Colin and Teck Ho (1999). "Experience-weighted attraction learning in normal form games". In: *Econometrica* 67.4, pp. 827–874.

Camerer, Colin F. (2013). "Experimental, cultural, and neural evidence of deliberate prosociality". In: *Trends in Cognitive Sciences* 17.3, pp. 106 – 108. ISSN: 1364-6613. DOI: `http://dx.doi.org/10.1016/j.tics.2013.01.009`. URL: `http://www.sciencedirect.com/science/article/pii/S1364661313000235`.

Camerer, Colin F, George Loewenstein, and Matthew Rabin (2011). *Advances in behavioral economics*. Princeton University Press.

Camilleri, Adrian R and Ben R Newell (2009). "The role of representation in experience-based choice". In: *Judgment and Decision Making* 4.7, p. 518.

— (2011a). "Description-and experience-based choice: Does equivalent information equal equivalent choice?" In: *Acta Psychologica* 136.3, pp. 276–284.

— (2011b). "When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms". In: *Psychonomic bulletin & review* 18.2, pp. 377–384.

— (2013). "The long and short of it: Closing the description-experience gap by taking the long-run view". In: *Cognition* 126.1, pp. 54–71.

Camilleri, Adrian Ryan (2011). "The Psychological Mechanisms Underpinning Experience-based Choice". PhD thesis. The University of New South Wales Sydney, Australia.

Casal, Sandro and Luigi Mittone (2016). "Social esteem versus social stigma: The role of anonymity in an income reporting game". In: *Journal of Economic Behavior & Organization* 124, pp. 55–66.

Casey, Jeff T and John T Scholz (1991a). "Beyond deterrence: behavioral decision theory and tax compliance". In: *Law and Society Review*, pp. 821–843.

— (1991b). "Boundary effects of vague risk information on taxpayer decisions". In: *Organizational Behavior and Human Decision Processes* 50.2, pp. 360–394.

Chakravarty, Sujoy, Glenn W Harrison, Ernan E Haruvy, and E Elisabet Rutström (2011). "Are you risk averse over other people's money?" In: *Southern Economic Journal* 77.4, pp. 901–913.

Cheung, Yin-Wong and Daniel Friedman (1998). "A comparison of learning and replicator dynamics using experimental data". In: *Journal of economic behavior & organization* 35.3, pp. 263–280.

Choo, Lawrence, MA Fonseca, and GD Myles (2016). "Do Students Behave Like Real Taxpayers? Experimental Evidence on Taxpayer Compliance From the Lab and From the Field". In: *Journal of Economic Behavior & Organization* 124, pp. 102–114.

Copeland, Phyllis V and Andrew D Cuccia (2002). "Multiple determinants of framing referents in tax reporting and compliance". In: *Organizational Behavior and Human Decision Processes* 88.1, pp. 499–526.

Cowell, Frank (1991). "Tax-evasion experiments: an economist's view". In:

Cowell, Frank A and James PF Gordon (1988). "Unwillingness to pay: Tax evasion and public good provision". In: *Journal of Public Economics* 36.3, pp. 305–321.

Crosetto, Paolo and Antonio Filippin (2013). "The bomb risk elicitation task". In: *Journal of Risk and Uncertainty* 47.1, pp. 31–65.

Davis, Jon S, Gary Hecht, and Jon D Perkins (2003). "Social behaviors, enforcement, and tax compliance dynamics". In: *The Accounting Review* 78.1, pp. 39–69.

Dawes, Robyn M (1998). "Behavioral decision making and judgment." In: *The handbook of social psychology.* McGraw-Hill, pp. 497–548.

DeBacker, Jason Matthew, Bradley T Heim, Anh Tran, and Alex Yuskavage (2015). "Legal enforcement and corporate behavior: An analysis of tax aggressiveness after an audit". In: *Journal of Law and Economics* 58.2, pp. 291–324.

Dhami, Sanjit and Ali Al-Nowaihi (2007). "Why do people pay taxes? Prospect theory versus expected utility theory". In: *Journal of Economic Behavior & Organization* 64.1, pp. 171–192.

Duffy, John (2006). "Agent-based models and human subject experiments". In: *Handbook of computational economics* 2, pp. 949–1011.

Durham, Yvonne, Tracy S Manly, and Christina Ritsema (2014). "The effects of income source, context, and income level on tax compliance decisions in a dynamic experiment". In: *Journal of Economic Psychology* 40, pp. 220–233.

Einhorn, Hillel J and Robin M Hogarth (1985). "Ambiguity and uncertainty in probabilistic inference." In: *Psychological review* 92.4, pp. 433–461.

Eisenhauer, Joseph G (2006). "The shadow price of morality". In: *Eastern Economic Journal* 32.3, pp. 437–456.

— (2008). "Ethical preferences, risk aversion, and taxpayer behavior". In: *The Journal of Socio-Economics* 37.1, pp. 45–63.

Erard, Brian and Jonathan S Feinstein (1994). "Honesty and evasion in the tax compliance game". In: *The RAND Journal of Economics*, pp. 1–19.

Erev, Ido (2007). "On the weighting of rare events and the economics of small decisions". In: *Developments on Experimental Economics.* Springer, pp. 59–73.

Erev, Ido, Eyal Ert, and Eldad Yechiam (2008). "Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions". In: *Journal of Behavioral Decision Making* 21.5, pp. 575–597.

Erev, Ido and Alvin E Roth (1998). "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria". In: *American Economic Review* 88.4, pp. 848–881.

— (2014). "Maximization, learning, and economic behavior". In: *Proceedings of the National Academy of Sciences* 111.Supplement 3, pp. 10818–10825.

Erev, Ido, Dotan Rodensky, Mark-Alain Levi, Michal Englard-Hershler, Hanna Admi, and Yoel Donchin (2010). "The value of gentle reminderon safe medical behaviour". In: *Quality and Safety in Health Care*, qshc–2009.

Eriksen, Kristoffer W and Ola Kvaløy (2009). "Myopic Investment Management*". In: *Review of Finance*, rfp019.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental economics* 10.2, pp. 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). "Are people conditionally cooperative? Evidence from a public goods experiment". In: *Economics Letters* 71.3, pp. 397–404.

Fischhoff, Baruch (1988). "Judgment and decision making." In:

Fischhoff, Baruch, Aandrew M Parker, Wändi Bruine de Bruin, Julie Downs, Claire Palmgren, Robyn Dawes, and Charles F Manski (2000). "Teen expectations for significant life events". In: *The Public Opinion Quarterly* 64.2, pp. 189–205.

Fox, Craig R and Liat Hadar (2006). "" Decisions from experience"= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)". In: *Judgment and Decision Making* 1.2, p. 159.

Friedland, Nehemia (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". In: *Journal of Applied Social Psychology* 12.1, pp. 54–59.

Friedman, Milton (1953). *Essays in positive economics.* University of Chicago Press.

Friedman, Milton and Leonard J Savage (1948). "The utility analysis of choices involving risk". In: *The journal of political economy*, pp. 279–304.

Garrido, Nicolás and Luigi Mittone (2008). "A description of experimental tax evasion behavior using finite automata: the case of chile and italy". In: *Cognitive and Experimental Economics Laboratory (CEEL) Working Papers* 809.

— (2013). "An agent based model for studying optimal tax collection policy using experimental data: The cases of Chile and Italy". In: *The Journal of Socio-Economics* 42, pp. 24–30.

Gemmell, Norman and Marisa Ratto (2012). "Behavioral responses to taxpayer audits: evidence from random taxpayer inquiries". In: *National Tax Journal* 65.1, pp. 33–58.

Gonzalez, Cleotilde and Varun Dutt (2011). "Instance-based learning: Integrating sampling and repeated decisions from experience." In: *Psychological review* 118.4, p. 523.

Gordon, James PP (1989). "Individual morality and reputation costs as deterrents to tax evasion". In: *European economic review* 33.4, pp. 797–805.

Guala, Francesco (1998). "Experiments as mediators in the non-laboratory sciences". In: *PHILOSOPHICA-GENT-* 62.2, pp. 57–76.

— (1999). "The problem of external validity (or parallelism) in experimental economics". In: *Social science information* 38.4, pp. 555–573.

Guala, Francesco (2002). “Models, simulations, and experiments”. In: *Model-based reasoning: Science, Technology, Values*. Springer, pp. 59–74.
— (2003). “Experimental localism and external validity”. In: *Philosophy of science* 70.5, pp. 1195–1205.
Guala, Francesco and Luigi Mittone (2005). “Experiments in economics: External validity and the robustness of phenomena”. In: *Journal of Economic Methodology* 12.4, pp. 495–515.
Hadar, Liat and Craig R Fox (2009). “Information asymmetry in decision from description versus decision from experience”. In: *Judgment and Decision Making* 4.4, p. 317.
Harris, Devin K (1988). “The subjective probability of involvement in an overtaking accident in the vicinity of a junction–road user behavior.” In: *In, J. A. Rothengatter R. De Bruin (Eds.), Road user behavior: Theory and research (pp. 163–168). Assen: Van Gorcum.*
Hartocollis, Anemona (2009). “Calorie postings dont change habits, study finds”. In: *The New York Times. October* 6.
Hashimzade, Nigar, Gareth D Myles, and Binh Tran-Nam (2013). “Applications of behavioural economics to tax evasion”. In: *Journal of Economic Surveys* 27.5, pp. 941–977.
Hau, Robin, Timothy J Pleskac, and Ralph Hertwig (2010). “Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities”. In: *Journal of Behavioral Decision Making* 23.1, pp. 48–68.
Hau, Robin, Timothy J Pleskac, Jürgen Kiefer, and Ralph Hertwig (2008). “The description–experience gap in risky choice: The role of sample size and experienced probabilities”. In: *Journal of Behavioral Decision Making* 21.5, pp. 493–518.
Herrnstein, Richard J, George F Loewenstein, Drazen Prelec, and William Vaughan (1993). “Utility maximization and melioration: Internalities in individual choice”. In: *Journal of behavioral decision making* 6.3, pp. 149–185.
Hertwig, Ralph and Ido Erev (2009). “The description–experience gap in risky choice”. In: *Trends in cognitive sciences* 13.12, pp. 517–523.
Hertwig, Ralph, Barron, Weber Greg, Elke U, and Ido Erev (2004). “Decisions from experience and the effect of rare events in risky choice”. In: *Psychological science* 15.8, pp. 534–539.
Hessing, Dick J, Henk Elffers, Henry SJ Robben, and Paul Webley (1992). “Does deterrence deter? Measuring the effect of deterrence on tax compliance in field studies and experimental studies”. In: *Why people pay taxes: Tax compliance and enforcement*, pp. 291–305.
Hokamp, Sascha (2014). “Dynamics of tax evasion with back auditing, social norm updating, and public goods provision–An agent-based simulation”. In: *Journal of Economic Psychology* 40, pp. 187–199.
Hokamp, Sascha and Michael Pickhardt (2010). “Income tax evasion in a society of heterogeneous agents–Evidence from an agent-based model”. In: *International Economic Journal* 24.4, pp. 541–553.
Hokamp, Sascha and Götz Seibold (2014a). “How much rationality tolerates the shadow economy?–An agent-based econophysics approach”. In: *Advances in Social Simulation*. Springer Berlin Heidelberg, pp. 119–128.

— (2014b). “Tax Compliance and Public Goods Provision–An Agent-based Econophysics Approach”. In: *Central European Journal of Economic Modelling and Econometrics,* 6.4, pp. 217–236.

Hsee, Christopher K and Elke U Weber (1997). “A fundamental prediction error: Self–others discrepancies in risk preference.” In: *Journal of experimental psychology: general* 126.1, p. 45.

Jessup, Ryan K, Anthony J Bishara, and Jerome R Busemeyer (2008). “Feedback produces divergence from prospect theory in descriptive choice”. In: *Psychological Science* 19.10, pp. 1015–1022.

Kahneman, Daniel and Amos Tversky (1979). “Prospect theory: An analysis of decision under risk”. In: *Econometrica: Journal of the econometric society* 47.2, pp. 263–291.

— (2000). *Choices, values, and frames.* Cambridge University Press.

Kareev, Yaakov (1995). “Through a narrow window: Working memory capacity and the detection of covariation”. In: *Cognition* 56.3, pp. 263–269.

— (2000). “Seven (indeed, plus or minus two) and the detection of correlations.” In: *Psychological review* 107.2, p. 397.

Kastlunger, Barbara, Erich Kirchler, Luigi Mittone, and Julia Pitters (2009). “Sequences of audits, tax compliance, and taxpaying strategies”. In: *Journal of Economic Psychology* 30.3, pp. 405–418.

Kirchler, Erich (2007). *The economic psychology of tax behaviour.* Cambridge University Press.

Kirchler, Erich, Erik Hoelzl, and Ingrid Wahl (2008). “Enforced versus voluntary tax compliance: The slippery slope framework”. In: *Journal of Economic Psychology* 29.2, pp. 210–225.

Knight, Frank H (1921). “Risk, uncertainty and profit”. In: *New York: Hart, Schaffner and Marx.*

Kogler, Christoph, Luigi Mittone, and Erich Kirchler (2016). “Delayed feedback on tax audits affects compliance and fairness perceptions”. In: *Journal of Economic Behavior & Organization* 124, pp. 81–87.

Koritzky, Gilly and Eldad Yechiam (2010). “On the robustness of description and experience based decision tasks to social desirability”. In: *Journal of Behavioral Decision Making* 23.1, pp. 83–99.

Korobow, Adam, Chris Johnson, and Robert Axtell (2007). “An agent–based model of tax compliance with social networks”. In: *National Tax Journal*, pp. 589–610.

Kvaløy, Ola, Kristoffer Eriksen, and Miguel Luzuriaga (2014). *Risk-taking with Other Peoples Money.* Tech. rep. University of Stavanger.

Levenson, Hanna (1972). “Distinctions within the concept of internal-external control: Development of a new scale.” In: *Proceedings of the Annual Convention of the American Psychological Association.* American Psychological Association.

Levinson, Wendy, Audiey Kao, Alma Kuby, and Ronald A Thisted (2005). “Not all patients want to participate in decision making”. In: *Journal of general internal medicine* 20.6, pp. 531–535.

Levitt, Steven D (2006). “The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions.” In: *Journal of Political Economy,* 114, 137.

Levitt, Steven D and John A List (2007). “What do laboratory experiments measuring social preferences reveal about the real world?” In: *The journal of economic perspectives* 21.2, pp. 153–174.

Lima, Francesco WS (2010). "Analysing and controlling the tax evasion dynamics via majority-vote model". In: *Journal of Physics: Conference Series*. Vol. 246. 1, pp. 1–12.

Lima, Francesco WS and Georg Zaklan (2008). "A multi-agent-based approach to tax morale". In: *International Journal of Modern Physics C: Computational Physics and Physical Computation* 19.12, pp. 1797–1808.

Lima, FWS (2012a). "Tax Evasion and Nonequilibrium Model on Apollonian Networks". In: *International Journal of Modern Physics C: Computational Physics and Physical Computation* 23.11.

— (2012b). "Tax evasion dynamics and Zaklan model on opinion-dependent network". In: *International Journal of Modern Physics C: Computational Physics and Physical Computation* 23.6.

Loewenstein, George and Peter Ubel (2010). "Economics behaving badly". In: *The New York Times* 14.

Lumpkin, James R (1988). "Establishing the validity of an abbreviated locus of control scale: Is a brief Levenson's scale any better?" In: *Psychological Reports* 63.2, pp. 519–523.

Malmendier, Ulrike and Stefan Nagel (2016). "Learning from inflation experiences". In: *The Quarterly Journal of Economics* 131.1, pp. 53–87.

Marchiori, Davide, Sibilla Di Guida, and Ido Erev (2013). "Noisy Retrievers and the Fourfold Reaction to Rare Events". In: *Department of Management, Università Ca'Foscari Venezia Working Paper* 3.

Mattila, Anna S (2010). "Do women like options more than men? An examination in the context of service recovery". In: *Journal of Services Marketing* 24.7, pp. 499–508.

McColl-Kennedy, Janet R, Catherine S Daus, and Beverley A Sparks (2003). "The role of gender in reactions to service failure and recovery". In: *Journal of Service Research* 6.1, pp. 66–82.

Méder, Zsombor Z, András Simonovits, and János Vincze (2012). "Tax morale and tax evasion: Social preferences and bounded rationality". In: *Economic Analysis and Policy* 42.2, pp. 257–272.

Melumad, Nahum D and Dilip Mookherjee (1989). "Delegation as commitment: the case of income tax audits". In: *The RAND Journal of Economics*, pp. 139–163.

Michael, J Graetz and L Wilde Louls (1985). "The economics of tax compliance: Facts and fantasy." In: *National Tax Journal* 38.3, p. 355.

Mittone, Luigi (1997). "Subjective versus objective probability: Results from seven experiments on fiscal evasion". In: *CEEL Working Paper 4-97*.

— (2001). "Vat evasion: an experimental approach". In:

— (2002). "Individual styles of tax evasion: an experimental study". In: *CEEL Working Paper 2-02*.

— (2006). "Dynamic behaviour in tax evasion: An experimental approach". In: *The Journal of Socio-Economics* 35.5, pp. 813–835.

Mittone, Luigi and Gian Paolo Jesi (2016). *Heuristic Driven Agents in Tax Evasion: an Agent-based Approach.* Tech. rep. Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia.

Mittone, Luigi, Fabrizio Panebianco, and Alessandro Santoro (2016). "The Bomb-Crater Effect of Tax Audits: Beyond Misperception of Chance". In: *No. 583*.

Mittone, Luigi and Paolo Patelli (2000). "Imitative behaviour in tax evasion". In: *Economic simulations in swarm: Agent-based modelling and object oriented programming*. Springer, pp. 133–158.

Moore, Edward F (1956). "Gedanken-experiments on sequential machines". In: *Automata studies* 34, pp. 129–153.

Muehlbacher, Stephan and Erich Kirchler (2016). "Taxperiments. About the external validity of laboratory experiments in tax compliance research". In: *Die Betriebswirtschaft (DBW)* 76, pp. 7–19.

Myles, Gareth D and Robin A Naylor (1996). "A model of tax evasion with group conformity and social customs". In: *European Journal of Political Economy* 12.1, pp. 49–66.

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

Oei, Shu-Yi and Diane M Ring (2016). "The Tax Lives of Uber Drivers: Evidence from Internet Discussion Forums". In: *Tulane Public Law Research Paper Forthcoming*.

Pahlke, Julius, Sebastian Strasser, and Ferdinand M Vieider (2015). "Responsibility Effects in Decision Making under Risk". WP.

Parke, Emily C (2014). "Experiments, simulations, and epistemic privilege". In: *Philosophy of Science* 81.4, pp. 516–536.

Pickhardt, Michael and Goetz Seibold (2014). "Income tax evasion dynamics: Evidence from an agent-based econophysics model". In: *Journal of Economic Psychology* 40, pp. 147–160.

Ping, Xu and Wang Tao (2007). "Cumulative prospect theory in taxpayer decision making: A theoretical model for withholding phenomenon". In: *2007 International Conference on Management Science and Engineering*. IEEE, pp. 1680–1685.

Piolatto, Amedeo and Matthew D Rablen (2014). "Prospect theory and tax evasion: a reconsideration of the Yitzhaki Puzzle". In:

Piolatto, Amedeo and Gwenola Trotin (2016). "Optimal Income Tax Enforcement under Prospect Theory". In: *Journal of Public Economic Theory* 18.1, pp. 29–41.

Pollmann, Monique MH, J Potters, and S T Trautmann (2014). "Risk taking by agents: The role of ex-ante and ex-post accountability". In: *Economics Letters* 123.3, pp. 387–390.

Quiggin, John (1982). "A theory of anticipated utility". In: *Journal of Economic Behavior & Organization* 3.4, pp. 323–343.

Rakow, Tim, Kali A Demes, and Ben R Newell (2008). "Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice". In: *Organizational Behavior and Human Decision Processes* 106.2, pp. 168–179.

Rakow, Tim and Katherine Miler (2009). "Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs". In: *Memory & cognition* 37.7, pp. 985–1000.

Rakow, Tim and Ben R Newell (2010). "Degrees of uncertainty: An overview and framework for future research on experience-based choice". In: *Journal of Behavioral Decision Making* 23.1, pp. 1–14.

Romera, Maria Elena (2000). "Using finite automata to represent mental models". In: *Unpublished Masters Thesis. San Jose, CA: San Jose State University*.

Rubinstein, Ariel (1986). “Finite automata play the repeated prisoner’s dilemma”. In: *Journal of economic theory* 39.1, pp. 83–96.

Sanchez, Isabel and Joel Sobel (1993). “Hierarchical design and enforcement of income tax policies”. In: *Journal of Public Economics* 50.3, pp. 345–369.

Sapp, Stephen G and Wendy J Harrod (1993). “Reliability and validity of a brief version of Levenson’s locus of control scale”. In: *Psychological Reports* 72.2, pp. 539–550.

Schepanski, Albert and Teri Shearer (1995). “A prospect theory account of the income tax withholding phenomenon”. In: *Organizational Behavior and Human Decision Processes* 63.2, pp. 174–186.

Scholz, John T and Neil Pinney (1995). “Duty, fear, and tax compliance: The heuristic basis of citizenship behavior”. In: *American Journal of Political Science* 39.2, pp. 490–512.

Schurr, Amos, Dotan Rodensky, and Ido Erev (2014a). “Internet app.” In: *¡http://fbm.bgu.ac.il/amos/Peak.html¿.*

— (2014b). “The effect of unpleasant experiences on evaluation and behavior”. In: *Journal of Economic Behavior & Organization* 106, pp. 1–9.

Schurr, Amos, Ilana Ritov, Yaakov Kareev, and Judith Avrahami (2012). “Is that the answer you had in mind? The effect of perspective on unethical behavior”. In: *Judgment and Decision Making* 7.6, p. 679.

Seibold, Götz and Michael Pickhardt (2013). “Lapse of time effects on tax evasion in an agent-based econophysics model”. In: *Physica A: Statistical Mechanics and its Applications* 392.9, pp. 2079–2087.

Selten, Reinhard (1998). “Aspiration adaptation theory”. In: *Journal of mathematical psychology* 42.2, pp. 191–214.

— (2001). “What is bounded rationality”. In: *Bounded rationality: The adaptive toolbox* 13, p. 36.

Sheffrin, S.M. and R.K. Triest (1992). “Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance”. In: *Why People Pay Taxes: Tax Compliance and Enforcement*. Ed. by J. Slemrod. Ann Arbor: University of Michigan Press, 193218.

Simon, Herbert A (1955). “A behavioral model of rational choice”. In: *The quarterly journal of economics*, pp. 99–118.

— (1956). “Rational choice and the structure of the environment.” In: *Psychological review* 63.2, p. 129.

Simonsohn, Uri, Niklas Karlsson, George Loewenstein, and Dan Ariely (2008). “The tree of experience in the forest of information: Overweighing experienced relative to observed information”. In: *Games and Economic Behavior* 62.1, pp. 263–286.

Sipser, Michael (2006). *Introduction to the Theory of Computation*. Vol. 2. Thomson Course Technology Boston.

Slemrod, Joel and Caroline Weber (2012). “Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy”. In: *International Tax and Public Finance* 19.1, pp. 25–53.

Snow, Arthur and Ronald S Warren (2005). “Ambiguity about audit probability, tax compliance, and taxpayer welfare”. In: *Economic Inquiry* 43.4, pp. 865–871.

Spicer, Michael W and Lee A Becker (1980). “Fiscal inequity and tax evasion: An experimental approach”. In: *National Tax Journal*, pp. 171–175.

Spicer, Michael W and Rodney E Hero (1985). "Tax evasion and heuristics: A research note". In: *Journal of Public Economics* 26.2, pp. 263–267.

Spicer, Michael W and J Everett Thomas (1982). "Audit probabilities and the tax evasion decision: An experimental approach". In: *journal of Economic Psychology* 2.3, pp. 241–245.

Srinivasan, T N (1973a). "Tax evasion: A model". In: *Journal of Public Economics* 2, pp. 339–346.

Srinivasan, Thirukodikaval N (1973b). "Tax evasion: A model". In: *Journal of public economics* 2.4, pp. 339–346.

Taber, Charles S and Richard J Timpone (1996). "Computational Modeling (Quantitative Applications in the Social Sciences)". In:

Taleb, Nassim Nicholas (2007). *The black swan: The impact of the highly improbable*. New York Random house.

Torgler, Benno (2002). "Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments". In: *Journal of Economic Surveys* 16.5, pp. 657–683.

Trotin, Gwenola (2010). *Tax evasion decision under cumulative prospect theory*. Tech. rep. Working Paper, EQUIPPE, Universit Charles-de-Gaulle Lille 3 and GREQAM-IDEP, Universit de la Mditerrane.

Tversky, Amos and Daniel Kahneman (1973). "Availability: A heuristic for judging frequency and probability". In: *Cognitive psychology* 5.2, pp. 207–232.

— (1985). "The framing of decisions and the psychology of choice". In: *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*. Springer, pp. 107–129.

— (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and uncertainty* 5.4, pp. 297–323.

Ungemach, Christoph, Nick Chater, and Neil Stewart (2009). "Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)?" In: *Psychological Science* 20.4, pp. 473–479.

Viscusi, W and Jahn K Hakes (2008). "Risk beliefs and smoking behavior". In: *Economic Inquiry* 46.1, pp. 45–59.

Viscusi, W Kip (1992). *Smoking: Making the risky decision*. Oxford University Press on Demand.

Watrin, Christoph and Robert Ullmann (2008). "Comparing direct and indirect taxation: The influence of framing on tax compliance". In: *The European Journal of Comparative Economics* 5.1, p. 33.

Weber, Elke U (2006). "Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet)". In: *Climatic Change* 77.1-2, pp. 103–120.

Weber, Elke U, Sharoni Shafir, and Ann-Renee Blais (2004). "Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation." In: *Psychological review* 111.2, p. 430.

Weber, Elke U, Ulf Böckenholt, Denis J Hilton, and Brian Wallace (1993). "Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19.5, pp. 1151–1164.

Webley, Paul (1987). "Audit probabilities and tax evasion in a business simulation". In: *Economics Letters* 25.3, pp. 267–270.

— (1991). *Tax evasion: An experimental approach*. Cambridge University Press.

Winsberg, Eric (2009). “A tale of two methods”. In: *Synthese* 169.3, pp. 575–592.

Wulff, Dirk U, Thomas T Hills, and Ralph Hertwig (2015). “Online product reviews and the description–experience gap”. In: *Journal of Behavioral Decision Making* 28.3, pp. 214–223.

Yechiam, Eldad, Greg Barron, and Ido Erev (2005). “The role of personal experience in contributing to different patterns of response to rare terrorist attacks”. In: *Journal of Conflict Resolution* 49.3, pp. 430–439.

Yechiam, Eldad, Ido Erev, and Greg Barron (2006). “The effect of experience on using a safety device”. In: *Safety science* 44.6, pp. 515–522.

Yitzhaki, Shlomo (1974). “Income tax evasion: A theoretical analysis”. In: *Journal of public economics* 3.2, pp. 201–202.

Zaklan, Georg, Francesco WS Lima, and Frank Westerhoff (2008). “Controlling tax evasion fluctuations”. In: *Physica A: Statistical Mechanics and its Applications* 387.23, pp. 5857–5861.

Zaklan, Georg, Frank Westerhoff, and Dietrich Stauffer (2009). “Analysing tax evasion dynamics via the Ising model”. In: *Journal of Economic Interaction and Coordination* 4.1, pp. 1–14.

Zohar, Dov and Ido Erev (2006). “On the difficulty of promoting workers’ safety behaviour: overcoming the underweighting of routine risks”. In: *International Journal of Risk Assessment and Management* 7.2, pp. 122–136.