Università degli Studi di Cagliari

DOTTORATO DI RICERCA

in Economia

Ciclo XXVII

Three Essays on the Generalizability of Experimental Results in Economics

Settore/i scientifico disciplinari di afferenza

SECS-P/01

Presentata da: **Tiziana Medda**

Coordinatore Dottorato Prof. Romano Piras

Tutor Prof. Vittorio Pelligra

Esame finale anno accademico 2014 – 2015

REGIONE AUTONOMA DELLA SARDEGNA

"La presente tesi è stata prodotta durante la frequenza del corso di dottorato in Economia dell'Università degli Studi di Cagliari, a.a. 2011/2015 - XXVII ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1 "Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell'ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali".

"Tiziana Medda gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.)".

Abstract

Although few economists today dismiss the use of the laboratory experiments, it would be a mistake to think that experimental methodology no longer represents a controversial issue in economics. One of the major criticisms is represented by the external validity of experimental data and concerns the transferability of results obtained in laboratory to the real world. The aim of this thesis is to tackle the issue of external validity focusing in particular on one aspect: the possible lack of representativeness of standard subjects pools usually used in economic research. The first experimental study compares the choices of undergraduates and subjects representative of population in different treatments and with different reward dimensions by exploiting the experimental design used by Pelligra and Stanca (2013) to investigate social preferences in a field experiment. Our results show that two samples follow a common behavioral pattern with the only exception of a significant difference in choices where self-interest may play a prominent role. In the second study we use a between-subjects design to compare the behavior of experienced and inexperienced subjects. We investigate whether the laboratory experience, built through repeated participation in experimental sessions, biases subjects' behavior in a set of representative simple games used to study social preferences. Our main finding shows how subjects having a high level of experience in lab experiments do not behave in a significantly different way from novices.

Contents

List of Figures

List of Tables

Acknowledgments

It is a pleasure for me to thank the many people who made this thesis possible. During the last four years, I have come across so many people who have supported me in different ways, so I want to thank them personally. Firstly, I want to thank my supervisor Vittorio Pelligra for the help in the preparation of this thesis. I would also like to thank Tommaso Reggiani and Luca Frigau for their precious support in the development of chapters 2 and 3.

I thank Prof. Enrique Fatas for his encouragement and stimulus in the early stages of the experimental project presented in Chapter 3; the staff of the CLER Lab of the University of Cologne for research assistantship; Andrea Isoni, Ro'i Zultan and the all participants at the "Methods" parallel session of 2015 ESA European Conference (Heidelberg), for their comments and suggestions.

I also want to show my gratitude to the Coordinator of the PhD Program, Prof. Romano Piras, for taking his precious time to answer my infinite questions and for his innate patience. I would also like to thank all participants at the internal seminars at the University of Cagliari for their suggestions and advices.

I am deeply grateful to Prof. Carla Massidda who showed me love for teaching and passion for research; her guidance as well as her academic experience, have been invaluable to me.

I would like to thank all Ph.D. colleagues that I have met during these years, with whom I shared pains and hopes: Massimo, Stefania, Alejandra, Alberto and mostly Elisa and Vania for their constant support. I want to thank the incredible staff of Economics Library, Tamara, Rosi and Gianfranco for their help and encouragement in all these years.

All my gratitude to the friends that have been there from the beginning and to those that I met during this journey. Without you, all this would have no sense.

I want to thank my dear friend, Emike, who tried to make this thesis readable.

Finally, my family. My sister, for being there for me, no matter what, and my parents for their unconditional love.

Summary

"*An honest skeptic then has the burden of stating what is different about the outside world that might change results obtained in the laboratory*"

(Friedman and Sunder, 1994)

The parallelism or external validity represents the most important methodological issue related to experimental economics and concerns the transferability of experimental results obtained in laboratory to the real world. Smith defines parallelism as: "propositions about the behaviour of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to non-laboratory microeconomies where similar ceteris paribus conditions hold defines a sufficient condition for the transferability of results from laboratory to field environments" (1982, p. 936). Given limitations of econometrics and field experiments, able to interpret the data only in terms of assumptions about preferences and subjects' behavior, the author describes the experimental methodology necessary to control these variables within a microeconomic system created in the lab and to observe and measure the message responses of agents and the outcomes resulting from these messages. However, despite Smith's confidence in parallelism, skepticism about the external validity of laboratory data has represented an obstacle to the acceptance of experimental economics as a "true" scientific discipline. The aim of this thesis is to tackle the issue of external validity focusing in particular on one aspect: the possible lack of representativeness of standard subjects sample usually used in the laboratory.

In the first Chapter, we survey the existing literature on all issues related to this topic. We make a critical review of principle papers dealing with the issues linked to the generalizability of experimental data obtained in the laboratory and concerning: the use of unrepresentative subjects' samples, the artificiality of the laboratory tasks, the provision of inadequate incentives and, the lack of sufficient opportunities to learn by experience for participants.

In the second Chapter, we investigate about the use of standard samples of college students in economic experiments.

The use of convenient pool of college students represents one of the most common criticisms related to the generalizability of laboratory data. Replicating in the lab a simple experiment originally run with a representative sample of the population, we show that, despite the differences between college students and subjects of the general population in terms of demographics, cognitive skills and personalities, the observed behavior of the two samples follows a common pattern in a set of binary dictator games focusing on several motives for altruistic behavior. We compare the choices of undergraduates and subjects representative of population in different treatments and with different reward dimensions by exploiting the experimental design used by Pelligra and Stanca (2013) to investigate social preferences in a field experiment. Our results show that two samples follow a common behavioral pattern with the only exception of a significant difference in choices where self-interest may play a prominent role. This gap seems to be related mainly to the academic background of the participants: our sample of undergraduates economics students differ in their degree of self-interested choices both from the representative group of the population and from its sub-sample of students.

In the third Chapter, we focus on another possible problem related to the representativeness of subjects usually considered in the experimental economic studies.

The ever-increasing number of experiments and the prevalent location of research centres in university campuses produced a peculiar category of participants: students with high level of laboratory experience built through repeated participation in experimental sessions. We investigate whether the experience accumulated in this way biases subjects' behavior in a set of representative simple games used to study social preferences (Dictator Game, Ultimatum Game, Trust Game, and Prisoner's Dilemma Game). Our main finding shows how subjects having a high level of experience in lab experiments do not behave in a significantly different way from novices.

Chapter 1

Parallelism: a Critical Review

"*In less than three decades economics has been transformed from a discipline where laboratory experimentation was considered impossible, useless, or at any rate largely irrelevant, into a science where some of the most exciting discoveries and developments are driven by experimental data.*"

(Guala 2008, p.2).

"*Ironically most objections raise questions that can be very well analysed with lab experiments, suggesting the wisdom of conducting more lab experiments, not fewer*"

(Falk and Heckman 2009, p.537).

1.1 Introduction

Although few economists today dismiss the use of the laboratory experiments, it would be a mistake to think that experimental methodology no longer represents a controversial issue in economics. One of the major criticisms is represented by the external validity of experimental data and concerns the transferability of results obtained in laboratory to the real world.

The aim of this chapter is to make a critical review of some very influential studies dealing with the main issues linked to the generalizability of experimental data obtained in the laboratory.

I shall discuss how experimentalists usually respond to the basic criticism that experiments do not say much about the 'real world', and more specifically about the use of unrepresentative subject pools and artificial tasks, the provision of inadequate incentives and the lack of sufficient opportunity to learn by experience.

Section 1.2 focuses on the main steps that led to the introduction of experimental method in economics. Section 1.3 presents the general problem of external validity. Sections from 1.4 to 1.7 discuss the most prominent problems related to the lack of generalizability of lab findings and specifically:

a) the unrepresentativeness of subject pool;
b) the artificiality of the laboratory tasks;
c) the inadequateness of incentives;
d) the lack of sufficient opportunities to learn by experience,

Section 8 sums up and concludes.

1.2 Experimental and Non-experimental methodology in Economics

During the twentieth century, many social science scholars were unanimous in considering economics as a theoretical science and experimental methodology as impractical in this field: economic research had to be carried out far from laboratories, whilst the experimental methodology was considered of exclusive domain of hard sciences.

Skepticism about use of experimental method in economics represented an obstacle to the acceptance of experimental economics as a "true" scientific discipline inasmuch as most scholars did believe (and some still do), that lab experiments could not provide very meaningful data for the economic issues (Starmer, 2006). There is no wonder that the first lab experiments in economics were run only in the late 1940s and that fewer than 10 experimental papers were published before 1965 (Falk and Heckman, 2009).

All the great scientists, from Galileo onwards, took care of defining how experiments had to be performed in the laboratory. Hard scientists, having as main aims of their research, description and prediction of phenomena, cannot disregard from the use of controlled experiments for their studies; lab experiments have the advantage of succeeding to control conditions more than in any other context. Social sciences, such as physical sciences, need to test their hypotheses and theories.

The first discussions about the use of experimental method in economics were characterized by a clear distinction between hard and social sciences. According to J. S. Mill: "The physical sciences are those which treat of the laws of matter, and of all complex phenomena in so far as dependent upon the laws of matter. The mental or moral sciences are those which treat of the laws of mind, and of all complex phenomena in so far as dependent upon the laws of mind" (1844, p. 317).

For the economist and philosopher, the laws of mind and laws of matter, cannot share the same routine analysis and method, and mostly, if on one hand the laws of matter are clearly unsuitable to guide social sciences, on the other hand, the same economic science really does not need it, taking advantage from 'introspection' (Guala, 2006) for understanding of own phenomena. Until that moment, both hard and social scientists did agree that "less harm is likely to be done by emphasising the

differences between the social and the natural sciences than by emphasising their similarities" (Robbins 1932, p. 112).

So, for all the XX century, economists of main Economic Schools, theoretically also conflicting from each other, have shared Mill's and Robbins' thought, by banishing any possible openness towards the experimental method. A pessimistic view about application of experimental method on social science was also expressed by Milton Friedman, one of the most influential economists of the second half of the 20th century. According to the author "unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences [….]. The inability to conduct so-called 'controlled experiments' does not, in my view, reflect a basic difference between the social and physical sciences both because it is not peculiar to the social sciences - witness astronomy - and because the distinction between a controlled experiment and uncontrolled experience is at best one of degree. No experiment can be completely controlled, and every experience is partly controlled, in the sense that some disturbing influences are relatively constant in the course of it" (1953, p. 101).

Lipsey perfectly sums up the mistrust about experimental method shared by previous academics: "it is rarely, if ever, possible to conduct controlled experiments with the economy. Thus economics must be a non-laboratory science" (1979, p. 39).

All these statements clearly highlight how the experimental methodology in economics is been deeply influenced by the way in which the same discipline was considered by the scientific community.

Initially, the necessity to fill the gap between the world of the abstracted theory and the data gathered in the real world was solved by incorporating some additional stochastic variables (Hey, 1991). In that way the theory becomes "a combination of the original economic theory and some assumptions about the nature of the additional stochastic variables; so, if 'the theory' survives the test, it could be because both the original economic theory and the assumptions about the stochastic variables are correct, or because both the original economic theory and the assumptions about the stochastic variables are incorrect" (Hey 1991, p. 8), in both cases the inference is inherently ambiguous.

Even though it is possible to identify some experimental or proto-experimental economic works since 1930, only in the 1990, Charles Plott, one of the pioneer of experimental economics, during the 61° annual meeting of the Southern Economic Association, posed the question about the use of experimental method in economics and positive about the answer wrote: "Economics is one of the few sciences that is fortunate to have both the field and the laboratory with which to work. [..] The

laboratory methodology, which has historically been absent, will grow and become an important partner in a joint effort to isolate the principles which govern economic behavior" (1991, p. 918).

The experimental economists generally pinpoint as first step towards the introduction of the experimental methodology in economics, the conference held in Santa Monica in 1952, but the lack of continuity in the research has made difficult the identification, for a long time, of a unified academic program. The 80s have seen a huge growth in the use of experimental method in economics but from an epistemological point of view, the Nobel Prize for Economic Science in 2002 to Vernon L. Smith has represented the true breakthrough point in the acknowledgement of economics as experimental discipline to all intents and purposes.

The economic science was protagonist of one of the greatest revolution from one methodological point of view. As highlighted in the presentation speech delivered by the Norwegian Committee during the awarding of the Nobel for the Economics in 2002: "Until recently, economics was widely regarded as a non-experimental science that had to rely on observation of real-world economies rather than controlled laboratory experiments. [….] Moreover, today's research increasingly relies on new data from laboratory experiments rather than on more traditional field data, that is, data obtained from observations of real economies" (Nobel Press, 2002:1).

1.3 The external validity problem. What is it that makes the experimental lab different from the real world?

The acceptance of the economics as experimental science, however, it has not been an obstacle-free process. Many critics have spoken out against the use of experiments adducing several reasons related to the lack of external validity of laboratory findings.

The parallelism or external validity is one of the most important methodological issues related to experimental economics and concerns the "ability to generalize from the research context to the settings that the research is intended to approximate" (Loewenstein, 1999: F26). Although, according to the author this methodological issue represents the 'Achilles Heel' of experimental economists, the majority "believe that in the certain features of their experiments, such as the incorporation of market institutions, the stationary replication, and carefully controlled incentives, make their experiments immune to the problem of external validity […]" (*ibidem*).

Among them we can include the father of Experimental Economics, V. L. Smith, who defines parallelism as: "propositions about the behaviour of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to non-laboratory microeconomies where

similar ceteris paribus conditions hold defines a sufficient condition for the transferability of results from laboratory to field environments" (1982, p. 936).

In agreement with Smith, since the abstractions of the laboratory are orders of magnitude smaller than those of economic theory, then it is possible to falsify any theory. If the aim of the experimentalist is to test hypotheses resulting from theory, then it is sufficient to respect four key precepts – non-satiation, saliency, dominance and privacy – to guarantee the generalizability of lab data. Inside the laboratory, saliency and non-satiation conditions will ensure the existence of microeconomies like 'real live economic systems', while, dominance and privacy conditions will permit to perform a working controlled laboratory experiment.

Respecting these precepts, relevant to the internal validity[1], experimentalists will be able to explore lab microeconomies in which real agents give-and-take messages through real institutions. It means that, similar conditions (ceteris paribus) which correspond to well-defined assumptions in lab and non-lab microeconomies, represent a sufficient requirement to draw valid inference from laboratory experiments, thus, as Smith points out, generalizing Shapley's parallelism definition, "as far as we can tell, the same physical laws prevail everywhere[2] (1964, p. 67).

The limitations that we can find in the econometrics and in the field experiments, can be overcome with the experimental methodology: a microeconomic system created in the lab is able to observe and measure the message responses of agents and the outcomes from these messages (Smith, 1982). However, despite Smith's 'confidence' in parallelism, arguments against the generalizability of results obtained in laboratory continue to have strong resonance.

As stressed by the Philosopher of Science Francesco Guala (2005), to write about the external validity is really challenging and so, whilst some experimentalists have intentionally avoided to address this methodological problem, other have proposed a set of rules able to guarantee a good generalizability of lab findings. But to face adequately the external validity issue, we need to consider the different purposes of experiments.

Guala and Mittone (2005) have examined that aspect by considering the three main objectives of experimental research identified in the Roth's taxonomy (1986): 'Speaking to theorists' which aim to test hypotheses from formally specified models and theories; 'Searching for facts' which aim to investigate phenomena or empirical regularities that cannot be explained by existing theoretical models and, 'Whispering in the ears of princes' which focus on illuminating or supporting policy-making process.

[1] Internal validity refers to the ability to draw confident causal conclusions from one's research (Loewenstein, 1999: F26).
[2] Smith's quote in Microeconomic Systems as an Experimental Science (1982).

As far as the first category, 'Speaking to theorists', is concerned, experiments can help to repel theories which do not succeed to survive in a simple experimental situation; a theory that cannot be completely specified, must be replaced. According to Plott: "General theories must apply to simple special cases. [….]. General models, such as those applied to the very complicated economies found in the wild, must apply to simple special cases. Models that do not apply to the simple special cases are not general and thus cannot be viewed as such" (1991, pp. 902-905).

Concerning the second element in the taxonomy, 'Whispering into the Ears of Princes', specific target of research should be shaped in order to fit the experimental prototype in the laboratory. However the economists usually are not able to shape the world as they want, but they may follow the opposite strategy: by starting from the experimental system and designing it to resemble to the target as much as possible (Guala and Mittone, 2005).

The last category identified by Roth, 'Searching for Facts', refers to experiments designed to collect data on interesting phenomena in the hope of identifying some pattern or empirical regularity. Even if from an ideal point of view all real phenomena should be explained through a theory, it may happen that these remain detached from any general principle. Several phenomena discovered in artificial environments are now widely discussed in the economic literature, such as violations of rationality.

However, transferring a phenomenon to the real-world requires very detailed knowledge of the domain of application, so the generalizability, it is possible only up to a point. As suggested by Guala and Mittone "the relevance of results may be indirect, and it is unreasonable to impose the requirement that the experimental validity of each single experiment be proven rigorously. In many cases, experimenters contribute to the 'library' of phenomena that the applied scientist will borrow and exploit opportunistically on a case-by-case basis" (2005, p. 15).

Moreover, the external validity of lab findings is strictly linked to the internal validity. The experimental validity of lab findings implicates a trade-off between two dimensions, internal and external: "The stronger an experimental design is with respect to one validity issue, the weaker it is likely to be with respect to the other" (Guala 2005, p. 144). For instance, a greater artificiality of the laboratory environment facilitates the internal validity of results but at the same time, it negatively strikes the external validity. Experimentalists are thus called to make a choice between these two dimensions: the simpler the experimental environment, the easier it is to identify the cause(s) responsible for a given phenomenon or effect and at the same time it distances it from real-life world.

In addition to the artificiality of lab environment, there are other differences between experimental and real-life conditions which may lead to the lack of validity of findings. The general criticism for which experiments do not say much about the 'real world' can be addressed focusing on four specific critiques: 1. the use of unrepresentative subject pools; 2. the use of artificial tasks; 3. the provision of inadequate incentives; 4. the lack of opportunity to learn by experience.

However, as stressed by Guala: "That an experimental and a real-world system should differ in some respects is just inevitable and rather uninteresting by itself; the interesting question is whether the differences are casually relevant" (2005, p. 157).

Until now, one of the best ways to investigate about these differences was to consider, case by case, each single question, by comparing the results obtained both in the laboratory and in the field performing the same experimental protocol.

1.4 Unrepresentativeness of subjects pools

The most common criticism about the lack of generalizability of lab findings concerns the representativeness of subjects' pools.

Until now three potential biases were largely explored: the **self-selection problem of participants,** the use of **college students** and, the **use of Educated people coming from Western Industrialized, Rich and Democratic world** (**WEIRD** acronym). Moreover, a relatively new potential source of bias linked to the representativeness of subjects' pool arises from the combined effect of 'location' and 'number': given the almost exclusive location of labs in university campuses and the ever-increasing number of experiments run in each of these labs, **subjects tend to accumulate experience through repeated participation in multiple experiments**.

Usually subjects who participate to experiments are university students. The massive presence of this category in research centres is confirmed by an analysis made on sixty papers published between 2001 and 2002 in the major journals of experimental economics[3]: 54 of 60 papers published had as protagonists university students (Danielson and Holm, 2007). This category certainly differs from other samples of population for demographics[4] [5], such as age, experience, status, social class and it diverges

[3] Experimental Economics, Games and Economic Behaviour and Journal of Economic Behavior and Organization.

[4] Guillén and Veszteg (2006), by analysing 8,755 observations coming from a laboratory (597 experimental sessions, 74 incentive based economic experiments recorded over more than 2 years), find out that demographic differences can explain only 4% of the variations on dependent variable considered (payments received). Carbone (2005) analyses demographic characteristics in relation with subjects' strategic behavior. She uses a unique experimental subject pool that participated in a life-cycle consumption experiment and finds that demographics have no effect on observed behavior.

[5] See Calder et al., 1981; Greenberg, 1987; Kraus, 1995.

for cognitive skills and they tend to be more homogeneous. Undergraduates show "unfinished" personalities (Carlson, 1971), less-crystallized attitudes and less-formulated senses of self, more unstable peer group relationships, stronger tendencies to comply with authority and stronger cognitive skills than to older adults (Sears, 1986).

In his second-order meta-analysis, Peterson (2001) has observed that students' answers were slightly more homogeneous than those of non-students, and the effect sizes differed for direction and magnitude with respect to those of non-students. Considering that, as stated by Schultz: "[…] approximately 80% of our research is performed on the 3% of the population currently enrolled in college" (1969, p. 218), a deep reflection is pretty necessary.

Moreover, "punctual college sophomore" volunteers are usually interested in research, quite willing to collaborate with the experimenter and in pursuit of social approval (Rosenthal and Rosnow, 1969). For the volunteer subject, it is important to feel himself as a "good subject": participant will be concerned about what their choices tell about them and, he/she could be tempted to validate the experimental hypothesis (Orne, 1969).

Rosenthal and Rosnow (1997) have identified and ranked 17 characteristics of volunteers. According to their ranking, volunteers tend to be: better educated and more intelligent, belonging to higher social class, more social approval-seeking and more sociable than non-volunteers; moreover, they have a tendency to be more arousal-seeking, unconventional, females, non-authoritarians, Jewish or Protestant and nonconforming and finally, they are likely to come from smaller towns, interested in religion, altruistic, self-disclosing, maladjusted, and younger with respect to non-volunteers.

All these elements can determine the first two potential biases in the experimental results related, first, to the unrepresentativeness of students with respect to the general population and, secondly, to the problem of self-selection. Even though the element of voluntariness needs to be clearly extended to all categories of subjects involved in research, there is a strong tie-up with figure of student and, for that reason self-selection and representativeness of subjects have been often addressed empirically together.

1.4.1 The self-selection problem of participants

Eckel and Grossman (2000), have examined the effect of recruitment method on behavior in a dictator game experiment, comparing results obtained with volunteers and pseudo volunteers in class (subjects recruited on the spot) and they find significant differences in the behavior between two samples. Their data show that pseudo volunteers are more generous on average than their volunteer

counterparts, and that non-monetary factors such as religion or altruistic preferences have a greater effect on the giving behavior of pseudo volunteers.

A different procedure to evaluate the selection bias issue was implemented by Cleave et al. (2010); authors conducted a classroom experiment with a population of 1,173 students, using a trust game and a lottery choice task to measure individual preferences. In a second time, all 1,173 students were invited to participate in a laboratory experiment. To calculate the extent and direction of selection bias, authors compared students' choices who decided to participate also in the lab experiment with those who did not. Their results show that social and risk preferences of students participating in the laboratory experiment are not significantly different from those belonging to student population.

The 'self-selection problem' has also been documented by Anderson et al. (2013); they examine how the social preferences measured in a laboratory experiment vary across three different samples of experimental subjects: undergraduate 'self-selected' students for the lab experiment; non-students sample recruited with similar procedure used for recruiting students; non-students sample recruited using a procedure that allowed very little self-selection. They find that self-selection does not distort the measuring of other-regarding preferences and the share of individuals exhibiting other-regarding concerns, is remarkably smaller among college students, even after controlling for observable differences in socio-demographic characteristics among subject pools. Their results also reject the more specific hypothesis that approval-seeking subjects are the ones most likely to self-select into experiments and they also show a large difference between self-selected college students and self-selected adults: the students appear considerably less pro-social.

Focusing on social preferences topic, Falk et al. (2013) have investigated whether laboratory experiments with student samples distort the importance of the same. In two different studies, authors have examined the problem of 'self-selection' and 'students' representativeness'. As far as it concerns the first one, by comparing the behavior of participating-students and non-participating students, they find out that with respect to their prosocial inclination, the former are not significantly different from non-participating-students.

Also Exadaktylos et al. (2013) examined issues related to students in laboratory whom voluntarily choose to participate focusing their attention on social preferences. In their study they employed data from a survey-experiment conducted with a representative sample in five experimental decisions in three canonical games: dictator, ultimatum and trust games. Their results confirm that self-selected students represent an appropriate subject pool for the study of social behavior.

See **Tables 1.1-1.2** for a summary of the experimental studies above cited.

1.4.2 College students

The concerns about the potential biases steaming from use of college students as standard sample in lab are fuelled by much evidence: really college students tend to behave differently with respect to non-students.

Gordon et al. (1986) analyse thirty-two studies which see groups of students and groups formed by non-students run identical protocols. In twelve studies[6], statistical tests of between-group differences show strong divergences between two samples; by contrast, authors find no big differences in the studies which do not employ statistical comparison. Authors suggest three ideas to improve external validity of lab findings: 1) using trained experimental subjects; 2) running clinical debriefings to reveal subjects' perceptions, understandings, and meanings associated with the research situations (Adair, 1984); 3) to employ subjects with demographic and interest profiles similar to the non-students to whom researchers wish to generalize. They conclude that the 'gatekeepers' of science should discourage the use of students unless they do not look like to the sample of interest.

More recent experimental studies have shown systematic differences between the behavior of college students and others specific categories.

Cooper et al. (1999) have compared the 'standard' experimental subject population (students) with an 'expert' subject population (Chinese managers and white-collar workers in state enterprises) addressing several open questions about the development of the ratchet effect in centrally planned economies and, more generally, about the evolution of strategic play in games. Contrary to the authors' expectations, students in their role as firms exhibit significantly higher initial levels of strategic play than older, more experienced managers; this surprising result appears due to the age and/or lower educational levels of older managers.

The suspicious that professionals' behavior may differ from non-professionals[7]' due to training, regulation, etc. led to Haigh and List (2005) to use a specific pool of professional traders recruited from the Chicago Board of Trade to investigate about the 'myopic loss aversion'; previous experimental studies[8] suggested that undergraduate students' behavior is consistent with this "myopic loss aversion" conjecture. Unexpectedly, authors find that professional traders exhibit myopic loss aversion to a greater extent than undergraduate students.

---

[6] Allen and Muchinsky (1984); Christensen-Szalanski et al. (1983); Churchill and Cooper (1971); Cornelius et al. (1984); Fleming (1969); Hakel et al. (1970a); Hakel et al. (1970b); Jago and Vroom (1982); Kavanagh (1975); Moskowitz (1971); Schneider (1982); Stow and Ross (1980).

[7] For a complete review about differences between professionals and non-professionals see G. R. Fréchette (2015).

[8] Thaler et al. (1997) and Gneezy and Potters (1997) have verified the existence of myopic loss aversion among undergraduate students.

A relevant finding about the difference between professionals and non-professionals (read students) is presented in an important work realized by Fehr and List (2004). They investigate experimentally how CEOs respond to incentives and how they provide the same in situations requiring trust and trustworthiness and, by comparing their behavior with the behavior of students, find that former are more trusting and show more trustworthiness than students. Moreover, most CEOs and students use the punishment threat, even though CEOs less.

In a social framing study, Carpenter et al. (2005) compare experiments conducted with the standard social framing, that is, undergraduates at small liberal arts college in Vermont (Middelbury College) and students of a junior college in Kansas City, to field experiments using Kansas City workers. They find that proposers in the Ultimatum Game in the two experiments in Kansas City make more generous offers than students at the Middlebury College, even controlling for demographic differences (it confirms the regional differences hypothesis); moreover their results show that Kansas City students offer significantly more than both Kansas City workers and Middlebury college students, respectively, whilst in the Dictator Game, are the workers the more generous among all samples. Furthermore, they find that both samples of students show a drop in the average allocations between two games, while the workers offer the same amount, on mean, in both games.

Bellemare and Kröger (2007) instead, explore behavior observed in an investment game between a representative sample of Dutch population and students' pool in the laboratory. Lab participants provide a lower bound on the levels of investments and of amounts returned with respect to the level estimated in the representative sample. Their results indicate that trust, trustworthiness and other social preferences, once identified in the lab, are likely to be present, with greater intensity, in the population as a whole.

Carpenter et al. (2008) run a Dictator Game in which students and random members of the community have to choose which charity to support and how much to donate. There are systematic differences between the choices of two groups: community members choose their own charity and donate significantly more (32% of them give the entire endowment). They conclude that students' sample does not appear representative in the specific context of the charitable giving and the Dictator Game.

Considering a wide set of standard experimental games (Dictator Game, Trust Game, Public Good Game, Beauty Contest and Second-price Auction), Belot et al. (2010) find out that, in general, students are more likely to behave as homo-economicus agents than non-students in games involving

other-regarding preferences, while there is not a significant difference between samples in the games that do not engage them (Beauty-contest and Second-price Auction).

Falk et al. (2010) in their second study, compare behavior of students and the general population using the same experimental protocol (variant of a trust game); their results show that first movers display a similar behavioral pattern, whilst, in terms of reciprocation (second movers player) higher investments are reciprocated with high repayments in both groups, but the level of reciprocation is lower for students than for non-students.

Cappelen et al. (2015) address the question whether lab experiments on student populations are useful to identify the motivational forces present in society at large by comparing the behavior of a nationally representative population with different student populations into the lab. Their results show that students may not be informative of the role of social preferences in the broader population: representative participants differ fundamentally from students both in their level of selfishness and in the relative importance assigned to different moral motives.

See **Tables 1.3-1.5** for a summary of the experimental studies above cited.

1.4.3 The WEIRD people

Henrich et al. (2010) have extensively explored the third possible source of lack of representativeness of samples. Specifically, authors wonder to what extent it is correct to assume a species-level generality of research findings considering that experimental samples usually come from western, educated, industrialized, rich and democratic societies (named with WEIRD acronym in their work). From an analysis carried out by Arnett (2008) between years 2003 and 2007 on six psychological sub-disciplines, is came to light that 68% of participants come from USA and the 96% from Western industrialized countries, particularly from North America, Europe, Australia and Israel, that represent only 12% of the world population.

The authors try to answer the important question whether the researchers are justified in assuming a species-level generality for their findings that regard, almost exclusively Western, Educated, Industrialized, Rich, and Democratic (WEIRD) people. They present a meta-analysis involving large-scale comparative experimentation on important psychological or behavioral variables and discuss four main contrasts: 1) differences between populations drawn from industrialized and small-scale societies in some basic psychological domains (visual perception, fairness and cooperation in economic decision-making, folkbiological reasoning, spatial cognition and potential differences); 2) Western with non-Western populations about four of the most studied domains (social decision making: fairness, cooperation, and punishment), independent versus interdependent self-concepts (and associated motivations), analytic versus holistic reasoning, and moral reasoning; 3) contemporary American versus the rest of the west population about individualism and related psychological phenomena; 4) typical contemporary American subjects versus other Americans (considering education for comparisons among contemporary adult Americans; spatial reasoning for comparisons subpopulations of American children; and psychological aspects for comparison between contemporary Americans and previous generations).

Their meta-analysis shows that members of WEIRD societies, including young children, are among the least representative populations one can find for generalizing about humans. Many of these findings involve domains that are associated with fundamental aspects of psychology, motivation, and behavior – hence, there are no findings a priori for claiming that a particular behavioral phenomenon is universal based on sampling from a single subpopulation.

Overall, all these comparative measures describe industrialized populations as outliers, and among them, Westerners appear far from "standard subjects" for several key dimensions, such as allocentric spatial reasoning and antisocial punishment. Moreover, among Westerners, Americans

diverge for psychological or behavioral measures. As stressed by Henrich et al. (2005), people are endowed with cultural learning capacities that allow them to acquire beliefs and preferences appropriate for the local social environment. As consequence of adaptive learning processes, different societies will tend to arrive at different social equilibria and will express differently preferences and beliefs; each person will lead in lab her/his preferences and beliefs acquired into decision-making situations in the real world. See **Table 1.6** for a summary of the study above cited.

1.4.4 Highly experienced subjects

The last and new concern linked to the unrepresentativeness of subjects usually recruited in research labs concerns the role of experience in experiments. Indeed, having a long history of participations in experimental sessions might alter subjects' behavior. The experience, built taking part repeatedly in experiments, could affect in some way, subjects' choice in lab.

This problem in economic experiments has been essentially neglected, with few notable exceptions: Benson and Faminov (1988) and Harrison et al. (1989) discussing IO experiments, Marwell and Ames (1980), Isaac et al. (1984) and Bolton (1991) in bargaining games experiments and more recently, Matthey and Regner (2013), Conte et al. (2014), Capraro and Cococcioni (2015) and Xue et al. (2015).

The first two papers document that highly experienced players are more capable at achieving tacit profitable collusion and they are more effective as monopolists than inexperienced ones, respectively. Harrison et al. (1985) use a design with simulated buyers and decreasing costs; monopolists with design experienced get an effectiveness monopoly value higher than value gained by inexperienced monopolists; moreover, experienced monopolists perform more homogeneously than inexperienced subjects. Benson and Faminov (1988) experimental design, explicitly considers 'experience' as a treatment variable. Two types of markets are run simultaneously: one type is composed by experienced participants (subjects with previous participations in similar but not identical markets) and the second one involves inexperienced subjects, all facing the same pricing decisions. Their results confirm that experience increases the likelihood of cooperation or tacit collusion and moreover suggest that the "experimentalists may be able to achieve equilibria without running the very large numbers of trials […]. [..] subjects assimilate their experience and can call upon it after a much longer time period" (Benson and Faminov 1988, p. 363).

Marwell and Ames (1980) and Isaac et al. (1984) both consider public good games situations and find no significant differences due to the different level of experience of participants.

The former study reports three replications of previous researches about the predictive utility of the free-rider hypothesis regarding the provision of public goods by groups. One of the three replications considers the role of experience and shows that experienced subjects do not behave very differently from inexperienced ones in that situation. Isaac et al. (1984) attempt to draw together the ideas of three previous works[9] about free-riding concept, by trying to reconcile divergences in their experimental

[9] Marwell and Ames (1980); Isaac, McCue and Plott (1982); Kim, Walker (1984).

results. As in the former experiment, in one treatment, authors compare inexperienced versus experienced subjects and no effect of experience is found.

Bolton (1991) finds a similar negative result in an experiment involving alternating-offer bargaining. His data shows no evidence that greater subjects experience determines pecuniary-equilibrium play and, with experience, the average of offers do not change, while standard errors decrease.

The more recent contribution of Matthey and Regner (2013) explore whether subjects' experience spills over between experiments. Their meta-analysis considers data from four different studies[10] involving allocation decisions in which the subjects' choices were combined with information about their past experiment participations stored in the online recruiting system used. They, more specifically, test the independence of history between experiments when there is no connection between experiments and the sessions are spread out over longer time spans. Their results show a negative correlation between the number of participations in lab sessions and generous behavior in the allocation situation, that is, subjects with a high number of participations tend to be less generous in allocation decisions. Therefore, giving decreases with the number of participations in experiments and so, the independence between experiments involving allocation decisions cannot be presumed if subjects participate in lab sessions repeatedly.

Conte et al. (2014) used, as Matthey and Regner (2013) the information on students' past participation in economic experiments stored in their database to analyse whether behavior in public goods games is affected by experience and history. In their study, the 'experience' measures the level of previous participation in social dilemma-type experiments, while 'history' denotes previous participations in experiments involving games different from social dilemmas. They find three main results: contributions by subjects and the expectation about others' contributions decrease with experience; a mixture model reveals that the proportion of unconditional co-operators decreases with experience, while that of selfish individuals increases and history also influences subjects' behavior although less than experience.

Focusing on cooperation in one-shot interaction, Capraro and Cococcioni (2015) study the history-dependent dynamic process. Many experimental researches suggest that the previous experience with economic games on cooperation and the intuition interact with each other such that experienced subjects are less cooperative than inexperienced subjects. They run a standard two-person Prisoner's Dilemma in which participants are randomly assigned to either of two conditions: 1) the

[10] Güth et al. (2009), Klempt and Pull (2009), Regner and Harth (2010), and Bracht and Regner (2011).

time pressure condition which measures intuitive cooperation; 2) the time delay condition which measures deliberate cooperation. Their main findings show that promoting intuition versus deliberation has no effect on cooperative behavior among inexperienced subjects living in a non-cooperative setting and that experienced subjects cooperate more than inexperienced subjects, but only under time pressure. These results suggest that cooperation is a learning process rather than an instinctive impulse or a self-controlled choice and that, experience operates primarily via the channel of intuition.

Finally, Xue et al. (2015) replicate and extend a simple riskless choice experiment reported by Hochman et al. (2014) as supporting loss aversion for money. One of five hypotheses concerns the possible role of experience with experiments. They test whether participants with greater experience in experiments will have higher maximization rates in prepayment treatments. They define "experienced" participants as those who had participated at least 10 times previously in experimental sessions, and "inexperienced" as those who had participated no more than 5 times and find that participants who have been in many economics experiments before do not choose differently than those who are relative novices. See **Tables 1.7-1.8** for a summary of the experimental studies above cited.

1.5 Artificial tasks

Another criticism to parallelism is related to the use of artificial tasks.

Smith identified two main components which determine a microeconomic system in the laboratory: 'environment' and 'institution'. The environment specifies each agent's preference (i.e. initiating circumstances, private tastes, and knowledge and skill endowments[11]), initial endowments and the technology; while the institution defines the language and the rules that can be used within the laboratory.

However, even though the economic environment is 'well defined' (i.e. rewards earned are significant enough to be pursued and participants have understood how the rewards are linked to the choice-options) the laboratory remains an artificial environment and decisional situations presented are generally abstract and unmatched in real life situations. For instance, as Starmer (1999) points out, in decision-making under risk experiments, participants are usual to make a series of one-off choices in unfamiliar context in which the only feedback about the success of choice strategy comes in the form of a payoff to a single task, randomly selected at the end of such experiments.

Furthermore, experimental tasks are clearly simpler than situations in the real world; for example, laboratory experiments often restrict the response mode to a single dimension, whereas real world settings almost always involve multiple response modes. Subjects in the experimental laboratory "try to make sense of the unfamiliar and incompletely defined experimental environment based on the instructions, cues and feedback they receive" (Zizzo 2010, p. 77) and address artificial restrictions on set of possible choices. This restriction on available choices can affect subjects' observed behavior.

Siakantaris (2000) argues that experimental economists, to achieve the best control over variables in order to exclude the effects of disturbing factors, are compelled to harshly specify the laboratory situation, raising what the author defines as the 'experimental economics trade-off': the better experimental economists do their job in controlling variables, the more they are threatened by a lack of parallelism and hence of the uselessness of their project.

The 'experimental economics trade-off' represents the trade-off between internal and external validity: the simpler the experimental environment, the easier it is to identify the cause(s) responsible for a given phenomenon or effect (Guala, 2005). In this way the experimenter can easily replicate the idealized assumptions of theoretical models and control variables. Indeed, according to Guala and Mittone (2005), adding realistic details to an experiment may imply increasing difficulties in the

[11] Smith (1982)

interpretation of the experimental results; the more artificial the laboratory environment is, better it is for the internal validity that benefits from greater abstraction and simplification.

However, as stressed by Schram (2005), "if the role of experiments shifts from testing theories to motivating the development of new theories, the 'mutual internal validity' of theory and experimental test has the danger of creating its own world" (ivi, p.234).

A clear example of artefact in lab is represented by standard dictator game; as explained clearly by Zizzo (2010) dictator games settings are highly artificial: participants are asked to give some amount of money, if they want, to strangers.

Situation like this are totally unlikely in the real world. In the everyday life few people or no one decide to give money at random strangers. Donations are usually made to family members, specific organisations or face-to-face to people requesting money, so dictator games results' can be misleading about the extent of faceless interpersonal altruism (Bardsley, 2008). Eckel and Grossman (1996) find that there is a significant increase in donations when we increase the extent to which a donation goes to a recipient generally agreed as "deserving." Hoffman et al. (1996) and Bohnet and Frey (1999) show that even if the dictator is not identified officially, seeing one's counterpart and getting information about the counterpart increases the possibility that a subject dictator will engage in positive reciprocity. Similar evidence are found when it is available a picture of the recipient (Burnham, 2003), other information are provided on recipients (Brañas-Garza, 2006) and there are visual suggestions of observation (Haley and Fessler, 2005). Specifically, Bardsley (2008) investigates whether dictator game can be considered a good tool to measure altruism, or whether it should be considered an artefact of experimentation. Author compares a standard dictator game to taking game (i.e. a dictator game with a modified set of options that include the opportunity to take instead to give or not something to the counterpart). His results show that a simple manipulation of the action set leads to drastic changes in behavior. Most of subjects' generosity appears to be reversible when there is sufficient opportunity to take[12]. As suggested by Guala and Mittone (2010), the Dictator Game experimental design "on its own, is probably too unusual and too abstract to trigger any real-life normative behavior. Indeed, the variability observed in experiments is probably due to the fact that all sort of norms (even conflicting ones) can be triggered by adding small cues to the basic design" (ivi, p. 584). The attention for the dictator games' usefulness should be shifted "away from the testing of preference theories, and towards theories of social norms" (*ibidem*).

[12] List (2007) has subsequently replicated the result that willingness to give erodes with taking options.

1.6 Inadequateness of incentives

The debate on monetary incentives in experiments figures among issues related to the external validity of research findings. The inadequateness of incentives in experiments may also affect subjects' behavior during experimental sessions and therefore making invalid data outside the lab.

'Actual choice versus hypothetical choice' has become one of the characteristic that distinguishes experiments published in economics journals from those in psychology journals. "Whereas […] economists almost always pay participants according to clearly defined performance criteria; psychologists usually pay a flat fee or grant a fixed amount of course credit" (Hertwig and Ortmann 2001, p. 384).

One of the first experiments on individual choice, realized by L.L. Thurston in 1931, was harshly criticized by W. Allen Wallis and Milton Friedman (1942) in relation to the use of hypothetical choices to experimentally elicit individual indifference curves. According to two authors, just actual stimuli can determine actual reactions by participants, whilst hypothetical stimuli clearly are not able to satisfy this goal (Roth, 1993).

Several economists[13] are in agreement that financial incentive makes clear the goal that has to be performed. There exists an interaction between motivations and incentives that determines agents' behavior, but while people's motivations are not controllable by the experimenter, incentives represent design decisions and it should be made considering the purpose of research. Usually, economic researchers carefully consider the structure of tasks related incentives, that is, how, how far and why they may affect subjects' behavior.

As summed up by Hertwig and Ortmann (2001), economists use financial incentives for at least four reasons: 1. Because they believe that salient payoff reduces variability (Davis and Holt, 1993); 2. For the assumption that saliency is easier to implement than other incentive; 3. Because for financial rewards there is no satiation over the course of an experiment; 4. And finally the most important reason, because the majority of economics experiments test the economic theory that requires, given its maximization assumptions, financial incentives. The same 'cognitive effort' is considered, in order to economists, as a scarce resource, so not-paid experimental subjects "will not invest cognitive effort to avoid making judgment errors [...][14]" (ivi, p. 391).

In their target article, two authors have examined some articles concerning financial incentives, published in the Journal of Behavioral Decision Making from 1988 to 1997. Results show that financial

[13] Harrison, 1992; Roth 1995; Smith, 1991; Smith and Walker 1993a; 1993b.
[14] Smith 1976; 1982; see also Harrison 1989 and 1992.

incentives seem to have a two-fold effect: the convergence of the data toward the performance criterion (even if they do not assure optimal decisions[15]) and the reduction of the data's variance; moreover, they matter more in some areas than in others (in researches on judgment and decision making).

However, overall their analysis shows that the effect sizes for financial incentives are very variable.

These findings are in line with Camerer and Hogarth (1999). They review 74 experimental papers focused on the level of financial performance-based incentive in order to understand when subjects should be paid and why. Their results show that incentives sometimes improve performance, but often do not. In some case higher levels of incentives have the largest effects in judgment and decision-tasks (like in judgment, prediction, problem-solving, recalling items from memory, or clerical tasks), sometimes instead, they do damage when problems are too difficult or intuitive. For instance, in auctions, risky choices and games, incentives do not affect mean performance, but often reduce variance in responses. Moreover, it may happen that in some unfamiliar situations, incentives may guide subjects from "favourable 'self-presentation' behaviour toward more realistic choices. (For example, when they are actually paid, subjects who dictate allocations of money to others are less generous and subjects choosing among gambles take less risk)" (ivi, p. 17).

Gintis, in the 'Commentary' of Hertwig and Ortmann (2001) target paper, points out how economists use game theory to design and interpret experiments. It is the Game Theory that gives a general framework for modelling strategic action and interaction: by defining 'the characteristics of players, the rules according to which they interact, the informational structure available to the agents, and the payoffs associated with particular strategic choices; so from this point of view, the specification of a careful financial incentive is a precondition to predict agent's behavior. However, it may happen that explicit payoffs do not enter in the subject's objective function; for example, in the Ultimatum Game, are fairness and reciprocity to enter in the subjects' objective functions, irrespective of explicit monetary payoffs. As suggested by the author, maybe, the most important reason to have monetary payoffs is to avoid the subject trying to indulge what she/he considers the wishes of the experimenter; financial incentives are able to attenuate this motivation in significant way (ivi, pp. 411-412).

Therefore, leaning exclusively on financial incentives to make clear the economic structure in the lab experiments might determine the missing of some significant phenomena. For instance, to the extent that people treat possessions differently from money, this would have been a hard effect to observe if the only payoffs available to subjects had been monetary (see Kahneman et al., 1990).

[15] These results are in line with works of Smith and Walker's (1993a), Grether 1980; Harless and Camerer 1994; Jamal and Sunder 1991.

To sum up with a Bardsley et al. (2010) statement: "there are objectives for which very high-related incentives would be important; others for which quite modest ones would be adequate; and others […] for which task-related incentives might be actually harmful".

A good experimenter should consider that there are situations in the real world in which profit maximization may have an important role but not decisive and, cases in which, financial incentives interact with nonpecuniary incentives in ways that are not understood (i.e. in a tournament situation).

"Although the use of monetary incentives in an experiment probably rarely decreases external validity, given that they are probably not the most important source of motivation in daily economic life, their contribution to external validity is likely to be minimal" (Loewenstein 1999, p. F32).

1.7 Insufficient opportunity to learn by experience

Lab experiments are often made up on a series of periods in which subject performs the same activity repeatedly. Stationary replication of activity is usually used by experimental economist as a technique for increasing external validity, because seems to be that subjects' behavior is more representative at the end of a series of repetitions. Indeed, during the first task, participants could be confused or they could not have understood roles or rules or, in a computerized experiment, how software works and thus, data obtained might be misleading. To prevent it, repeated trials are usually run and feedback of their previous performance provided. Plott (1996) in the analysis about the 'discovered preference hypothesis', tries to explain how "with practice and experiences, under conditions of substantial incentives, and with the accumulating information that it is obtained from the process of choice, the attitudes stabilize in the sense of a consistent decision rule, reflecting the preferences that were discovered through the process" (ivi, p. 228). So, the repetition could be useful to understand whether the data are results of confusion and inexperience (Ledyard, 1995).

However, an experimental study conducted by Hey (2001) seeks to determine whether such noise is relatively transitory and decays with experience; his results show a high degree of variability in subjects' responses and subjects' behavior during the experiment. For some subjects replications of the tasks improve the consistency of their decisions while for other the noise remains high and the identification of the underlying preference function remains difficult.

Moreover, some consider stationary repetition useful because subjects may apply what they learn to their behavior in other situations. This phenomenon called "transfer of learning" appears weak and its weakness was proved by psychological and economic studies. As far as it concerns psychological works, Bassok et al. (1995) results show that trained subjects through several repetitions

give wrong answers when the problem is changed even only superficially. So, "transfer of learning" across situations is very weak even when subjects are aware that their previous experience is relevant for a second task. From the side of experimental economics, in an auction experiment, Kagel and Levin (1986) find that even if subjects who learned the rule after several repetitions in a 3-person auction, are able to improve their performance in that situation (that is, avoiding winner's curse), they are not able to generalize what they have learned in a slightly different situation (placed in a 6-person auction instead to lower their bids, they raise them).

Furthermore, stationary replications are unusual in the economic life; in fact there are few real-world settings in which subjects are engaged with repetitive situations. So, if on one side stationary repetitions can reduce confusion and make clear the microsystem built in lab, on the other side the same tend "to repress certain types of psychological motives, such as fairness, that may play a prominent role in early-period play" (Loewenstein 1999, p. F29).

1.8 Conclusions

The use of lab experiments in economics has enormously grown over the last three decades and is still growing, even if there is some resistance among academics who argue that experimental data lack generalizability. In this chapter we have reviewed the more salient articles dealing with the main problems related to the external validity and, we have not found a univocal answer: some studies do confirm biases linked to all these issues, others instead, do not.

The choice of particular samples, tasks, incentives or methods, is ultimately a matter of the underlying research question and as suggested by Falk and Heckman (2009) the wiser choice is always to conduct more lab experiments not fewer, to address all possible elements of bias which may undermine, in terms of robustness, the results coming from the lab and their generalizability to the real word.

Table 1.1 Self-selection of participants: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between samples
Eckel & Grossman (2000)	Lab subjects (Volunteers) vs. Classroom subjects (Pseudo Volunteers)	168 subjects out to 181 made useable decisions in a dictator game experiment with a charity as recipient. 81 subjects played four sessions conducted during class time (Pseudo Volunteers, PV) and 87 subjects were recruited to run a lab experiment (Volunteers, V). Each subject selected a single charity, from a list of ten, to receive any contributions made and completed twelve allocation decision problems, determining how many of the tokens in her/his endowment to keep and how many to pass to her charity partner.	1) In the 'PV' sessions, 16 subjects out to 81 contributed with the entire endowment to their charities for all 12 decisions, while only 1 out of 87 of the 'V' subjects gave everything. At the same time, none of the 'V' and only two of the 'PV' passed nothing; 2) The 'PV' are significantly more generous in their contributions to the charities in all treatments. In every case, the 'PV' donated between 22 and 50 % more tokens than did the 'V'. The means are significantly different (p-values < 0.006) in every case; 3) They reject the null hypothesis of no difference in the distributions of contributions made by 'PV' and 'V' (Epps-Singleton test, p-value < 0.03); 4) Non-monetary factors such as religion or altruistic preferences have a greater effect on the giving behaviour of Pseudo Volunteers.	**Yes**
Cleave et al. (2010)	Lab subjects vs. Classroom subjects *(same population)	Authors used two experimental tasks to measure individual preferences: a trust Game and a lottery choice task. Overall 1,451 students attended the classroom experiment (with 254 students assigned to a control treatment). Of the 1,197 students, 1,173 were agreed to participate in the classroom experiment. After 3 months it was sending an e-mail to the students who had expressed an interest in participating in experiments (12 percent of the students from the classrooms). So they also participated to the lab experiment.	1) On average, they do not detect any significant selection bias based on observable characteristics; 2) Students who send less in the trust game are more likely to self-select into the lab experiment, but any significant difference is found in the likelihood to participate in the lab experiment based on either the lottery choice or the percent returned in the trust game; 3) Of the 9 pair-wise tests which examine whether there is selection bias between subgroups, they only detect one significant and one marginally significant difference in selection; women are significantly more likely than men to participate the riskier the lottery they choose, and commerce students are marginally more likely to participate than non-commerce students	**No**
Falk et al. (2010) 'study 1'	Participating in lab students vs. Non-participating in lab students	They use a naturally occurring donation decision as a measure of participants' and non-participants' prosocial inclination; as a proxy they used students' decisions about whether or not they want to contribute a pre-determined amount to two social funds which provide charitable services (required each semester by the University). A subsample of the students in the lab dataset performs a modified dictator game.	1) The raw data analysis does not reveal any significant difference in prosocial inclinations between participants and non-participants; 2) Students with stronger prosocial inclinations are neither more likely to participate in experiments (extensive margin), nor do they participate more often (intensive margin).	**No**

Table 1.2 Self-selection of participants: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between samples
Anderson et al. (2013)	Self-selected students vs. Self-selected Non student adults vs. Non-Self-Selected Trainee Truckers	All subjects (1261 participants) are exposed to the same experimental protocol. They belong to three different population which differ whether subjects are undergraduate college students or not, and/or in the procedures used to recruit them. In the experiment subjects play a social dilemma game (Two-Person Sending Task) exactly once, and are asked to make decisions in both roles.	1) The pro-social inclinations of subjects who self-selected into the experiments do not appear to be significantly different from the inclinations of those who had instead very little opportunity to self-select into the study; 2) Data also reject the more specific hypothesis that approval-seeking subjects are the ones most likely to select into experiments; 3) the share of individuals exhibiting other-regarding concerns, is remarkably smaller among college students, even after controlling for observable differences in socio-demographic characteristics among subject pools. They observe a large difference between self-selected college students and self-selected adults: the students appear considerably less pro-social.	**No**
Exadaktylos et al. (2013)	Students Volunteers vs. Non-Students Volunteers Students Volunteers vs. Students Non-Volunteers Non-Students Volunteers vs. Non-Students Non-Volunteers Students Volunteers vs. Students Non-Volunteers, Non-Students Volunteers, Non-Students Non-Volunteers Students Volunteers vs. Non-Students Non-Volunteers	The study employs data from a survey-experiment conducted with a representative sample of a city's population (765 participants). Authors used behavioural data from five experimental decisions in three canonical games: dictator, ultimatum and trust games. The dataset includes students and non-students as well as volunteers and non-volunteers.	1) Students Volunteers vs. Non-Students Volunteers: students are more strategic players mostly because they make less generous DG offers. However, these differences are never larger than 6% of the pie; 2) Students Volunteers vs. Students Non-Volunteers and Non-Students Volunteers vs. Non-Students Non-Volunteers: volunteers are more likely to both trust and to reciprocate the trust than non-volunteers in the TG. However, the first difference vanishes when making pairwise comparisons within groups. That is, the aggregate effect is not specifically attributable to either students or non-students; 3) Finally, the behaviour of Students Volunteers is different from the rest of the sample only regarding UG offers, and at marginally significant levels; comparison between Students Volunteers and Non-Students Non-Volunteers yields only one (marginally) significant result as well. Self-selected students increase their offers between DG and UG more than non-self-selected, non-students.	**No**

Table 1.3 Representativeness of students samples: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between samples
Cooper et al. (1999)	Chinese Students vs. Chinese Managers	They examine strategic interactions between firms and planners in China in a ratchet effect game, comparing behaviour between: (i) students and managers with field experience in this situation, (ii) standard monetary incentives versus increased monetary incentives, and (iii) sessions conducted "in context", making explicit reference to interactions between planners and managers, and those without any such references. Each experimental session employ between 12 and 16 subjects. Subjects are randomly divided into two groups, firms and central planners. A total of 36 games are played in each session, with subjects switching roles after every 6 games.	As far as it concerns students' and managers' behaviour: 1) a fivefold increase in incentives significantly increases initial levels of strategic play. There are relatively large and statistically significant incentive effects on firms' play, with significantly more strategic play in early rounds of high-pay student sessions compared to standard-pay student sessions. Over time, the standard-pay students are able to erase this difference, suggesting that experience and incentives act as substitutes for each other; 2) games played in context generate greater levels of strategic play for managers, with minimal impact on students; context has a much larger and more consistent effect on managers than on students. In addition, this impact materialize faster and is more robust for managers in their role as central planners than in their role as firms. Contrary to the authors' expectations, students in the role as firms exhibit significantly higher initial levels of strategic play than older, more experienced managers; this surprising result is due to the age and/or lower educational levels of older managers.	**Yes**
Fehr & List (2004)	Students vs. CEOs	Authors study the hidden costs and returns of incentives in situations requiring trust and trustworthiness using two versions of a Trust game (standard and with punishment (TWP)). They recruit 126 subjects for the student treatments from the undergraduate student body at the University of Costa Rica and 76 CEOs.	As far as it concerns the comparison between students and CEOs: 1) CEO principals transfer more money than students. Moreover, for any given transfer level, CEO agents pay back more money than students; 2) the majority of CEO and student principals use the punishment option in the TWP condition. However, CEOs use the punishment option less often; 3) CEO agents pay back more money if the punishment option is not available than when it is available and used. Student agents pay back the same amount of money irrespective of whether the option is not available or available and used. Neither CEO nor student principals who use the punishment option gain in material terms if they are deprived of the option; 4) CEOs consistently achieve higher efficiency levels.	**Yes**

Table 1.4 Representativeness of students samples: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between samples
Haigh & List (2005)	Undergraduate Students vs. Professional Traders	In a between subjects design, undergraduate students and professional traders play a lottery to investigate about the 'myopic loss aversion' in two distinct treatments: Treatment with frequent feedback (F) and Treatment with no frequent feedback (I) . Authors used a specific pool of professional traders recruited from the Chicago Board of Trade.	1) In F traders bet less than students, whereas in I they bet more than students. Comparing student and trader data in I, they find that traders bet significantly more than students; 2) in F, while traders consistently bet less than students, the differences are never statistically significant at conventional levels; 3) combining the results of F and I: traders fall prey to MLA to a greater degree than students	**Yes**
Carpenter et al. (2005)	Undergraduate Students (Middelbury College) vs. Non- traditional Students (Kansas Community College) vs. Workers	In a social framing study, they compare three different samples: undergraduates at small liberal arts college in Vermont (Middelbury College), students of a junior college in Kansas City (Kansas Community College), and workers of a publishing distribution warehouse in Kansas City (using an identical protocol in the field). Participants play an Ultimatum game and a Dictator game.	1) In the Ultimatum Game, offers of Non-traditional Students are higher both than workers and students. In the Dictator game employees are more generous than both typologies students; 2) proposers in the Ultimatum Game in the two experiments in Kansas City made more generous offers than students at the Middlebury College, even controlling for demographic differences (it confirmed the regional differences hypothesis); 3) Kansas City students offer significantly more than both Kansas City workers and Middlebury college students, respectively, whilst in the Dictator Game, are the workers the more generous among all samples; 4) both samples of students show a drop in the average allocations between two games, while the workers offer the same amount, on mean, in both games.	**Yes**
Bellemare & Kroger (2007)	Students vs. Non-Students Representative Sample of Dutch population	Participants play an investment game in which a sender and a responder are both endowed with 500 points; the sender can send money to the responder from his endowment (he/she got a set of 11 investment possibilities). The amount sent is doubled by the experimenters and added to the endowment of the responder. Responders make their decisions using the strategy method.	1) The sample distributions of investments are significantly different between two samples; 2) senders have an estimated 50% probability of making a loss when playing against a random responder from the Dutch population, as opposed to a significantly higher probability when interacting with a responder from the laboratory; 3) the return ratio distribution is more dispersed in the student sample than in the representative sample. Overall, students' behaviour in lab differs from that in the representative sample, with students who make significantly lower investments and rewarding significantly less investments than representative participants.	**Yes**

Table 1.5 Representativeness of students samples: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between samples
Carpenter et al. (2008)	Students vs. Community members	They run a Dictator Game in which students and random members of the community have to choose which charity to support and how much to donate. Firstly, participants have to choose among 13 charities to take the role of recipient (if participants do not like any of the 13 options, they can add one to own choice); once chosen a charity, they are then asked to divide $100 between the charity and themselves.	1) Students allocate significantly less to charity. If it is considered the fraction of people giving away all of the money, the proportions test confirms that students are much less likely to allocate all the money to charity; 2) community members choose their own charity and donate significantly more (32% of them give the entire endowment). 3) Overall, students' sample does not appear representative in the specific context of the charitable giving and the Dictator Game.	**Yes**
Belot et al. (2010)	Students vs. Non Students	They run six experimental sessions: two with students, two with nonstudents and two with a mixed population (a total of 128 subjects). In each session, each subject play five different games. Three games invoke other regarding preferences: the Dictator Game (non-strategic and one-shot); the Trust Game (strategic and one-shot); and the Public Good Game (strategic and repeated). Two games, instead, involve cognitive capabilities and no other-regarding preferences: Beauty-contest (or Guessing Game); and the Second-price sealed bid auction.	1) Students are more likely to behave as homo-economicus agents than non-students in games involving other-regarding preferences (Dictator Game, Trust Game and Public Good Game). These differences persist even when controlling for demographics, cognitive ability and risk preferences; 2) in games that do not engage other-regarding preferences (Beauty-contest and Second-price Auction) there is limited evidence of differences in behaviour between subject pools. In none of the five games there is evidence of significant differences in comprehension between students and non-students; 3) within subject analyses indicate that students are highly consistent in their other-regarding preferences while non-student subjects are inconsistent across other-regarding games.	**Yes/No**
Falk et al. (2010) ***'study 2'***	Students vs. Representative sample	Authors conduct two identical trust experiments using distinct subject pools, students and a representative sample of the population of the city of Zurich, to investigate whether students and non-students exhibit different prosocial inclinations. 1296 participants run the experiment, 295 recruited from the student pool, 1001 recruited from the general population; authors use the same recruitment procedure, the same instructions, the same decision process and the same financial incentives for participants in both experiments.	1) By examining 'trusting behaviour' of first movers, a simple comparison between the two groups reveals only a small difference across the two subject pools not statistically significant; 2) by investigating about second movers' behaviour, they find that for every possible first mover transfer, students make lower average repayments than non-students. All differences are statistically significant; 3) overall, students' and non-students' reciprocation pattern is very similar; the only difference being that students reciprocate on a lower absolute level.	**Yes/No**
Cappelen et al. (2015)	Students vs. Representative sample Economics Students vs. Non-Economics Students	Authors address the question whether lab experiments on student populations are useful to identify the motivational forces present in society at large by comparing the behaviour of a nationally representative population with different student populations into the lab. Participants play four standard dictator games (two as dictator and two as passive recipient) and ten trust games (five as sender and five as responder).	1) Their study demonstrates that student subject groups may not be representative of the social preferences in society at large. They differ from a representative group of non-students both in their level of selfishness and in the relative importance assigned to different moral motives; 2) economics students and non-economics students differ in their level of selfishness.	**Yes**

Table 1.6 The WEIRD people: Summary Enrich et al. (2010) meta-analysis

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between groups
Henrich et al. (2010)	1) Industrialized societies vs. Small-scale societies 2) Western vs. Non-Western societies 3) Contemporary Americans vs. Rest of the West 4) Typical contemporary American subjects vs. Other Americans	Authors present a meta-analysis about a comparison between WEIRD (Western, Educated, Industrialized, Rich, and Democratic) people and other population. They first discuss the evidence about differences between populations drawn from industrialized societies and small-scale societies in some basic psychological domains (visual perception, fairness and cooperation in economic decision-making, folkbiological reasoning, spatial cognition and potential differences). As second contrast, they review data which compare Western with non-Western populations about four of the most studied domains: social decision making (fairness, cooperation, and punishment), independent versus interdependent self-concepts (and associated motivations), analytic versus holistic reasoning, and moral reasoning. The third contrast regards the contemporary American versus the rest of the west population about individualism and related psychological phenomena. The last one comparison investigates about the differences between typical contemporary American subjects and other Americans: considering education for comparisons among contemporary adult Americans; spatial reasoning for comparisons subpopulations of American children; and psychological aspects for comparison between contemporary Americans and previous generations.	1) Although there are several domains in which the data from small-scale societies appear similar to that from industrialized societies, comparative projects involving visual illusions, social motivations (fairness), folkbiological cognition, and spatial cognition all show industrialized populations as outliers; 2) Westerners emerge as unusual – frequent global outliers – on several key dimensions (although robust patterns have emerged among industrialized societies). Many differences are not merely differences in the magnitude of effects but often show qualitative differences which involve effects reversals or novel phenomena such as allocentric spatial reasoning and antisocial punishment; 3) Few research programs have explicitly sought to compare Americans with other Westerners on psychological or behavioural measures. However, those phenomena for which there are sufficient data to make cross-population comparisons, reveal that American participants are exceptional even within the unusual population of Westerners – outliers among outliers; 4) Numerous findings from multiple disciplines indicate that, in addition to many similarities, there are differences among typical subjects and the rest of the American population in unexpected domains. In some of these domains the data from American undergraduates represent even more dramatic departures from the patterns identified in non-Western samples. Further, contemporary American college students appear further removed along some of these dimensions than did their predecessors a few decades earlier. Typical subjects may be outliers within an outlier population.	**Yes**

Table 1.7 The role of experience lab: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between types
Marwell & Ames (1980)	Experienced participants vs. Inexperienced participants	The study reports three replications of previous researches about the predictive utility of the free-rider hypothesis regarding the provision of public goods by groups. One out the three replications considers the role of experience. Authors examined the effects of previous experience on investment behaviour.	1) Overall this study shows that experienced subjects do not behave differently from inexperienced subjects in this situation; 2) The mean investment of all subjects run under conditions in this study it is 94 tokens, while it is 106 for the experienced subjects. Neither the means nor variances differed significantly (F-test).	**No**
Isaac et al. (1984)	Experienced participants vs. Inexperienced participants	In this experiment, given a specific endowment of resources, participants face the decision of allocating between an individual exchange (private good) and a group exchange (public good). It involves three treatment effects: (1) group size, (2) per capita return from the group investment, and (3) inexperienced versus experienced subjects. They define as experienced group of participants a group in which all subjects have previously participated at least one experiment in this particular experimental environment, although not with a group of the same composition of individuals. Their design provides 40 paired observations where the only difference within a pair is experience..	As far as it concerns the experience treatment: 1) Free riding increases with experience, however they are not able to claim with any degree of confidence that this difference cannot have been random.	**No**
Benson & Faminov (1988)	Experienced participants vs. Inexperienced participants	In this experiment, authors investigate about the role of experience of participants in posted price oligopoly markets. The purpose of experimental markets was to determine if previous participations influence subsequent pricing decisions. Three experimental sessions are run in three consecutive weekdays; in each session eight experienced subjects are paired in four duopolies, and eight inexperienced subjects are paired in other four markets.	1) Comparison of data from the two market types show that experience increases the likelihood of cooperation or tacit collusion; 2) Participants assimilate their experience and call upon it after a much longer time period (one month in this case).	**Yes**
Bolton (1991)	Experienced participants vs. Inexperienced participants	Author run an experiment involving alternating-offer bargaining; it comprised ten cells, each distinguished by three treatment variables: structure, subject experience, and discount factors.	As far as it concerns the experience hypothesis: 1) the average observed of opening offers for each round of experienced subjects is virtually identical to that of the final rounds obtained with inexperienced subjects; 2) the standard errors are smaller for experienced subjects than for inexperienced subjects, making it possible to reject the hypothesis of pecuniary-equilibrium play with experienced subjects at all conventional levels of significance	**No**

Table 1.8 The role of experience lab: Experimental studies

Author(s)	Experimental details	Experiment	General Results	Evidence of difference between types
Matthey & Regner (2013)	Experienced participants vs. Inexperienced participants	To assess the influence of previous experiment participation on subjects' behaviour, authors analyse data from four different studies, all involving allocation decisions: Dictator game variant by Güth et al. (2009); three dictator and ultimatum game variants by Klempt and Pull (2009); mini trust game by Regner and Harth (2010) and mini trust game by Bracht and Regner (2011).	In the four studies, subjects who participated in more experiments show significantly less generous behaviour than subjects who participated in less experiments; regression analysis shows a negative correlation between the number of previous participations in experiments and generous behaviour in the allocation situation.	**Yes**
Conte et al. (2014)	Experienced participants vs. Inexperienced participants	The basic decision situation is a linear public goods game. Authors investigate whether and how contribution choices and their dynamics in public goods experiments are affected by i) previous participation in social dilemma-type experiments, which will be referred to as experience; and ii) previous participation in experiments different from the social dilemmas, which will be referred to as history.	1) On average, the experienced systematically contribute with smaller amounts than the inexperienced; 2) on average, the experienced subjects systematically expect the other participants to contribute smaller amounts than the inexperienced; 3) on average, the experienced's beliefs tend to be more accurate than those of the inexperienced from the middle of the game; 4) history influences the proportions of the population but it affects less than experience.	**Yes**
Capraro & Cococcioni (2015)	Experienced participants vs. Inexperienced participants	Authors wonder about cooperation decision-making in one-shot interactions is a history-dependent dynamic process; several evidences suggest that cooperation decision-making in one-shot interaction is most likely when players have a history in participation in similar games sessions. They run a standard two-person Prisoner's Dilemma in which participants are randomly assigned to either of two conditions: 1) the time pressure condition which measures intuitive cooperation; 2) in the time delay condition which measures deliberate cooperation. They use a self-reported measure of experience; participants are asked to answer to what extent they have previously participated in other studies like that (e.g., exchanging money with strangers)?" using a 5 point Likert-scale from 'Never' to 'Several times': a subject was considered inexperienced if he or she answered 'never'.	As far as it concerns the experience topic, they find two major results: 1) promoting intuition versus deliberation has no effect on cooperative behaviour among inexperienced subjects living in a non-cooperative setting; 2) experienced subjects cooperate more than inexperienced subjects, but only under time pressure. These results suggest that cooperation is a learning process, rather than an instinctive impulse or a self-controlled choice, and that experience operates primarily via the channel of intuition.	**Yes**
Xue et al. (2015)	Experienced participants vs. Inexperienced participants	Authors replicate and extend a simple riskless choice experiment reported by Hochman et al. (2014) as supporting loss aversion for money. One out of five hypotheses concerns the possible role of experience with experiments. They test whether participants with greater experience in experiments will have higher maximization rates in prepayment treatments. They defined "experienced" participants as those who had participated at least 10 times previously in experimental sessions, and "inexperienced" as those who had participated no more than 5 times.	Concerning to experience's effect, participants who have been in many economics experiments before do not choose differently than those who are relative novices.	**No**

Chapter 2

On the Representativeness of Standard Laboratory Subjects: An Experimental Study with Simple Binary Dictator Games

Abstract. The use of convenient pool of college students in economic experiments represents one of the most common criticisms related to the generalizability of laboratory data. Replicating in the lab a simple experiment originally run with a representative sample of the population, we show that, despite the differences between college students and subjects of the general population in terms of demographics, cognitive skills and personalities, the observed behavior of the two samples follows a common pattern in a set of binary dictator games focusing on several motives for altruistic behavior.

We compare the choices of undergraduates and subjects representative of population in different treatments and with different reward dimensions by exploiting the experimental design used by Pelligra and Stanca (2013) to investigate social preferences in a field experiment. Two samples follow a common behavioral pattern with the only exception of a significant difference in choices where self-interest may play a prominent role. This gap seems to be related mainly to the academic background of the participants: our sample of undergraduates economics students differ in their degree of self-interested choices both from the representative group of the population and from its sub-sample of students.

Keywords: methodology, external validity, experiments, prosocial behavior.

JEL codes: C91, D03.

2.1 Introduction

One of the most common criticisms about generalizability of experimental data concerns the representativeness of subjects pool generally used in economics research. Some scholars agree that the external validity of research findings might be frustrated by the use of college students as standard sample in experiments. Their massive presence as participants is clearly related to the usual location of research centers in universities campuses[16], to their low opportunity cost and for some peculiarities[17] which make them the best sample to test general economic theories.

College students certainly differ from other samples of population for demographics[18,19], cognitive skills and tend to be more homogeneous. Undergraduates show "unfinished" personalities (Carlson, 1971), less-crystallized attitudes and less-formulated senses of self, more unstable peer group relationships, stronger tendencies to comply with authority and stronger cognitive skills than to older adults (Sears, 1986) and give slightly more homogeneous answers than non-students (Peterson, 2001). Moreover, 'punctual college sophomore' volunteers are usually interested in research, quite willing to collaborate with the experimenter, in pursuit of social approval (Rosenthal and Rosnow, 1969) and concerned about what their choices tell about them and, he/she could be tempted to validate the experimental hypothesis (Orne, 1969).

The lack of representativeness of this convenient pool has been largely explored. One of the first investigations was realized by Gordon et al. (1986). They compare thirty-two studies in which groups of students and non-students have participated in identical experiments and find strong divergences between two samples in twelve studies[20].

The first reflections on topic focused on the possible differences between professionals' and non-professionals'[21] (read students) behavior. Cooper et al. (1999) experimentally analyze the 'standard' sample of students addressing several open questions about the development of the ratchet

[16] According to the list drafted by the Laboratoire Montpelliérain d´Economie Théorique et Appliquée, only 2 out of 173 experimental labs listed are located outside universities campus and, only one is totally independent by academic research activities.

[17] Tendency to have a steep learning curve (Friedman and Cassar, 2004) and to be intelligent and educated (Gätcher, 2010).

[18] Guillén and Veszteg (2006), by analysing data from a single laboratory comprising 8,755 observations (597 experimental sessions from 74 incentive based economic experiments recorded over more than 2 years), find out that demographic differences can explain only 4% of the variations on dependent variable considered (payments received). Carbone (2005) analyses demographic characteristics in relation with subjects' strategic behavior. She uses a unique experimental subject pool that participated in a life-cycle consumption experiment and finds that demographics have no effect on observed behavior.

[19] See Calder et al., 1981; Greenberg, 1987; Kraus, 1995.

[20] Allen and Muchinsky (1984); Christensen-Szalanski et al. (1983); Churchill and Cooper (1971); Cornelius et al. (1984); Fleming (1969); Hakel et al. (1970a); Hakel et al. (1970b); Jago and Vroom (1982); Kavanagh (1975); Moskowitz (1971); Schneider (1982); Stow and Ross (1980).

[21] For a complete review about differences between professionals and non-professionals see G. R. Fréchette (2015).

effect in centrally planned economies; Chinese students exhibit significantly higher initial levels of strategic play than older and more experienced Chinese managers.

The suspicious that professionals' behavior may differ from non-professionals is confirmed by Haigh's and List's (2005) work, in which professional traders show greater myopic loss aversion than undergraduate students. Differences between the two types of groups are also been found by Fehr and List (2004); they investigate experimentally how CEOs respond to incentives and how they provide the same in situations requiring trust and trustworthiness. By comparing their behavior with the behavior of students comes to light that the former are more trusting, show more trustworthiness and use the punishment threat less than students.

The ever-increasing interest for the possible bias related to the use of undergraduates' sample in research is become then, very popular in the studies on social preferences. No wonder saw that "among the papers published on social preferences in the top five economics journals[22] from 2000 to 2010, only four out of 24 papers reported from experiments on non-student populations" (Cappelen et al., 2015).

In a social framing study, Carpenter et al. (2005) compare three different samples: students at Middlebury College, non-traditional students at Kansas City Kansas Community College, and employees at distribution center in Kansas City; they find that proposers in the Ultimatum Game in the two experiments run in Kansas City make more generous offers than Middlebury College students, even controlling for demographic differences (confirming the regional differences hypothesis); Kansas City students offer significantly more than both Kansas City workers and Middlebury college students (in that order), whilst in the Dictator Game, are the workers the more generous among all samples. Furthermore, their results show that both samples of students have a drop in the average allocations between two games, while the workers offer the same amount, on mean, in both games.

Bellemare and Kröger (2007), instead, explore behavior observed in an investment game between a representative sample of Dutch population and students' pool in the laboratory. Their results indicate that trust, trustworthiness and other social preferences, once identified in the lab, are likely to be present, with greater intensity, in the population as a whole.

In a charitable giving game, Carpenter et al. (2008) discover systematic differences between random members of the community and students; citizens choose their own charity and donate significantly more (32% of them give the entire endowment).

Whilst, considering a wide set of standard experimental games (Dictator Game, Trust Game, Public Good Game, Beauty Contest and Second-price Auction), Belot et al. (2010) find out that, in

[22] American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies.

general, students are more likely to behave as homo-economicus agents than non-students in games involving other-regarding preferences, while there is not a significant difference between samples in the games that do not engage them (Beauty-contest and Second-price Auction).

In one of the two studies published in 2013 by Falk et al., the comparison between the behavior of students and the general population using the same experimental protocol (variant of a trust game) show that the first movers display a similar behavioral pattern, whilst, in terms of reciprocation (second movers play) higher investments are reciprocated with high repayments in both groups, but the level of reciprocation is lower for students than for non-students.

In one of the last studies published on that topic, Cappelen et al. (2015) address the question whether lab experiments on student populations are useful to identify the motivational forces present in society at large by comparing the behavior of a nationally representative population with different student populations into the lab. Their results show that students may not be informative of the role of social preferences in the broader population: representative participants differ fundamentally from students both in their level of selfishness and in the relative importance assigned to different moral motives.

As in Cappelen et al. (2015), the aim of our investigation is to verify to what extent experimental results obtained with undergraduates may be representative of the motivational forces of population. We exploit the experimental design and use the results obtained by Pelligra and Stanca (2013) in an artefactual field experiment with a representative sample of adults of Sardinian Region. We replicate the reference experimental design with a sample of 240 undergraduates of the same region in a class experiment. Our contribution enriches the present status of the literature under two main respects: first, we exactly follow the reference protocol and experimental design to make a punctual comparison between two different samples belonging to the same region (no cultural difference is present); second, we use an alternative methodology to disentangle behavior components' from the choices implemented in the binary dictator games.

In the first part of our analysis we compare results obtained with the two different samples in a set of binary-choice dictator games within four treatments and for two reward dimensions, whilst in the second one, we make an analysis of differences between samples for single behavioral component.

Our main results show that there is a common pattern of behavior between students and representative subject of population at least as far as it concerns choices that do not engage self-interest motives. Our sample of undergraduates appears more sensitive to the loss of a possible payoff and

show high level of self-interest that significantly differs from the level scored by representative subjects and from the sub-sample of students belonging to the same representative population.

The remainder of the paper is organized as follows. Section 2.2 describes the experimental procedures and provides specific details about the reference design. In Section 2.3 the testable hypotheses are derived. Section 2.4 presents the statistical analyses and the result. Section 2.5 concludes.

2.2 Experimental procedures

In this section we describe the experimental tasks and treatments.

In order to make a comparison between experimental findings obtained with undergraduates and representative subjects of population, we exactly replicate the experimental design implemented in a field experiment by Pelligra & Stanca (2013), using a sample of 240 undergraduates.

The experimental design involves four treatments based on binary-choice dictator game with two reward dimensions, small (T1-T4) and large (T1L-T4L). Overall, each participant is asked to make eight decisions, one with small and another one with large reward for each treatment. One of the two allocations is considered as a benchmark and has always equal payoffs for the two players (400, 400). The alternative allocations instead vary along the four treatments.

Table 2.1 summarizes all treatments by considering what type of information comes out when players choose alternative allocations in terms of preferences for efficiency, equality, self-interest and competition. Briefly, treatments T1 and T2, consider decisions when giving is costless and keeping out self-interest as motive of choices. Specifically, in T1 the choice of giving corresponds to the benchmark option (400, 400), which increases both efficiency and equality, while the alternative allocation (400, 300) is guided by competitive motives. In T2 is the alternative allocation (400, 500) that corresponds to the choice of giving and increases efficiency and inequality with respect to the benchmark allocation; while, equality and efficiency motives act in opposite direction from each other in this treatment. Treatments T3 and T4 explore decisions when giving and no giving are costly respectively and, can be guided by self-interest motives. Precisely, the choice of giving in T3 (350, 500) and in T4 (450, 500) increases efficiency and disadvantageous inequality.

The same considerations are valid for four large size rewards Dictator Games.

Table 2.1 Treatments T1-T4 (small reward) and T1L-T4L (large reward).

Treatments		Preferences			
	Alternative Allocations	Efficiency	Equality	Self-interest	Competition
T1 - T1L	(400, 300) - (400, 0)	-	-	=	+
T2 - T2L	(400, 500) - (400, 800)	+	-	=	-
T3 - T3L	(350, 500) - (350, 800)	+	-	-	-
T4 - T4L	(450, 500) - (450, 800)	+	-	+	-

Notes: Column 2 reports allocations that can be chosen as an alternative to the benchmark (400, 400) for small and large reward size respectively; columns from 3 to 6 report the sign of the corresponding choice.

We conducted five experimental sessions on January 2011 at the University of Cagliari[23] (Sardinia). We run classroom experiments with 240 students recruited during the hour lessons at the Department of Business and Economics. We used this procedure as we wanted to minimize the effect both of self-selection and of heterogeneity in the cultural background. All the participants were in first year of their BA in economics. They were informed that participation was voluntary and, all those did not wish to participate, left the classroom. Participants run paper and pencil exercises and each of them was randomly matched with a different partner in each situation (perfect stranger matching). A double blind anonymity procedure (subject vs. subject and subjects vs. experimenters) was maintained during and after the classroom experiments about the participants' decisions. Instructions with a general description of experiment and an explanation of the payment mechanism were also read aloud.

We used the same incentive system of the Pelligra and Stanca's design. Once all sessions were completed, one of the couples-matched was randomly selected together with one of the 8 situations played. A random assignment determined who had to be considered the dictator and who the recipient. Dictator's action was then implemented and the corresponding payoff paid in cash; dictator and receiver won both € 400.

2.3 Testable Hypotheses

According to Gächter (2010), students are often the perfect subject pool for answering some fundamental research questions: they are, on average, educated, intelligent and used to learning. All

[23] The instructions are provided in the appendix A.

these characteristics are valuable considering that often, economic theories and especially the theories of social preferences, assume cognitive sophistication.

Following Gächter's rationale and given our focus on motives for altruistic behavior, we can derive the following testable hypotheses.

1st Hypothesis: Samples comparison

Hp. 1: The first hypothesis concerns the proportions of subjects who decide to give within treatments and for both reward dimensions. *Our hypothesis is that the population proportion of undergraduates who choose to give is equal to the population proportion of representative subjects who make the same decision.*

We test this first hypothesis within all treatments and considering both reward dimensions (T1-T4/T1L-T4L) by using Two-sided Wilcoxon-Mann-Whitney test.

2nd Hypothesis: Samples comparison for single behavioral components

Hp. 2: For the second hypothesis we exploit Pelligra's and Stanca's design by focusing on motives of altruistic behavior behind participants' choices. Once disentangled different motives (efficiency, equality, self-interest and competition) on the decision to give, we verify whether behavioral components affect representative subjects' and undergraduates' choices differently.

Our hypothesis is that undergraduates and representative subjects are affected in their choices from behavioral components in the same way.

We test the second hypothesis comparing the total levels of each behavioral component between samples by using Two-sided Wilcoxon-Mann-Whitney test.

2.4 Results

First, in Table 2.2 we summarize the characteristics of the participants. We enrolled a total of 240 undergraduates students (141 female and 99 male), while the sample of representative subjects of Sardinian population is composed by 611 participants (323 female and 288 male). In addition to a series of questions on socio-demographic characteristics, both samples are asked to answer about pro-social behavior, as captured by donations to non-profit organizations and participation to voluntary activity. Moreover we asked a self-reported measure of trust, satisfaction for own life and satisfaction about own financial situation.

Table 2.2: Characteristics of participants

	Undergraduates (n = 240)	**Representative subjects** (n = 611)
Characteristics		
Female, # (%)	141 (59%)	323 (53%)
Family Dimension, mean (min-max)	4.2 (1-10)	3.4 (1-8)
Voluntarism, # (%)	61 (25%)	141 (23%)
No Trust, # (%)	198 (83%)	413 (68%)
Donations, # (%)	100 (42%)	433 (71%)
Life satisfaction, mean (min-max)	7.3 (1-10)	7.6 (1-10)
Financial satisfaction, mean (min-max)	6.4 (1-10)	6.3 (1-10)

2.4.1 Non parametric analysis (Hp. 1)

Figures 2.1-2.4 plot participants' behavior by subjects' samples: Undergraduates and Representative subjects. From a first visual inspection of the graphs, it turns to be quite evident that two samples do show the same behavioral pattern as far as it concerns the direction of choices in all four treatments and for both reward dimensions.

In T1 and T1L, where giving is costless and increases efficiency and equality, the 78% of undergraduates decide to give in T1, and this fraction rises to 86% in T1L. In the representative sample of population, the proportion of givers is the 83% and increases by 6 points percentage in T1L (see Figure 2.1). By comparing independent observations, we do not find any difference statistically significant between samples. We cannot reject the hypothesis of equality between samples' distributions both for T1 and T1L (p=0.11 and p=0.25, respectively, Two-sided Wilcoxon-Mann-Whitney test).

Result 1*: When giving is not costly and increases efficiency and equality, we do not find any difference statistically significant between two proportions of givers in the two samples.*

In T2 and T2L, where giving is still costless and there exists a trade-off between efficiency and equality, the fraction of undergraduates givers is equal to 49% and decreases to 43%, respectively. In the representative sample of population the proportion of givers decreases from 47% in T2 to 43% in T2L (see Figure 2.2). Once again, we do not find any difference statistically significant between the proportion of givers in two samples both for T2 and T2L (p=0.70 and p=0.87, respectively, Two-sided Wilcoxon-Mann-Whitney test).

Result 2: *When giving is not costly and increases efficiency and inequality, we do not find difference statistically significant between two samples.*

In T3 and T3L, where the decision of giving is costly and increases efficiency and inequality disadvantageous, the proportion of students who decide to give in T3 is 11% and increases to 17.5 % in T3L. Instead, as we can see from Figure 2.3, the 29% of representative subjects opts for giving in T3 and 28% in T3L. Even though, students' and representative subjects' decisions go to the same direction, the fraction of students who choose to give is statistically lower than the proportion of representative agents who make the same decision. The statistical comparison between the proportions of givers of two different samples points out a significant difference both in T3 and T3L (p<0.00 and p=0.00, respectively, Two-sided Wilcoxon-Mann-Whitney test).

Result 3*: When giving is costly, increases efficiency and inequality and decreases self-interest, we do find difference statistically significant between two samples.*

As far as it concerns T4 and T4L, in which decisions of not giving are now costly and decrease efficiency and inequality, students' and representative subjects' choices go to the same direction, but the proportion of college students and representative agents who decide to give are statistically different in T4 (p<0.00, Two-sided Wilcoxon-Mann-Whitney test) but *not* in T4L (p=0.15, Two-sided Wilcoxon-Mann-Whitney test). The proportion of students who decide to give in T4 is 68% and decreases to 56 % in T4L, while, the 55% of representative subjects opts for giving in T4 and 50% in T4L. See Figure 2.4.

***Result 4**: When not giving is costly, increases efficiency, inequality and self-interest, we do find difference statistically significant between two samples in T4 but not in T4L.*

These numbers show negligible differences between two samples for the first two treatments in both reward dimensions and reject any significant behavioral difference.

We find significant differences only in the last two treatments between samples and for treatment 3 for both reward dimensions. Our sample of undergraduates appears more sensitive to the possibility to lose an available payoff with respect to the representative sample of population.

Figures 2.1 - 2.4

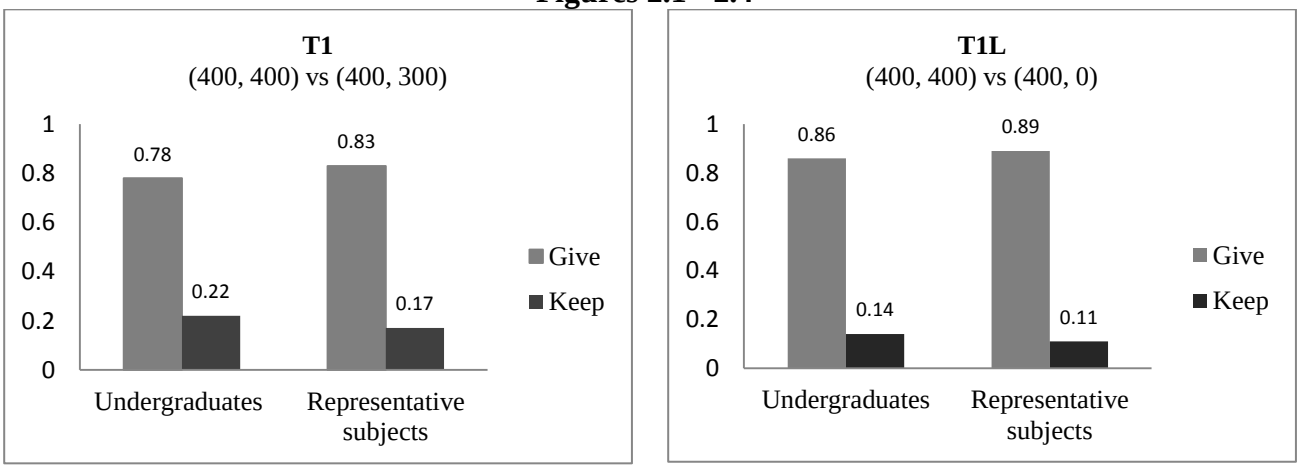


Figure 2.1 Treatment T1-T1L

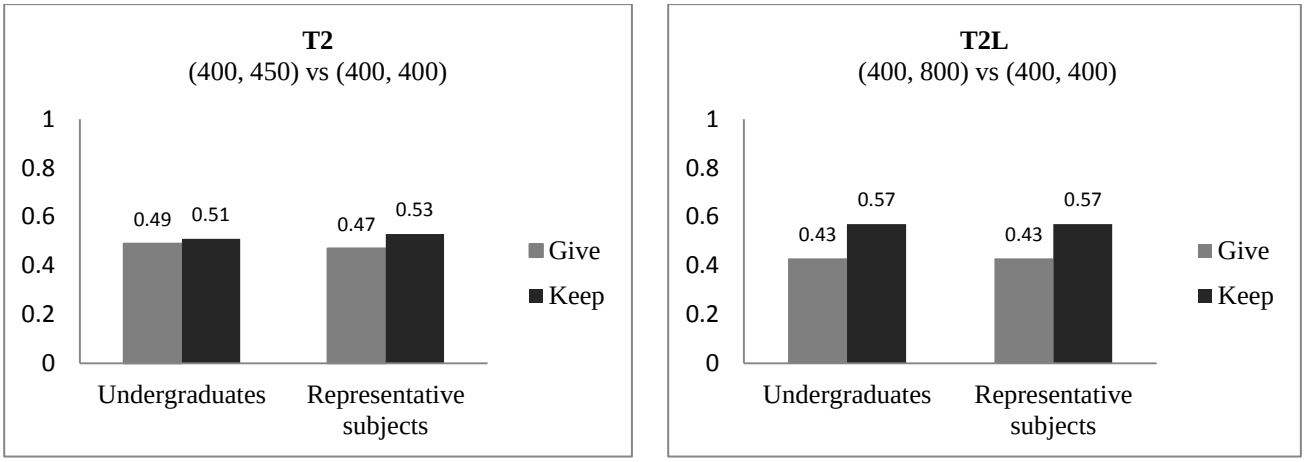
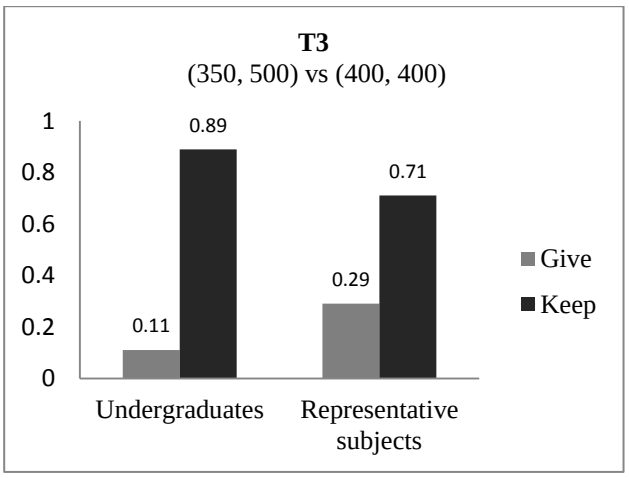


Figure 2.2 Treatment T2-T2L



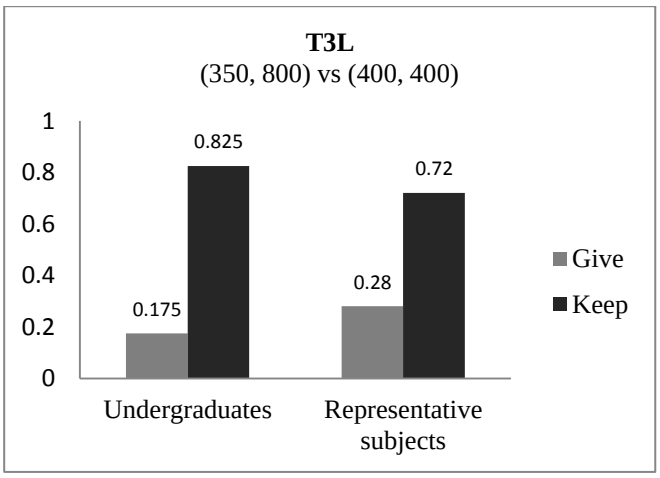
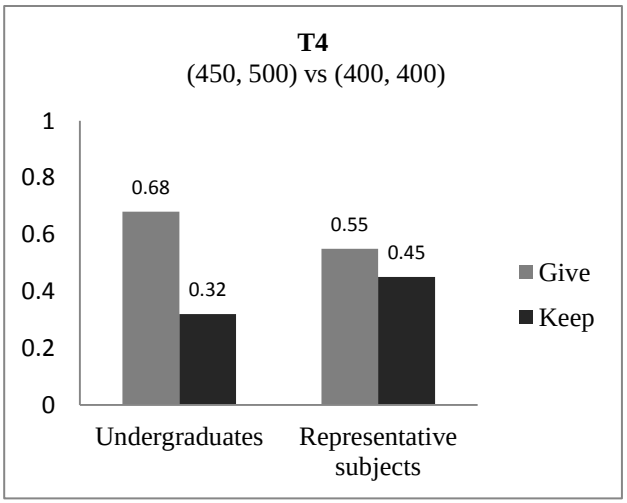


Figure 2.3 Treatment T3-T3L



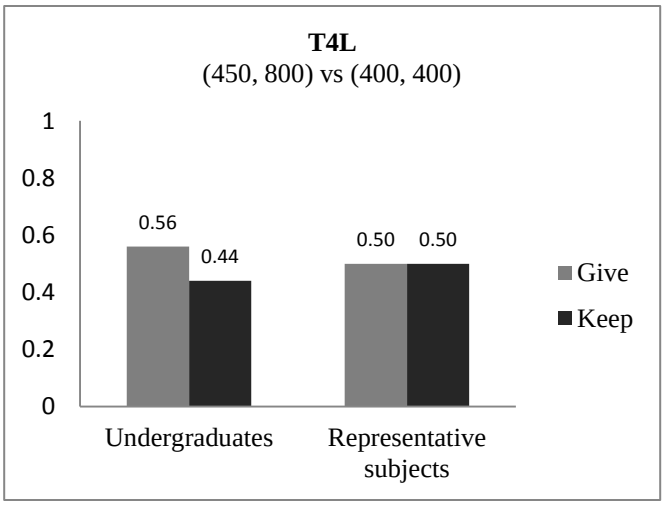


Figure 2.4 Treatment T4-T4L

2.4.2 Non parametric analysis (Hp.2)

Now we check whether the different components of behavior, singularly, affect representative subjects' and undergraduates' choices in a similar way. We have disentangled different motives by exploiting the subjects' decisions in the binary-choice dictator games: a weight that can be -1, 0 or +1 is assigned to each component along treatments in relation to the predicted effect behind the choice of the alternative allocation, as shown in Table 2.1.

For instance, for the choice of the alternative allocation in the treatment T1-T1L, in which the predicted effects are positive for the competition, negative for the equality and the efficiency and null for the self-interest (absent in this treatment), we assign the weight +1, - 1, -1 and 0, respectively for each component, and so on for the other treatments (read the Table 2.3 for each treatment horizontally).

After that, we define the total level for each behavioral component, counting for each sample the number of points scored in all treatments (read Table 2.3 for each behavioral component vertically).

Table 2.3 Components weights

Treatments T1-T4 (small reward) and T1L-T4L (large reward).

		Components weights			
Treatments	Alternative Allocations	**Efficiency**	**Equality**	**Self-interest**	**Competition**
T1 - T1L	(400, 300) - (400, 0)	-1	-1	0	1
T2 - T2L	(400, 500) - (400, 800)	1	-1	0	-1
T3 - T3L	(350, 500) - (350, 800)	1	-1	-1	-1
T4 - T4L	(450, 500) - (450, 800)	1	-1	1	-1
		Total level of efficiency	**Total level of equality**	**Total level of self-interest**	**Total level of competition**

In the 2nd column are reported alternative allocations for all four treatments and for both reward size (small and large); from 3rd to 6th columns are reported weights for each component within treatments. Self-interest weight in T1– T1L and T2 – T2L is equal to zero because it does not play any role in these treatments.

The total levels of efficiency, equality and competition scored by the two pools of subjects result statistically indistinguishable from each other at any conventional level and in both reward dimensions. Instead we find a difference statistically significant for the self-interest motive for both reward dimensions (p=0.000 in both dimensions; Two-sided Wilcoxon-Mann-Whitney test).
Self-interest component, on average, is statistically higher in undergraduates than in representative agents: 1.142 vs 0.517 (small reward dimension) and 0.767 vs. 0.451 (large reward dimension).

Results 5*: When we compare each behavioral component between samples, we find that only the total level of self-interest is statistically different between undergraduates and representative agents for both reward dimensions* (see Tables 2.4-2.5).

Table 2.4. Components Comparison: Small reward dimension

Behavioral Components	Samples	Mean	Std. Dev	MWU - Z	p-value
Total Efficiency Level	Undergraduates	*0.125*	*2.307*	*0.584*	*0.559*
	Representative Subjects	*0.275*	*2.35*		
Total Equality Level	Undergraduates	*0.992*	*1.892*	*0.363*	*0.717*
	Representative Subjects	*1.031*	*2.449*		
Total Competition Level	Undergraduates	*-0.125*	*2.307*	*-0.584*	*0.559*
	Representative Subjects	*-0.275*	*2.305*		
Total Self-Interest Level	Undergraduates	*1.142*	*1.073*	*-7.572*	*0*
	Representative Subjects	*0.517*	*1.078*		

Notes: Columns (5) and (6) are the results of a Wilcoxon-Mann-Whitney Nonparametric test.

Table 2.5. Components Comparison: Large reward dimension

Behavioral Components	Samples	Mean	Std. Dev	MWU - Z	p-value
Total Efficiency Level	Undergraduates	*0.05*	*2.515*	*0.74*	*0.459*
	Representative Subjects	*0.193*	*2.423*		
Total Equality Level	Undergraduates	*1.383*	*2.146*	*0.126*	*0.899*
	Representative Subjects	*1.355*	*2.437*		
Total Competition Level	Undergraduates	*-0.05*	*2.515*	*-0.74*	*0.459*
	Representative Subjects	*-0.0193*	*2.423*		
Total Self-Interest Level	Undergraduates	*0.767*	*1.057*	*-4.047*	*0*
	Representative Subjects	*0.451*	*1.008*		

Notes: Columns (5) and (6) are the results of a Wilcoxon-Mann-Whitney Nonparametric test.

In order to check whether the difference in the behavioral component of self-interest between samples is related to specific demographic characteristics of subjects, as age or schooling level, we compare the sample of undergraduates with the sub-sample of students belonging to representative subjects pool. The result of Wilcoxon-Mann-Whitney Nonparametric test show a difference statistically significant in the small reward dimension between these two sample of students (p=0.017).

However, we do not find any difference significant between students of representative sample and the rest of the same sample (p=0.809 for the small reward).

Non parametric analysis suggests that being a student *per se* it does not seem to be the reason of difference found; the high level of self-interest component appears as an own characteristic of our sample of undergraduates. Even though we do not have details about the majors of students belonging to representative sample, we may speculate about this result, confirming the findings obtained by Cappelen et al. (2011) for which students in economics and students in other disciplines mainly differ in their level of self-interest; non-economics students appear less selfish than students in economics and thus make choices more in line with what we observe in the representative group.

2.4.3 Regression analysis

So as to check the robustness of the result delivered by the non-parametric analysis in the section 2.4.2, we report further GLM regression analyses (Table 2.6 for small reward and Table 2.7 for large reward dimensions) that allow to assess if and how the estimates of behavioral components may be affected by use of student samples by controlling for richer set of individual-specific factors that might influence the behavioral outcomes observed.

In models (**1**) (**2**) (**3**) (**4**) we focus on the behavioral components in which the level of efficiency, the level of equality, the level of competition and the level of self-interest represent the response variables, respectively.

All these response variables are analysed at the light of our main experimental variable, '**Undergraduates sample**'; it is a dummy variable that identifies the sample identity, if undergraduates or representative subjects of Sardinian population. To follow, '**Family Dimension**' identifies the number of family components self-stated; '**Voluntarism**' is a dummy variable that identifies whether subjects declare to make voluntary work; '**No Trust**' is a dummy variable that identifies a measure of trusting self-reported (participants are asked to answer if they have trust in others or have not); '**Donation**' is a dummy variable that identifies whether subjects declare to make donation or not; '**Gender**' (male=1) is a dummy variable that is equal to one if the subject is a male; '**Life satisfaction**' is a self-reported measure of satisfaction: participants are asked to answer how much they are satisfied of their life using a scale from '1' to '10'; and finally, '**Financial satisfaction**' is a self-reported measure of satisfaction as well: in this case participants are asked to answer how much they are satisfied of their financial situation, using a scale from '1' to '10'.

a) Small Reward Dimension

The GLM estimates for the Undergraduates sample dummy variable reject any statistically significant differential effect generated by the group identity - with respect to the representative subjects group – concerning each behavioral component with the only exception of the level of self-interest (see Table 2.6).

As far as it concerns Model (1), the level of efficiency is affected significantly by dummies 'No Trust' and 'Gender'. The level of efficiency is lower in the subjects who have no trust in the people, while, men seem to be more efficient with respect to women. In Model (2), we see that women seem to be more equalitarian than men. The level of competition, Model (3), is influenced by 'No Trust' and 'Gender' dummies; it is higher in the subjects who have stated do not trust of people, while, women seem to be more competitive with respect to men. In Model (4), the level of self-interest is *not* influenced by individual characteristics and beliefs.

This regression analysis confirms the result delivered by the non-parametric analysis: self-interest is the only behavioral component in which undergraduates sample appears different to representative sample.

b) Large Reward Dimension

Considering the large reward dimension, the GLM estimates confirm that the Undergraduates sample dummy variable is statistically significant only with respect to the level of self-interest (see Table 2.7).

As far as it concerns Model (1), the level of efficiency is affected significantly by dummies 'No Trust', 'Donations' and 'Gender'. In Model (2), we see that women seem to be more equalitarian than men. The level of competition, Model (3), is influenced by 'No Trust' and 'Gender' dummies; it is higher in the subjects who have stated do not trust of people, while, women seem to be more competitive with respect to men. In the Model (4), the level of self-interest is not influenced by individual characteristics and beliefs. As in the Small reward Dimension analysis the result delivered by the non-parametric analysis is confirmed by regression analysis: two samples differ only for the behavioral component of self-interest.

Table 2.6. Behavioral outcomes. GLM regressions. Small Reward Dimension

	(1)	(2)	(3)	(4)
OUTCOMES:	**Total level of efficiency**	**Total level of equality**	**Total level of competition**	**Total level of self-interest**
Undergraduates sample (Dummy)	**0.112**	**-0.202**	**-0.112**	**0.649*****
	(0.197)	**(0.194)**	**(0.197)**	**(0.092)**
Family Dimension	-0.064	0.09	0.064	0.042
	(0.067)	(0.066)	(0.067)	(0.031)
Voluntarism	0.111	-0.233	-0.111	-0.04
	(0.189)	(0.186)	(0.189)	(0.088)
No Trust	-0.484*	0.001	0.484*	-0.101
	(0.189)	(0.186)	(0.189)	(0.088)
Donations	0.302	-0.263	-0.302	0.095
	(0.182)	(0.179)	(0.182)	(0.085)
Gender (Male=1)	0.566***	-0.618***	-0.566***	0.098
	(0.165)	(0.162)	(0.165)	(0.077)
Life satisfaction	0.053	-0.026	-0.053	-0.001
	(0.059)	(0.058)	(0.059)	(0.028)
Financial satisfaction	0.011	-0.045	-0.011	-0.002
	(0.046)	(0.045)	(0.046)	(0.021)
Intercept	-0.172	1.731***	0.172	0.349
	(0.529)	(0.52)	(0.529)	(0.246)
Observations	851	851	851	851
Null deviance	4443.5 on 809 d.f.	4273.8 on 809 d.f.	4443.5 on 809 d.f.	1006.07 on 809 d.f.
Residual deviance	4293.1 on 801 d.f.	4151.4 on 801 d.f.	4293.1 on 801 d.f.	931.76 on 801 d.f.
Number of Fisher Scoring iterations	2	2	2	2

Notes: ***, **, * for significant level at the 0.001, 0.01 and 0.05 level, respectively. Standard errors are reported in parenthesis.

Table 2.7. Behavioral outcomes. GLM regressions. Large Reward Dimension

OUTCOMES:	(1) **Total level of efficiency**	(2) **Total level of equality**	(3) **Total level of competition**	(4) **Total level of self-interest**
Undergraduates sample (Dummy)	**0.213**	**-0.213**	**-0.213**	**0.377*** **
	(0.204)	**(0.196)**	**(0.204)**	**(0.088)**
Family Dimension	-0.118	0.028	0.118	-0.019
	(0.069)	(0.067)	(0.069)	(0.029)
Voluntarism	0.142	-0.097	-0.142	-0.029
	(0.196)	(0.189)	(0.196)	(0.084)
No Trust	-0.575**	0.337	0.575**	-0.041
	(0.196)	(0.188)	(0.196)	(0.084)
Donations	0.441*	-0.436*	-0.441*	0.107
	(0.188)	(0.181)	(0.188)	(0.081)
Gender (Male=1)	0.717***	-0.672***	-0.717***	0.021
	(0.171)	(0.164)	(0.171)	(0.073)
Life satisfaction	0.093	-0.049	-0.093	-0.017
	(0.061)	(0.059)	(0.061)	(0.026)
Financial satisfaction	0.013	-0.023	-0.013	-0.006
	(0.048)	(0.046)	(0.048)	(0.021)
Intercept	-0.468	2.193***	0.468	0.629**
	(0.548)	(0.527)	(0.548)	(0.235)
Observations	851	851	851	851
Null deviance	4861.0 on 809 d.f.	4434.2 on 809 d.f.	4861.0 on 809 d.f	871.16 on 809 d.f
Residual deviance	4600.5 on 801 d.f.	4265.1 on 801 d.f.	4600.5 on 801 d.f	849.77 on 801 d.f
Number of Fisher Scoring iterations	2	2	2	2

Notes: ***, **, * for significant level at the 0.001, 0.01 and 0.05 level, respectively. Standard errors are reported in parenthesis.

2.5 Conclusions

A rigorous comparison between undergraduates and a representative sample of population is surely "a [*good*] starting point for investigating on social preferences generalizability to the other social groups" (Gächter 2010, p. 93), but mostly it represents a valid way to answer to methodological questions about the representativeness of samples usually used in laboratory.

As we know the most of our knowledge about social preferences comes from experimental findings obtained with samples of students. Our results confirm the goodness of that pool for similar studies: undergraduates show a similar pattern of behavior with respect to a representative sample of population of the same region in a binary-choice dictator game.

In the first part of our analysis, the simple comparison of samples choices in each treatment shows a very similar directional pattern for all treatments. The treatments (T1-T1L), (T2-T2L) and (T4L) show, on average, negligible differences and the non-parametric analysis rejects any significant behavioral difference between two samples. Instead, we find, in treatments T3-T3L and T4, that the proportion of undergraduates who choose the alternative allocations is statistically different with respect to the representative subjects sample; when giving is costly the proportion of undergraduates who choose to give is statistically lower than the proportion of representative subjects who make the same choice, and when not giving is costly, is the proportion of undergraduates who choose to give higher than the proportion of representative subjects who make same choice. We note a higher sensitivity of undergraduates about the possibility to lose a part of own payoff.

Finally, when we check for single behavioral component between samples, we find difference only in the level of self-interest. Our sample of undergraduates show a statistically higher level of self-interest with respect to both the representative subjects sample and the sub-sample of students belonging to the same group, confirming a possible academic background effect.

Appendix A: Experimental Instructions

By answering this section of the questionnaire, you can win a substantial prize in euro. You have to answer a series of 8 questions. If your name will be extracted among those who participate in this survey, one of the 8 answers you provide will be randomly selected to determine the corresponding prize. Note that you will be matched to another randomly selected participant, who will also win a prize that will depend on your choices.

Example:

Which would you choose between the following two options?

(A) you win 200 euros and the other subject wins 200 euros

(B) you win 300 euros and the other subject wins 250 euros

By choosing A, you would win 200 euro and the other subject would win 200 euro.

By choosing B, you would win 300 euro and the other subject would win 250 euro.

Let us now turn to the actual questions, that might determine your actual win.

[The order of the 8 choices was randomized]

1. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 400 euros and the other subject wins 300 euros

2. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 400 euros and the other subject wins 500 euros

3. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 350 euros and the other subject wins 500 euros

4. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 450 euros and the other subject wins 500 euros

5. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 400 euros and the other subject wins 0 euros

6. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 400 euros and the other subject wins 800 euros

7. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 350 euros and the other subject wins 800 euros

8. Your choice:

(a) you win 400 euros and the other subject wins 400 euros

(b) you win 450 euros and the other subject wins 800 euros**.**

Chapter 3

Does Experience Affect Fairness and Reciprocity in Lab Experiments?

Abstract. One of the most common criticisms about the external validity of lab experiments in economics concerns the representativeness of subjects usually considered in these studies. The ever-increasing number of experiments and the prevalent location of research centres in university campuses produced a peculiar category of participants: Students with high level of laboratory experience built through repeated participations in experimental sessions. We investigate whether the experience accumulated in this way biases subjects' behavior in a set of representative simple games used to study social preferences (Dictator Game, Ultimatum Game, Trust Game, and Prisoner's Dilemma Game). Our main finding shows how subjects having a high level of experience in lab experiments do not behave in a significantly different way from novices.

Keywords: Experimental Methodology, External Validity, Subjects' Experience, Lab Experiment.
JEL codes: D03, D83, C91, C92.

3.1 Introduction

Since its first appearance, experimental analysis of economic behavior has provoked sceptical reactions and criticisms. One of the major aims of the critics has always been the external validity of laboratory results. In particular, a reason behind the lack of generalizability of the conclusions obtained in the lab is associated with the predominant use of students as experimental subjects. The extensive use of pools of students, in fact, may generate problems related to their intrinsic characteristics (the majority of them are college students[24] coming from western industrialized, rich and democratic countries[25] – WEIRD) but the voluntary basis of the enrolment process, may also produce self-selection that in turn may lead to the formation of experimental pools with peculiar characteristics[26].

A further potential source of bias arises from the combined effect of 'location' and 'number': given the prevalent location of labs in university campuses[27] and the ever-increasing number of experiments run in each of these labs, in fact, subjects tend to accumulate laboratory experience through repeated participation in experimental sessions (Friedman and Cassar, 2004).

In this paper we investigate whether having a long record of participations in experimental sessions (**H types** – $participations \geq 15$) alters subjects' behavior in a set of representative experimental games on fairness and reciprocity with respect to a benchmark group made by subjects with low experience (**L types** – $1 \leq participations \leq 5$) in laboratory experiments. We fail to observe any systematic behavioral difference between the two groups.

The role of experience in economic experiments has been largely neglected, with few notable exceptions: Harrison et al. (1987) and Benson and Faminov (1988) discussing IO experiments, Marwell and Ames (1980), Isaac et al. (1984) and Bolton (1991) in bargaining games experiments and more recently, Matthey and Regner (2013), Conte et al. (2014), Capraro and Cococcioni (2015) and Xue et al. (2015). The first two papers document that highly experienced players are more effective as monopolists and are more capable at achieving profitable tacit collusion than inexperienced ones. Marwell and Ames (1980) and Isaac et al. (1984) both consider public good games situations and find no significant differences due to the different level of experience of participants. Bolton (1991) finds a similar negative result in an experiment involving alternating-offer bargaining.

---

[24] See Cooper et al. (1999); Peterson (2001); Fehr and List, (2004); Carpenter et al. (2005); Bellemare and Kroger (2007; Danielson and Holm, (2007); Carpenter et al., 2008; Alatas et al. (2009); Belot et al. (2010); Cappelen et al. (2015); Anderson et al. (2013); Fréchette (2015).

[25] Henrich et al. (2010).

[26] See Eckel and Grossman (2000); Falk et al. (2010); Cleave et al. (2012); Exadaktylos et al. (2013); among others.

[27] According to the list drafted by Laboratoire Montpelliérain d´Economie Théorique et Appliquée's, only 2 out of 173 experimental economics labs in the world are not located in university campus and only one is independent and not related to academic activities (http://leem.lameta.univ-montp1.fr/).

The more recent contributions of Matthey and Regner (2013) explore whether subjects' experience spills over between experiments. Their meta-analysis considers data from four different studies and their results show that subjects with a higher number of participations tend to be less generous in allocation decisions. However this holds true only if participations in experiments that involved games similar to those used in the four studies are considered. Frequency in participation *per se* in laboratory sessions is not taken into account. On the other hand, Conte et al. (2014) focus specifically on the effect of *'experience'* and *'history'*. 'Experience' measures the level of previous participation in public goods games experiments, while 'history' denotes previous participations in experiments involving different games. They find that at the aggregate level, subjects with 'experience' contribute smaller amounts, expect that other players contribute less and hold more accurate beliefs when compared with subjects without it. They show that also 'history' influences subjects' behavior although less than 'experience'.

Focusing on cooperation in one-shot interaction, Capraro and Cococcioni (2015) analyse the history-dependent dynamic process. Many experimental studies suggest that cooperative decision-making in one-shot interactions is most likely a history-dependent dynamic process. They run a standard two-person Prisoner's Dilemma in which participants are randomly assigned to either of two conditions: (i) Time pressure condition which measures intuitive cooperation and (ii) time delay condition which measures deliberate cooperation. Their main findings show that promoting intuition versus deliberation has no effect on cooperative behavior among inexperienced subjects playing in a non-cooperative setting and that experienced subjects cooperate more than inexperienced subjects, but only under time pressure. These results suggest that cooperation is a learning process, rather than an instinctive impulse.

Xue et al. (2015) replicate and extend a simple riskless choice experiment originally devised by Hochman et al. (2014). One of their five hypotheses concerns the possible role of experience. They test whether participants with greater experience in experiments will have higher maximization rates in prepayment treatments. Their results show that individuals who have been participating in many economics experiments before do not choose differently than those who are novices.

Our contribution enriches the current state of the literature and departs from the most relevant contributions (Matthey and Regner, 2013; Conte at al., 2014; Capraro and Cococcioni, 2015), in two main respects: First, Matthey and Regner (2013), run a meta-analysis of four previous studies using dictator, ultimatum and mini-trust games; on the other hand Conte at al. (2014) and Capraro and Cococcioni (2015) run a proper experiment but focusing only on public good games and Prisoner's

Dilemma, respectively. We, on the contrary, design and run a controlled experiment considering a richer set of different games. Second, our contribution differs as far as the quantitative definition of 'experience' is concerned. We denote, in fact, subjects in our pool as experienced only if they have a considerable large number of lab participations (at least fifteen). Matthey and Regner (2013) consider subjects with at most thirteen participations, Conte et al. (2014) define as experienced subjects with at least another participation in experiment with a public good game or a prisoner's dilemma. In Capraro and Cococcioni (2015) there is no hard information about the actual number of participations in experiments but a self-reported measure of experience: Participants are asked to answer to what extent they have previously participated in other studies like that (e.g., exchanging money with strangers), using a 5 points Likert-scale from 'Never' to 'Several times': a subject was considered inexperienced if he or she answered 'never'.

The remainder of the paper is organized as follows. In section 3.2 the experimental procedures are described. Section 3.3 provides specific details about the games and their parametrization. In Section 3.4 the testable hypotheses are derived. Section 3.5 presents the statistical analyses and the main result. Section 3.6 reports further results generated by two side-manipulations. Section 3.7 concludes.

3.2 Experimental Procedures

Our main goal is to investigate the effect of a high level of experience in lab experiments on decision-making in simple representative experimental games focusing on fairness and reciprocity. Exploiting data stored in the recruitment ORSEE system (Greiner, 2015) of the University of Cologne, by design we recruited both *high experience subjects* **(H types)**, that is, individuals with at least 15 participations in experiments and *low experience subjects* **(L types)**, namely, individuals having between 1 and 5 previous participations[28]. These two clusters have been chosen by design in order to assure, on one hand, (i) an adequate number of subjects in each group and, on the other hand, (ii) a sharp difference in the level of individual experience between the two groups.

In our main experimental condition (C1)[29] each participant is asked to make his/her decisions in four standard experimental games without receiving any information about the level of lab experience

[28] Subjects having no experience at all have not been recruited. Since this category of subjects is totally inexperienced with lab experiments, this might be a cause for naïve outcomes as these subjects are often times stressed by the completely new environment they are exposed to, and they are not familiar with the standard procedures.
[29] Further side-manipulation, C2 and C3 are reported in section 6.

of the counterpart[30] he/she is randomly re-matched with in the different games (perfect stranger matching). In practice the simple fact of having accumulated *high experience* **(H types)** vs *low experience* **(L types)** represents the main experimental variable.

We elicited individual behavior in four games: Dictator Game, Ultimatum Game, Trust Game, and Prisoner's Dilemma Game[31]. In the Dictator Game, Ultimatum Game and Trust Game, all the subjects played both roles A (dictator/proposer/trustor) and role B (receiver/responder/trustee) in a strategy-method fashion, and subsequent stages were not announced in advance.

In order to implement an incentive-compatible payment mechanism at the end of each experimental session, only one game and one decision in each session were randomly selected and the corresponding payoff was paid in cash. If Dictator Game, Ultimatum Game and Trust Game were randomly selected for the payments, a random assignment determined which one of the members of the matched couple must be actually considered as player in role A, the other one is considered as player in role B. A's action (Dictator/Proposer/Trustor) is then implemented. Finally, if the Prisoner's Dilemma was randomly selected for the payments, players' actions were implemented.

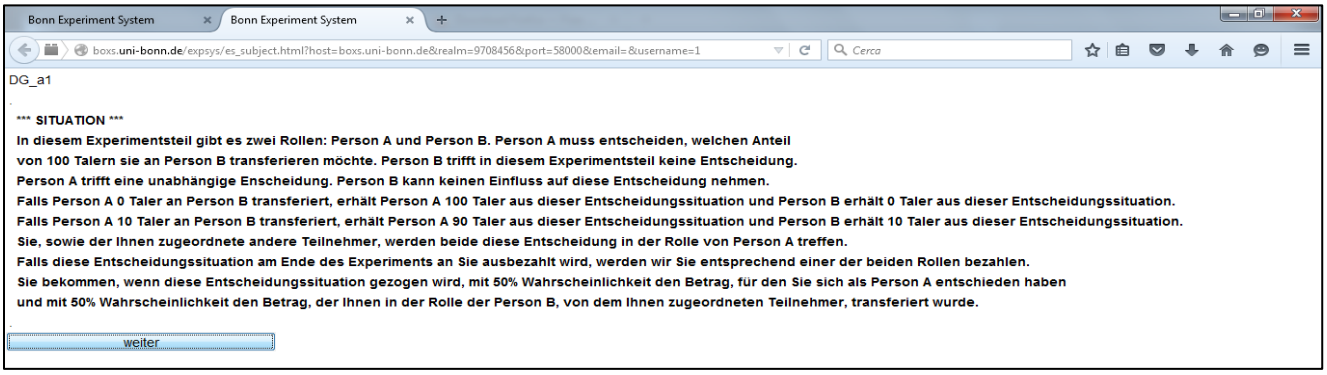
All subjects received 2.50 Euros as show-up fee and got an average experimental payment of 7.50 Euros for a 45 minutes lab session including post-experimental surveys and debriefing. The exchange rate between ECU and Euros was 6 ECU=1 Euro. Six experimental sessions were conducted on January 15th and 16th 2015 at the University of Cologne[32]. The experimental protocol was implemented using the Bonn Experiment System (Seithe, 2012) (Screenshot 1A-B). A double blind anonymity procedure (subject vs subject and subject vs experimenters) was maintained during and after the experiment both about the participants' decisions and the payments. No feedback or results were received by participants before the end of the session.

[30] In the debriefing questionnaire, it is asked to self-report about the number of experiments subjects had already participated in the past. The correlation between this self-reported measure and the actual record provided by ORSEE is 0.89. This shows how subjects are quite aware about their own individual level of experience in laboratory experiments.

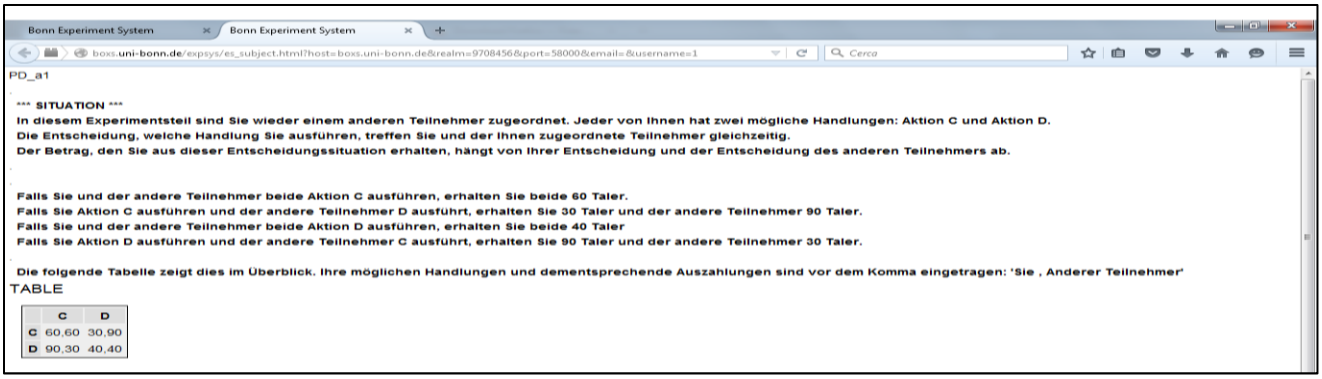
[31] The logical sequence "Dictator Game →Ultimatum Game → Trust Game", moving from the baseline case (DG) to the more sophisticated (TG) interaction, it is implemented in order to favour the comprehension of the games and to avoid confusion. The Prisoner's dilemma game is placed at the end of sequence in order to reduce priming effect and because of the different nature of its dynamics.

[32] The instructions are provided in the Appendix B.

Screenshot 1 (A-B). Screen shots of the computer interface.



Screenshot 1A



Screenshot 1B

3.3 The Games

We elicited individual behavior in four games: Dictator Game, Ultimatum Game, Trust Game, and Prisoner's Dilemma Game. In the Dictator Game, Ultimatum Game and Trust Game, all the subjects played both role A (dictator/proposer/trustor) and role B (receiver/responder/trustee) in a strategy-method fashion, and subsequent stages were not announced in advance.

In Table 1 (A/B) the main parameters of the games are specified and summarized.

In the **Dictator Game,** dictator A is endowed with 100 ECU. She is asked to send any amount between 0 and 100 (in steps of 10) to receiver B who gets such amount.

In the first stage of the **Ultimatum Game**, each subject first plays the role of proposer A. Proposer A is endowed with 100 ECU. She is asked to send any amount of it between 0 and 100 (in steps of 10) to responder B who gets such amount. If B accepts the offer, the two subjects share the endowment as proposed by player A. If B rejects the share offered by A, the entire endowment goes back to the experimenter and the subjects would get 0 ECU each.

In the second stage, each subject plays the strategy method – whether to accept or not offers of 0-10-20-30-40-50-60-70-80-90-100 ECU – for the role of responder B.

In the first stage of the **Trust Game**, each subject first plays the role of trustor A. Trustor A is endowed with 50 ECU. She is asked to send any amount of it between 0 and 50 (in steps of 10) to trustee B. The amount transferred by the trustor A to the trustee B is multiplied by a factor of 3 by the experimenter and sent to B. In the second stage, each subject plays the strategy method for the role of trustee B stating the amount returned to A in the different cases in which she gets 30-60-90-120-150 ECU.

Finally, in the **Prisoner's Dilemma** there are two players and each of them has two possible actions: cooperating or defecting. In order to play the game, both players simultaneously choose one of two actions. The key feature of such game is that for each player, the choice to defect has a higher payoff regardless of the choice made by the other player. That is, we used the classic form of Prisoner's Dilemma where cooperating is strictly dominated by defecting, so that the only Nash equilibrium is for all players to defect. Their earnings depend on both players' actions: if both players decide to cooperate both of them get 60 ECU; if both players decide to defect both of them get 40 ECU; and if one of two players chooses to defect and the other chooses to cooperate, their earnings will be 90 ECU and 30 ECU respectively.

For all the different games, control questions are administrated in order to tests the full comprehension of each game (see Appendix B).

Table 3.1 (A/B): Parametrization of the games

Table 3.1 A Dictator Game, Ultimatum Game and Trust Game

	Dictator Game	**Ultimatum Game**	**Trust Game**
Endowment role A	100 ECU	100 ECU	50 ECU
Endowment role B	0 ECU	0 ECU	0 ECU
Efficiency rate			x3
Role A→ Role B	0-100 ECU (steps of 10)	0-100 ECU (steps of 10)	0-50 ECU (steps of 10)
Role B→ Role A		Accept/Reject *strategy method	0-150 ECU (steps of 10) *strategy method

Table 3.1 B. Prisoner's Dilemma Game

		Other	
		Cooperate	**Defect**
You	**Cooperate**	60, 60	30, 90
	Defect	90, 30	40, 40

3.4 Testable Hypotheses

According to Friedman and Cassar (2004), subjects' behavior changes over time as they get used to the experimental setting. This fact represents an issue both in terms of *intra-session* learning and *inter-sessions* experience accumulation. As far as it regards intra-session learning, Binmore and Shaked (2010) observe how the fact of getting used to the experimental setting leads the subjects to converge to behavioral patterns closer to the 'homo economicus' ones. Nevertheless, it is not clear if this is also the case for inter-sessions experience accumulation in which subjects get used to the experimental setting by participating in several experimental sessions. Following Binmore's rationale, and given our focus on fairness and reciprocity, we can derive the following testable hypotheses for the different games in object:

hp.1: In the Dictator Game the average dictator's offer for the **H types** is smaller than the average dictator's offer for the **L types**.

hp.2a: In the Ultimatum Game the average proposer's offer for **H types** is smaller than the average proposer's offer for **L types.**

hp.2b: In the Ultimatum Game the minimum acceptance offer for **H types responders** is smaller than the minimum acceptance offer for the **L types responders**.

hp.3a: In the Trust Game the average trust rate for **H types** is smaller than the average trust rate for **L types**.

hp.3b: In the Trust Game the average level of trustworthiness for **H types trustees** is smaller average level of trustworthiness for **L types trustees**.

hp.4: In the Prisoner's Dilemma game, the defection rate is larger for **H types** subjects than the defection rate for **L types**.

3.5 Results

First, in Table 3.2 we summarize the characteristics of the participants. We enrolled a total of 134 subjects (77 female and 57 male), aged on average 25 and balanced for the level of laboratory experience: 67 **H type** subjects (min 15, max 86 previous participations in experiments, avg. 31) and, 67 **L type** subjects (min 1, max 5 previous participations in experiments, avg. 3). Data show that, apart for the degree of laboratory experience, the two pools are fairly homogeneous.

In this section we discuss the effect of a *high experience* level exerts on individuals' choices with respect to a *low experience* level.

In our main experimental condition **(C1)**, where players have no information about the level of experience of the partner, any difference in behavior between **H type** subjects and **L type** subjects can be interpreted as the effect of subjects' own level of experience. By comparing their choices in a between-subjects fashion, we can test the general hypothesis about whether experience, *per se*, systematically modifies subjects' behavior.

Table 3.2 Comparability of the two experimental groups

Characteristics	**H types (n = 67)**	**L types (n = 67)**	**delta: Δ (H-L) *P-value***
Experience, mean (min-max)	31.2 (15-86)	3.01 (1-5)	0
Female, # (%)	37 (55.2 %)	40 (59.7 %)	0.6
Age, mean (min-max)	25.7 (19-60)	24.1 (18-65)	0.14
Behavioural Economics classes, # (%)	12 (18 %)	11 (16.4 %)	0.82
Games Theory classes, # (%)	23 (34.4 %)	18 (26.9 %)	0.35

3.5.1 Non-parametric analysis

Figures 3.1a-f plot participants' behavior by subjects' groups: **H** and **L** types. From a first visual inspection of the graphs, it turns to be quite evident that **H** and **L** types do not show different behavioral patterns in all the four different experimental games and this is confirmed by non-parametric tests.

In the **Dictator Game**, out of a budget of 100 ECU, **H** types allocate on average 27.5 ECU to the counterpart, and **L** types allocate 31.1 ECU. The giving rates **(hp.1)** of the two different pools of subjects are statistically indistinguishable from each other at any conventional level (p=0.42, Two-sided Wilcoxon-Mann-Whitney test) (see fig. 1a). The average 1/3 giving rate observed in our experiment is in line with the consolidated result reported in the literature (Engel, 2011).

In the **Ultimatum Game**, out of a budget of 100 ECU for player in role 'A', the average amount sent to 'B' is 39.6 ECU for **H** and 37.9 ECU for **L** types. The giving rates **(hp.2a)** of the two different pools of subjects are statistically indistinguishable from each other at any conventional level (p=0.45, Two-sided Wilcoxon-Mann-Whitney test) (see fig. 1c). The average proposers' offers of about 40% we observe in our ultimatum bargaining experiment are in line with the common results reported in the literature (Güth and Kocher, 2014). For the same game, the average minimum acceptable offer – mao – for players in role 'B' is 29.3 ECU for **H** types and 26.6 ECU for **L** types. The minimum acceptable offers **(hp.2b)** - about 30% of the endowment - in the two different pools of subjects, are statistically indistinguishable from each other at any conventional level (p=0.31, Two-sided Wilcoxon-Mann-Whitney test) (see fig. 1d). Also in this case, the result meets the standard finding of the literature (Güth and Kocher, 2014).

In the **Trust Game**, given the endowment of 50 ECU for player in role 'A', the average amount invested is 23.6 ECU for **H** types and 21.3 ECU for **L** types. The trust rates **(hp.3a)** of the two groups are statistically indistinguishable from each other at any conventional level (p=0.34, Two-sided Wilcoxon-Mann-Whitney test) (see fig. 1e). The fact that on average subjects invest about 45% of their endowments is in line with the findings of the literature (Johnson and Mislin, 2011). Finally, the average return rate **(hp.3b)** in the same game equals to 0.196 for **H** types and to 0.229 for **L** types (p=0.28, Two-sided Wilcoxon-Mann-Whitney test) (see fig. 1f).

In the **Prisoner Dilemma** the rate of defection is 57% for the **H** types and 63% for the **L** types. The defection rates **(hp.4)** in the two different pools of subjects are statistically indistinguishable from each other at any conventional level (p=0.50, χ^2-test) (see fig. 1b). The average 60% defection rate we

observe in our experiment is consistent with the common results reported in the literature (Brosig, 2002).

These numbers show negligible differences between high and low experienced subjects for all the four games and the non-parametric analysis rejects any significant behavioral difference between **H types** and the benchmark group based on **L types.**

See **Table 3.4** for a summary of results of all games (C1).

Figures 3.1a-f Choices in the four games by experience levels

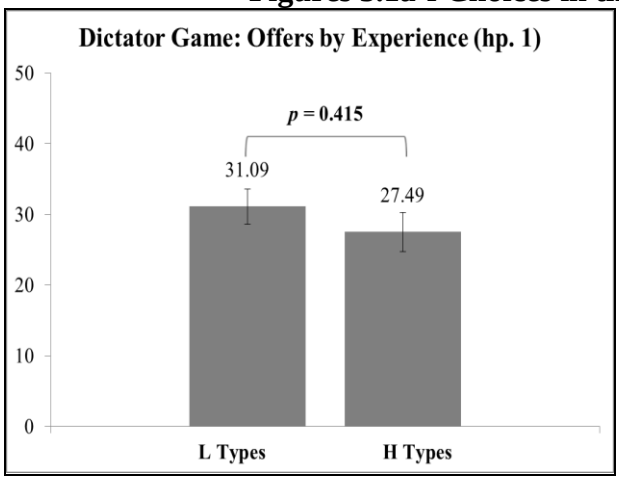


Figure 3.1a: Dictator Game

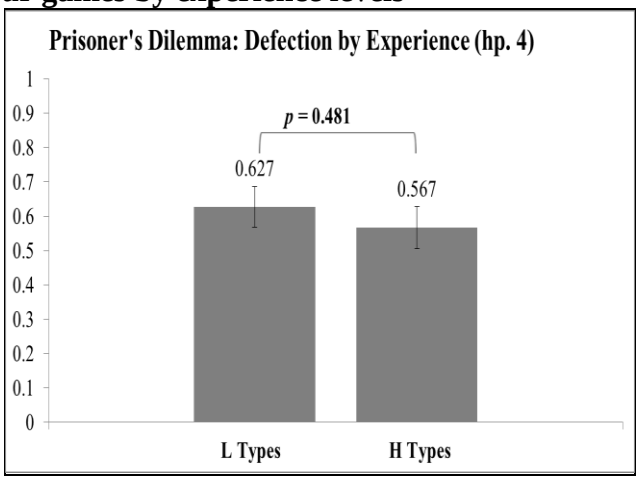


Figure 3.1b: Prisoner's Dilemma

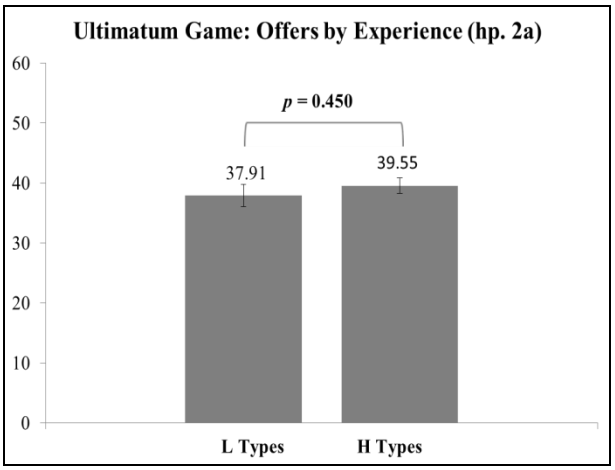


Figure 3.1c: Ultimatum Game (Proposer)

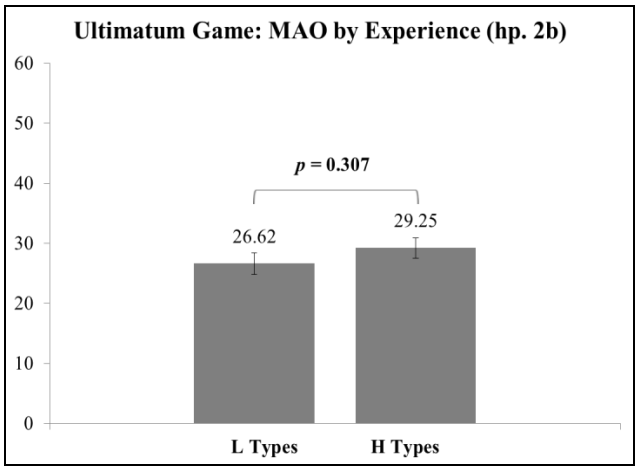


Figure 3.1d: Ultimatum Game (Responder)

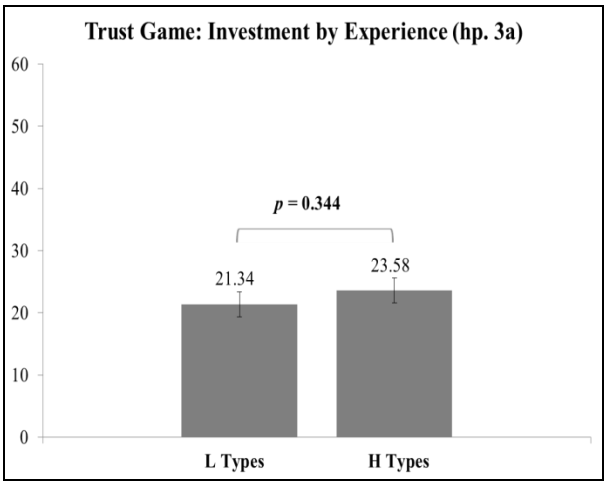


Figure 3.1e: Trust Game (Trustor)

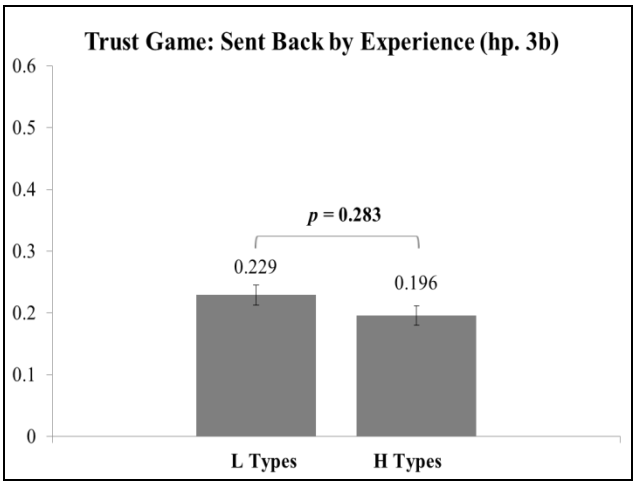


Figure 3.1f: Trust Game (Trustee)

3.5.2 Regression Analysis

In order to check the robustness of the non-result delivered by the non-parametric analysis, **Table 3.3** reports further **OLS regression analyses** that allow to assess the differential effect caused by a high level of laboratory experience controlling for richer set of individual-specific factors that might influence the behavioral outcomes observed in the lab.

Model (1) focuses on the **Dictator Game**: The offer in the dictator game represents the outcome variable. Models (2) and (3) address the **Ultimatum Game**: In model (2) the offer in the ultimatum game represents the outcome variable while in model (3) the minimum accepted offer – mao – is the outcome variable. Models (4) and (5) analyse individual choices in the **Trust Game**: In model (4) the outcome variable is the amount transferred by the trustee to the trustor while the mean return rate from trustor to the trustee is the dependent variable in model (5). In column (6) the propensity to defect in a **Prisoner's Dilemma** game is analysed. The outcome variable is the probability of defection in a linear probability model (all coefficients – by construction – can be interpreted as marginal effects).

All these outcome variables are analysed at the light of our main experimental variable (H type dummy variable) as well as a set of individual-specific control variables that might influence the behavioral outcomes. **H type** is a dummy variable that identifies the high individual level of experience in lab experiments. **Risk Attitude** identifies the elicited individual level of risk-aversion considering the switch point from risky bets to safer ones in standard Holt and Laury (2002) test involving 15 pairs of lotteries. **Experimental Economics Class** is a dummy variable that identifies subjects that have been exposed to an experimental economics class. **Game Theory Class** is a further dummy variable that identifies subjects that have received training in game theory. **Gender** (male=1) is a dummy variable that is equal to one if the subject is a male. **Age** is self-reported in the post experimental questionnaire. **Non-German nationality**, since the vast majority of the experimental subjects are Germans, this dummy variable identifies the 13% of the subjects that declared a nationality other than the German. We also include **Other Demographics** mostly referred to the economic status of the subjects, living conditions and marital status.

The OLS estimates for the **H type** dummy variable reject any statistically significant differential effect generated by a higher level of experience – with respect to the baseline group (Constant) of low experienced subjects (L types) – in all the different games and roles. Offers in the dictator game and in the ultimatum game seem to be marginally negatively affected by lower individual levels of risk attitude, but this negative effect is not detected when we focus on the public good

interaction or on the prisoner's dilemma where – in principle – risk attitude should represent a more substantial issue.

Training in game theory or in experimental economics does not systematically bias subjects' behavior compared to baseline group of subjects who have not been exposed to such training. Males seem to be significantly more generous than females when transferring a share of their endowments to their counterparts in the ultimatum game and trust game. The opposite when we look males' behavior in the ultimatum game. No clear gender-based pattern can be established. The age of the subjects, does not substantially affect the observed outcomes in all the different games.

These regression analyses, taking into account a wider set of control variables that might affect the observed behavioral outcomes, confirms the non-result delivered by the non-parametric analysis: H types do not behave significantly differently with respect to L types.

Table 3.3. Behavioural outcomes (C1). OLS regressions.

	(1)	(2)	(3)	(4)	(5)	(6)
	DG offer	**UG offer**	**UG mao**	**TG trust**	**TG avg. trustworth.**	**PD defection**
OUTCOMES:	**(Hp.1)**	**(Hp.2a)**	**(Hp.2b)**	**(Hp.3a)**	**(Hp.3b)**	**(Hp.4)**
H type (Dummy)	**-3.789**	**2.041**	**3.378**	**3.145**	**-3.485**	**-0.078**
	(3.886)	**(2.302)**	**(2.543)**	**(2.904)**	**(2.523)**	**(0.09)**
Risk Attitude	-1.123**	-0.755**	0.442	0.264	-0.352	-0.011
	(0.509)	(0.301)	(0.335)	(0.38)	(0.33)	(0.012)
Experimental Economics Class	-1.785	1.142	6.347*	-6.284	0.648	-0.166
	(5.485)	(3.248)	(3.566)	(4.099)	(3.561)	(0.127)
Game Theory Class	-2.006	-1.337	1.065	1.507	0.547	0.064
	(4.603)	(2.726)	(3.001)	(3.44)	(2.988)	(0.107)
Gender (Male=1)	-6.928*	4.277*	3.899	8.051***	1.049	0.053
	(3.943)	(2.335)	(2.6)	(2.947)	(2.56)	(0.091)
Age	-0.127	0.055	0.038	-0.489*	0.338	-0.001
	(0.367)	(0.217)	(0.239)	(0.274)	(0.238)	(0.009)
Non-German nationality	-1.799	-6.919**	-1.631	-2.246	-1.862	0.091
	(5.588)	(3.31)	(3.628)	(4.177)	(3.628)	(0.129)
Other Demographics	yes	yes	yes	yes	yes	yes
Constant	41.359***	39.993***	16.881**	30.437***	18.816***	0.887***
	(10.431)	(6.178)	(6.772)	(7.796)	(6.772)	(0.242)
Observations	134	134	132	134	134	134
R-squared	0.109	0.139	0.145	0.136	0.049	0.081

Notes: Three stars, two stars and one star for significant level at the 1%, 5% and 10% level respectively. Standard errors are reported in parenthesis. The sample size relevant for the regression in column (4) is 132 instead of 134, because two subjects stated inconsistent choices in terms of minimum acceptance offer in the Ultimatum Game.

3.6 Further manipulations

In order to further enrich our comprehension about the effect of accumulated experience in lab experiments on subjects' behavioral outcomes, two additional within-subject manipulations (condition C2 and condition C3) have been devised.

While under the condition **C1,** subjects having different levels of experience were purely randomly paired, and no further information was given to them, in the two subsequent unannounced sets of interactions **C2** and **C3** – the ordering of this two side-manipulations was randomized while C1 was always implemented at first in order to keep it totally independent from the other two variations[33]– the following information concerning the level of experience of the counterpart was determined by design and revealed in the instructions for all the four games:

C2: for *H* [*L*]: *"…in this situation you will face a different counterpart who has a HIGH [LOW] level of experience. That is, a subject who has participated in many [in few] experiments";*

C3: for *H* [*L*]: *"…in this situation you will face a different counterpart who has a LOW [HIGH] level of experience. That is, a subject who has participated in few [in many] experiments".*

In order to avoid multiple testing issues, in this section we rely only on regression analyses. As first step, the within-subject first-difference of the outcomes – computed contrasting each behavioral outcome in the main condition C1 against the corresponding action under the manipulated condition **C2** – are assessed. The differential effect generated by the interaction between experienced subjects paired together **(H: vs H)**, compared to baseline pairs of low experienced players **(L: vs L)** are captured by the coefficients for **H type** and the **Constant**, respectively.

As second step, the same exercise is performed contrasting behavioral outcome in the main condition C1 against the corresponding action under the condition **C3** in order to isolate the differential effect generated by high experience subjects interacting with low experienced ones **(H: vs L)** and vice versa **(L: vs H).** In both the cases, control variables are included in the regression analysis.

[33] Sequences: i) C1 / C2 / C3; ii) C1 / C3 / C2 .

3.6.1 C1 vs C2: Pairs with homogeneous levels of experience (H: vs H / L: vs L)

Table 3.5 reports about the change of the behavior when subjects are informed that they are now interacting with an opponent having the same level of experience (condition C2) compared to the baseline behavior elicited under condition C1. For all games, the estimates of the **Constant** are not significant at any conventional statistical level. This means that the individual behavior of L types does not change when they are exposed to C2 compared to the baseline behavior pictured under the C1 condition.

Similarly, the coefficients for the dummy variable **H type** are not significant at any conventional, except for a marginal negative effect (*p-value*=0.06) detected for the Minimum Acceptable Offer in the Ultimatum game. The general pattern across games, shows how highly experienced subjects do not behave differently from less experienced ones.

The control variables do not show any consistent pattern of significant effects on subjects' behavior.

3.6.2 C1 vs C3: Pairs with heterogeneous levels of experience (H: vs L / L: vs H)

Table 3.6 reports about the change of the behavior when subjects are informed that they are now interacting with an opponent having a different level of experience (condition C3) compared to the baseline behavior elicited under condition C1. For all games, the estimates of the **Constant** are not significant at any conventional statistical level. This means that the individual behavior of low experienced subjects does not change when they are exposed to C3 compared to the baseline behavior pictured under condition C1.

Similarly, the coefficients for the dummy variable **H type** are not significant at any conventional. Also in this third case, this shows how highly experienced subjects do not behave differently from less experienced ones.

The control variables do not show any consistent pattern of significant effects on subjects' behavior.

3.7 Conclusions

The generalizability of conclusions drawn from lab experiments is still a debated issue in economics. It is of course a multifaceted problem that refers to many dimensions of the experimental practices and methods: the artificiality of the situations considered in the lab, the small size of the incentives, the lack of representativeness of the experimental subjects, are only few examples. In particular when we consider the reliability of the conclusions drawn from experiments involving convenience pool of students we should also ask whether the repeated participation into different experiments by these subjects might have a lasting effect on their behavioral tendencies in the lab. Were this to be true, in fact, experienced subjects would constitute an even less representative pool whose behavior patterns could not be reliably generalized. In this paper we addressed precisely this point. By design we investigated whether having repeatedly taken part in previous experiments consistently modifies individuals' behavior in a set of widely used games focusing on fairness and reciprocity: Dictator Game, Ultimatum Game, Trust Game, and Prisoner's Dilemma Game. We considered a between-subjects design to compare the behavior of high experienced and low experienced subjects in the four games. Our data show that a high level of experience per se does not influence subjects' behavior compared to a benchmark pool of low experienced subjects.

Table 3.4. Experience Effect. Summary and Results (all games - C1)

Games	Subjects	Mean	Std. Dev	Median	Min	Max	MWU - Z	p-value
Dictator Game	High	27.5	22.6	30	0	80	0.815	0.415
	Low	31.1	20.6	40	0	60		
Ultimatum Game Proposer	High	39.6	10.4	40	10	60	-0.755	0.45
	Low	37.9	10.4	40	0	100		
Ultimatum Game *MAO* Responder	High	29.3	13.9	30	0	50	-1.022	0.307
	Low	26.6	14.7	30	0	50		
Trust Game Trustor	High	23.6	16.2	20	0	50	-0.946	0.344
	Low	21.3	16.6	20	0	50		
Trust Game Trustee *% Average resent*	High	19.6	12.7	23.3	0	42	1.072	0.283
	Low	22.9	13.4	27.1	0	50		

Notes: Columns (8) and (9) are the results of a Wilcoxon-Mann-Whitney Nonparametric test.

Games	Subjects	Defection Share	Std. Dev	Median	Min	Max	X^2	p-value
Prisoner's Dilemma	High	0.57	0.5	1	0	1	0.496	0.481
	Low	0.63	0.49	1	0	1		

Notes: Columns (8) and (9) are the results of Chi-square test, Pearson chi2(1).

Table 3.5. Within-subject first-differences of the outcomes: (ΔC1 - C2). OLS regressions.

	(1)	(2)	(3)	(4)	(5)	(6)
Δ C1-C2	***Δ C1-C2***	***Δ C1-C2***	***Δ C1-C2***	***Δ C1-C2***	***Δ C1-C2***	***Δ C1-C2***
OUTCOMES:	**DG offer**	**UG offer**	**UG mao**	**TG trust**	**TG avg. trustworth.**	**PD Defection**
H type	**0.971**	**-0.619**	**-3.331***	**-0.047**	**1.071**	**0.073**
[H: vs H]	**(3.405)**	**(2.334)**	**(1.749)**	**(1.913)**	**(1.345)**	**(0.075)**
Risk Attitude	-0.148	-0.530*	-0.37	-0.216	0.07	-0.004
	(0.446)	(0.306)	(0.23)	(0.25)	(0.176)	(0.01)
Experimental Economics Class	-0.146	5.394	-0.498	2.216	2.097	-0.075
	(4.806)	(3.294)	(2.446)	(2.7)	(1.898)	(0.105)
Game Theory Class	-0.079	1.467	4.613**	0.779	1.005	0.003
	(4.034)	(2.765)	(2.061)	(2.266)	(1.593)	(0.089)
Gender (Male=1)	-2.969	1.08	1.713	2.291	-2.602*	-0.027
	(3.455)	(2.368)	(1.787)	(1.941)	(1.365)	(0.076)
Age	0.971	-0.619	-3.331*	-0.047	1.071	0.073
	(3.405)	(2.334)	(1.749)	(1.913)	(1.345)	(0.075)
Non-German nationality	0.305	-0.969	5.191**	3.92	-2.212	-0.175
	(4.897)	(3.356)	(2.49)	(2.75)	(1.934)	(0.107)
Other Demographics	yes	yes	yes	yes	yes	yes
Constant	**7.9**	**0.571**	**1.966**	**1.102**	**-4.537**	**0.33**
[L: vs L]	**(9.141)**	**(6.266)**	**(4.648)**	**(5.134)**	**(3.61)**	**(0.201)**
Observations	134	134	131	134	134	134
R-squared	0.026	0.09	0.131	0.051	0.137	0.071

Notes: Three stars, two stars and one star for significant level at the 1%, 5% and 10% level respectively. Standard errors are reported in parenthesis. The sample size relevant for the regression in column (4) is 131 instead of 134, because three subjects stated inconsistent choices in terms of minimum acceptable offer in the Ultimatum Game.

Table 3.6. Within-subject first-differences of the outcomes: (ΔC1 - C3). OLS regressions.

	(1)	(2)	(3)	(4)	(5)	(6)
Δ C1-C3	***Δ C1-C3***	***Δ C1-C3***	***Δ C1-C3***	***Δ C1-C3***	***Δ C1-C3***	***Δ C1-C3***
OUTCOMES:	**DG offer**	**UG offer**	**UG mao**	**TG trust**	**TG avg. trustworth.**	**PD defection**
High type	**-1.694**	**0.433**	**-0.82**	**3.281**	**1.329**	**-0.062**
[H: vs L]	**(3.709)**	**(2.57)**	**(1.577)**	**(2.156)**	**(1.608)**	**(0.09)**
Risk Attitude	-0.238	-0.622*	-0.034	0.144	0.024	-0.003
	(0.485)	(0.336)	(0.207)	(0.282)	(0.211)	(0.012)
Experimental Economics Class	-1.442	5.998	3.313	1.617	1.461	0.016
	(5.234)	(3.627)	(2.211)	(3.043)	(2.27)	(0.126)
Game Theory Class	0.977	-1.846	0.329	0.45	3.828**	0.135
	(4.393)	(3.044)	(1.861)	(2.554)	(1.905)	(0.106)
Gender (Male=1)	-5.148	0.573	1.273	3.916*	-0.591	0.134
	(3.763)	(2.607)	(1.613)	(2.188)	(1.632)	(0.091)
Age	-0.17	-0.002	0.041	-0.124	0.174	-0.003
	(0.35)	(0.243)	(0.148)	(0.204)	(0.152)	(0.008)
Non-German nationality	4.155	1.03	1.314	1.028	-3.286	-0.058
	(5.333)	(3.695)	(2.25)	(3.1)	(2.313)	(0.129)
Other Demographics	yes	yes	yes	yes	yes	yes
Constant	**11.789**	**5.319**	**-2.242**	**0.515**	**-2.587**	**-0.075**
[L: vs H]	**(9.955)**	**(6.898)**	**(4.2)**	**(5.788)**	**(4.317)**	**(0.24)**
Observations	134	134	132	134	134	134
R-squared	0.036	0.078	0.043	0.071	0.106	0.05

Notes: Three stars, two stars and one star for significant level at the 1%, 5% and 10% level respectively. Standard errors are reported in parenthesis. The sample size relevant for the regression in column (4) is 132 instead of 134, because two subjects stated inconsistent choices in terms of minimum acceptable offer in the Ultimatum Game.

Appendix B: Experimental Instructions

GENERAL INSTRUCTIONS

Welcome and thank you for participating in this experiment. The aim of this study is to investigate how people make decisions in particular situations. Feel free to ask questions at any time before the session begins, we will answer you privately. From now until the end of the session, unauthorized communication of any nature with other participants is prohibited. Decisions have to be made individually and in private.

This is not a test, so there are no right or wrong answers. Just think about what is best for you and act accordingly. Your decisions will be strictly anonymous and could not be linked to you in any way. The data collected will be used only for scientific purposes and stored for the duration of this study. At the end of the session one of situations will be randomly selected and you will be paid in cash according to the choice you made in that particular situation (the rules of the specific situation are explained in details below). You will also receive you € 2.5 as show-up fee. Note that in each situation, you will be paired with a different person. You will not be told to whom you are matched with during or after the experiment, and he or she will not be told who you are either during or after the experiment. The experiment involves three phases, and overall, it will last approximately 60 minutes.

Phase 1 [Main experimental condition C1]

In this part of the experiment you will be paired in each situation with a different person who will also get a reward that will depend on you choice or on the combination of your and his/her own choice. You will not be told to whom you are matched with during or after the experiment, and he/she will not be told who you are either during or after the experiment.

Situation 1

In this task there are two subjects: Person A and Person B. Person A has to decide what portion, if any, of 100 experimental points he/she wants to transfer to Person B. Person B does not have any decision to make, the final distribution of points depends only on Person A. If Person A transfers 0 points to Person B, Person A keeps all his/her endowment of 100 experimental points and Person B will get 0 points; if Person A transfers 10 points to Person B, Person A will get 90 points and Person B 10, and so on. You will play the role of Person A and Person B. At the end we will randomly pick one of the two roles and, if this game will be selected, you will be paid accordingly.

Now you are Person A. Your decision is a simple one: what portion, if any, of 100 experimental points, you want to transfer to Person B? Remember, your choice can be anywhere from 0 to 100, in 10 points increments (points will be converted in cash: 6 experimental points = €1).

Now it is time to make your decision.

How much do you want transfer to Person B? [___]

Situation 2

There are two players in this game: 'YOU' and the 'Other' player. Each of you has two options: Action C and Action D. In order to play the game, both of you simultaneously choose one of your actions. Remember your earnings depend both on your choice and the other player's choice. Your choices will give you the chance to get a certain number of experimental points that will be converted in money (6 experimental points = €1). If you and the other player play C, both of you will get 60 experimental points; if you play C and the other player plays D, you will get 30 points and the other player will get 90 points; if you and the other player play D, both of you will get 40 points; if you play D and the other player plays C, you will get 90 experimental points and the other player will get 30 points.

The table below summarizes the game (players, actions and payoffs). Your payoffs are indicated before the comma, other's payoff after.

		Other	
		Cooperate	**Defect**
You	**Cooperate**	60, 60	30, 90
	Defect	90, 30	40, 40

Control Questions

Now verify if you have understood the game. If the other player plays D and you play C:

- You get...
- Other player gets...

If the other player plays C and you play D:

- You get...
- Other player gets...

Which action would you like to play, Action C or Action D? __

Situation 3

There are two players in this game: Person A and Person B. Person A has 100 experimental points as initial endowment and he/she must decide how much to send to Person B. In turn, person B may accept or reject Person A's offer. If Person B accepts, he/she gets the money that Person A sent, and Person A keeps the remaining points (100 minus the amount sent); if Person B rejects Person A's offer, both get nothing (0 points). Example: if Person A sends 10 points to Person B and Person B rejects that offer, Both A and B get nothing; if Person B accepts Person A's offer, he/she will get 10 points and Person A 90 (100 minus 10). You will make decisions both as Person A and as Person B.

You will play both the role of Person A and Person B. At the end we will randomly pick one of the two roles and, if this game will be selected, you will be paid accordingly.

Now you are Person A. Your decision is a simple one: how many, if any, of 100 experimental points you want to transfer to Person B? Remember, your choice can be anywhere from 0 to 100, in 10 points increments (points will be converted in cash: 6 experimental points = €1).

Control questions.

Now verify if you have understood the game.

You are Person A and assume that your initial endowment is 10 points.

1. You send to Person B 3 points. Person B accepts your offer. How much do you get?

- You get…
- Person B gets…

2. You send to Person B 4 points. Person B rejects your offer. How much do you get?

- You get…
- Person B gets…

You are Person B and assume that initial endowment of Person A is 10 points.

3. Person A sends you 2 points. You reject that offer. How much do you get?

- You get…
- Person A gets…

4. Person A sends you 4 points. You accept the offer. How much do you get?

- You get…
- Person A gets…

You are Person A. Your initial endowment is 100 experimental points.

How much of your initial endowment do you want to send to Person B (in 10 points increments)?

—

You are Person B. Initial endowment of person A is 100 experimental points. Person A sends you:

- 0, do you accept or reject? [YES / NO]
- 10, do you accept or reject? [YES / NO]
- 20, do you accept or reject? [YES / NO]
- 30, do you accept or reject? [YES / NO]
- 40, do you accept or reject? [YES / NO]
- 50, do you accept or reject? [YES / NO]
- 60, do you accept or reject? [YES / NO]
- 70, do you accept or reject? [YES / NO]
- 80, do you accept or reject? [YES / NO]
- 90, do you accept or reject? [YES / NO]
- 100, do you accept or reject? [YES / NO]

Situation 4

There are two players in this game: Person A and Person B. Person A has 50 experimental points as initial endowment and he/she have to decide how much of this amount, if any, he/she wants to send to person B. Person B will receive that amount multiplied by 3. For instance, if Person A sends 10 points to Person B, Person B will receive 30 points; if Person A sends 20 points, Person B will receive 60 points, and so on. In turn, Person B will have to decide how much of amount received, if any, he/she wants to send back to Person A. You will make decisions both as Person A and as Person B. If this situation will be drawn to be paid at the end of this session, we will pay you for one of two roles (A or B), selected randomly (6 experimental points = €1). Person A's earnings will be equal to: initial endowment minus (-) no. points sent to Person B plus (+) points received back by Person B. Person B's earnings will be equal to: points sent by Person A multiplied by 3 minus (-) points sent back to person A.

Control questions.

Now verify if you have understood the game.
You are Person A and suppose that your initial endowment is 10 points.
1. If you send 3 points to Person B, how much does person B get?
- Person B gets...
2. If you send 4 points to Person B and Person B re sends you 0, how much do you get?
- You get...
- Person B gets...

Now you are Person B and suppose that initial endowment of person A is 10 points.
3. Person A sends you 2 points, how much do you receive?
- You get...

4. Person A sends you 5 points and you re-send 0 points, how much do you get?
- You get...
- Person A gets...

You are Person A and your initial endowment is 50 experimental points.

How much of your initial endowment of 50 experimental points do you want to send to Person B (in 10 points increments)? ___

You are Person B and the initial endowment of Person A is 50 experimental points. If Person A sends you:

- 10 (so you receive 30). How many points do you send back in increments of 10 experimental points? __
- 20 (so you receive 60). How many points do you send back in increments of 10 experimental points? __
- 30 (so you receive 90). How many points do you send back in increments of 10 experimental points? __
- 40 (so you receive 120). How many points do you send back in increments of 10 experimental points? __
- 50 (so you receive 150). How many points do you send back in increments of 10 experimental points? __

Phase 2 – [C2 condition]

As in condition C1 plus information about the other player's level of experience (same level)

Now starts a new phase of the experiment.

(for Low types)

You have a LOW level of experience in laboratory, that is, you have participated in few experiments and, in this part of the experiment you will be paired in each situation with a different person who has your same level of experience and who will also get a reward that will depend on your own choice or both of you choices, depending on the situation. You will not be told who you are matched with during or after the experiment and he or she will not be told who you are either during or after the experiment.

Control questions.

Your level of experience is...

The level of experience of your partner is...

(for High types)

You have an HIGH level of experience in laboratory, that is, you have already participated in many experiments and, in this part of the experiment you will be paired in each situation with a different person who has your same level of experience and, who will also get a reward that will depend on your own choice or both of you choices, depending on the situation. You will not be told who you are matched with during or after the experiment, and he or she will not be told who you are either during or after the experiment.

Control questions.

Your level of experience is...

The level of experience of your partner is...

Situation 1

(for Low types)

As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a LOW level of experience as well.

In this task there are two subjects: Person A and Person B. Person A has to decide what portion, if any, of 100 experimental points he/she wants to transfer to Person B. Person B does not have any decision to make, the final distribution of points depends only on Person A. If Person A transfers 0 points to Person B, Person A keeps all his/her endowment of 100 experimental points and Person B will get 0 points; if Person A transfers 10 points to Person B, Person A will get 90 points and Person B 10, and so on. You will play the role of Person A and Person

B. At the end we will randomly pick one of the two roles and, if this game will be selected, you will be paid accordingly.

Now you are Person A. Your decision is a simple one: what portion, if any, of 100 experimental points, you want to transfer to Person B? Remember, your choice can be anywhere from 0 to 100, in 10 points increments (points will be converted in cash: 6 experimental points = €1).

Now it is time to make your decision. How much do you want transfer to Person B? ___

(for High types)

As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person B who has a HIGH level of experience as well.

In this task there are two subjects: Person A and Person B. Person A has to decide what portion, if any, of 100 experimental points he/she wants to transfer to Person B. Person B does not have any decision to make, the final distribution of points depends only on Person A. If Person A transfers 0 points to Person B, Person A keeps all his/her endowment of 100 experimental points and Person B will get 0 points; if Person A transfers 10 points to Person B, Person A will get 90 points and Person B 10, and so on. You will play the role of Person A and Person B. At the end we will randomly pick one of the two roles and, if this game will be selected, you will be paid accordingly.

Now you are Person A. Your decision is a simple one: what portion, if any, of 100 experimental points, you want to transfer to Person B? Remember, your choice can be anywhere from 0 to 100, in 10 points increments (points will be converted in cash: 6 experimental points = €1).

Now it is time to make your decision. How much do you want transfer to Person B? ___

Situation 2

There are two players in the game: you and another player you are paired with.

(for Low types)

You have LOW experience and the other player has LOW experience as well.

Now look the table below and make your decision. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face a player with a **LOW** level of experience as well.

Other

You		**Cooperate**	**Defect**
	Cooperate	60, 60	30, 90
	Defect	90, 30	40, 40

Which action would you like to play, Action C or Action D? __

(for High types)

You have HIGH experience and the other player has HIGH experience as well.

Now look the table below and make your decision. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face a player with HIGH level of experience as well.

Other

You		**Cooperate**	**Defect**
	Cooperate	60, 60	30, 90
	Defect	90, 30	40, 40

Which action would you like to play, Action C or Action D? __

Situation 3

There are two players in the game: Person A and Person B.

(for Low types)

Person A has LOW experience and person B has LOW experience as well.

(for High types)

Person A has HIGH experience and person B has HIGH experience as well.

(for Low types)

You are Person A. Your initial endowment is 100 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a LOW level of experience as well.

How much of your initial endowment do you want to send to Person B (in 10 points increments)?

(for High types)

You are Person A. Your initial endowment is 100 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person B who has a HIGH level of experience as well.

How much of your initial endowment do you want to send to Person B (in 10 points increments)?

(for Low types)

You are Person B. Initial endowment of person A is 100 experimental points.

As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and in this situation you will have to face Person A who has a LOW level of experience as well. Person A sends you:

- 0, do you accept or reject? [YES / NO]
- 10, do you accept or reject? [YES / NO]
- 20, do you accept or reject? [YES / NO]
- 30, do you accept or reject? [YES / NO]
- 40, do you accept or reject? [YES / NO]
- 50, do you accept or reject? [YES / NO]
- 60, do you accept or reject? [YES / NO]
- 70, do you accept or reject? [YES / NO]
- 80, do you accept or reject? [YES / NO]
- 90, do you accept or reject? [YES / NO]
- 100, do you accept or reject? [YES / NO]

(for High types)

You are Person B. Initial endowment of person A is 100 experimental points.

As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and in this situation you will have to face Person A who has a HIGH level of experience as well. Person A sends you:

- 0, do you accept or reject? [YES / NO]

- 10, do you accept or reject? [YES / NO]
- 20, do you accept or reject? [YES / NO]
- 30, do you accept or reject? [YES / NO]
- 40, do you accept or reject? [YES / NO]
- 50, do you accept or reject? [YES / NO]
- 60, do you accept or reject? [YES / NO]
- 70, do you accept or reject? [YES / NO]
- 80, do you accept or reject? [YES / NO]
- 90, do you accept or reject? [YES / NO]
- 100, do you accept or reject? [YES / NO]

Situation 4

There are two players in the game: Person A and Person B.

(for Low types)

Person A has LOW experience and person B has LOW experience as well.

(for High types)

Person A has HIGH experience and person B has HIGH experience as well.

(for Low types)

You are Person A and your initial endowment is 50 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a LOW level of experience as well.

How much of your initial endowment of 50 experimental points do you want to send to Person B (in 10 points increments)?

(for High types)

You are Person A and your initial endowment is 50 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person B who has a HIGH level of experience as well.

How much of your initial endowment of 50 experimental points do you want to send to Person B (in 10 points increments)?

(for Low types)

You are Person B and the initial endowment of Person A is 50 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person A who has a LOW level of experience as well. If Person A sends you:

- 10 (so you receive 30). How many points do you send back in increments of 10 experimental points ?__
- 20 (so you receive 60). How many points do you send back in increments of 10 experimental points? __
- 30 (so you receive 90). How many points do you send back in increments of 10 experimental points? __
- 40 (so you receive 120). How many points do you send back in increments of 10 experimental points? __
- 50 (so you receive 150). How many points do you send back in increments of 10 experimental points? __

(for High types)

You are Person B and the initial endowment of Person A is 50 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person A who has a HIGH level of experience as well. If Person A sends you:

- 10 (so you receive 30). How many points do you send back in increments of 10 experimental points? __
- 20 (so you receive 60). How many points do you send back in increments of 10 experimental points? __
- 30 (so you receive 90). How many points do you send back in increments of 10 experimental points? __
- 40 (so you receive 120). How many points do you send back in increments of 10 experimental points? __
- 50 (so you receive 150). How many points do you send back in increments of 10 experimental points? __

Phase 3 – [C3 condition]

As in condition C1 plus information about the other player's level of experience (different level).

Now starts a new phase of the experiment.

(for Low types)

You have a LOW level of experience in laboratory, that is, you have participated in few experiments and, in this part of the experiment you will be paired in each situation with a different person who has HIGH level of experience, who will also get a reward that will depend on your own choice or both of you choices, depending on the situation. You will not be told who you are matched with during or after the experiment, and he or she will not be told who you are either during or after the experiment.

Control questions.

Your level of experience is...

The level of experience of your partner is...

(for High types)

You have a HIGH level of experience in laboratory, that is, you have participated in many experiments and, in this part of the experiment you will be paired in each situation with a different person who has LOW level of experience, who will also get a reward that will depend on your own choice or both of you choices, depending on the situation. You will not be told who you are matched with during or after the experiment, and he or she will not be told who you are either during or after the experiment.

Control questions.

Your level of experience is...

The level of experience of your partner is...

Situation 1

(for Low types)

As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a HIGH level of experience, that is, a subject who has participated in many experiments.

Now it is time to make your decision. How much do you want transfer to Person B? ___

(for High types)

As you know, you have a HIGH level of laboratory experience, that is, you have participated in

many experiments and, in this situation you will have to face Person B who has a LOW level of experience, that is, a subject who has participated in few experiments.

Now it is time to make your decision. How much do you want transfer to Person B? ___

Situation 2

There are two players in the game: you and another player you are paired with.

(for Low types)

You have LOW experience and the other player has HIGH experience.

Now look the table below and make your decision. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face a player with a HIGH level of experience, that is, a subject who has participated in many experiments.

Other

		Cooperate	**Defect**
You	**Cooperate**	60, 60	30, 90
	Defect	90, 30	40, 40

Which action would you like to play, Action C or Action D? ____

(for High types)

You have HIGH experience and the other player has LOW experience.

Now look the table below and make your decision. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face a player with a LOW level of experience, that is, a subject who has participated in few experiments.

Other

		Cooperate	**Defect**
You	**Cooperate**	60, 60	30, 90
	Defect	90, 30	40, 40

Which action would you like to play, Action C or Action D? ____

Situation 3

There are two players in the game: Person A and Person B.

(for Low types)

Person A has LOW experience and person B has HIGH experience.

(for High types)

Person A has HIGH experience and person B has LOW experience.

(for Low types)

You are Person A. Your initial endowment is 100 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a HIGH level of experience, that is, she/he has participated in many experiments.

How much of your initial endowment do you want to send to Person B (in 10 points increments)?

(for High types)

You are Person A. Your initial endowment is 100 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person B who has a LOW level of experience, that is, she/he has participated in few experiments.

How much of your initial endowment do you want to send to Person B (in 10 points increments)?

(for Low types)

You are Person B. Initial endowment of person A is 100 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and in this situation you will have to face Person A who has a HIGH level of experience, that is, she/he has participated in many experiments.

Person A sends you:

- 0, do you accept or reject? [YES / NO]
- 10, do you accept or reject? [YES / NO]
- 20, do you accept or reject? [YES / NO]
- 30, do you accept or reject? [YES / NO]
- 40, do you accept or reject? [YES / NO]
- 50, do you accept or reject? [YES / NO]
- 60, do you accept or reject? [YES / NO]
- 70, do you accept or reject? [YES / NO]
- 80, do you accept or reject? [YES / NO]

- 90, do you accept or reject? [YES / NO]
- 100, do you accept or reject? [YES / NO]

(for High types)

You are Person B. Initial endowment of person A is 100 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and in this situation you will have to face Person A who has a LOW level of experience, that is, she/he has participated in few experiments. Person A sends you:

- 0, do you accept or reject? [YES / NO]
- 10, do you accept or reject? [YES / NO]
- 20, do you accept or reject? [YES / NO]
- 30, do you accept or reject? [YES / NO]
- 40, do you accept or reject? [YES / NO]
- 50, do you accept or reject? [YES / NO]
- 60, do you accept or reject? [YES / NO]
- 70, do you accept or reject? [YES / NO]
- 80, do you accept or reject? [YES / NO]
- 90, do you accept or reject? [YES / NO]
- 100, do you accept or reject? [YES / NO]

Situation 4

There are two players in the game: Person A and Person B.

(for Low types)

Person A has LOW experience and person B has HIGH experience.

(for High types)

Person A has HIGH experience and person B has LOW experience.

(for Low types)

You are Person A and your initial endowment is 50 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person B who has a HIGH level of experience that is, she/he has participated in many experiments.

How much of your initial endowment of 50 experimental points do you want to send to Person B (in 10 points increments)?

(for High types)

You are Person A and your initial endowment is 50 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person B who has a LOW level of experience that is, she/he has participated in few experiments.

How much of your initial endowment of 50 experimental points do you want to send to Person B (in 10 points increments)?

(for Low types)

You are Person B and the initial endowment of Person A is 50 experimental points. As you know, you have a LOW level of laboratory experience, that is, you have participated in few experiments and, in this situation you will have to face Person A who has a HIGH level of experience, that is, she/he has participated in many experiments. If Person A sends you:

- 10 (so you receive 30). How many points do you send back in increments of 10 experimental points? __
- 20 (so you receive 60). How many points do you send back in increments of 10 experimental points? __
- 30 (so you receive 90). How many points do you send back in increments of 10 experimental points? __
- 40 (so you receive 120). How many points do you send back in increments of 10 experimental points? __
- 50 (so you receive 150). How many points do you send back in increments of 10 experimental points? __

(for High types)

You are Person B and the initial endowment of Person A is 50 experimental points. As you know, you have a HIGH level of laboratory experience, that is, you have participated in many experiments and, in this situation you will have to face Person A who has a LOW level of experience, that is, she/he has participated in few experiments.

If Person A sends you:

- 10 (so you receive 30). How many points do you send back in increments of 10 experimental points? __
- 20 (so you receive 60). How many points do you send back in increments of 10 experimental points? __
- 30 (so you receive 90). How many points do you send back in increments of 10 experimental points? __
- 40 (so you receive 120). How many points do you send back in increments of 10 experimental points? __
- 50 (so you receive 150). How many points do you send back in increments of 10 experimental points? __

Bibliography

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69, 334-345.

Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N., and Gangadharan, L. (2009). Subject pool effects in a corruption experiment: A comparison of Indonesian public servants and Indonesian students. *Experimental Economics*, 12(1), 113-132.

Allen, J. S., and Muchinsky, P. M. (1984). Assessing raters' policies in evaluating proposed services for transporting the physically handicapped. *Journal of Applied Psychology*, 69, 3-11.

Anderson, J., Burks, S. V., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D., Potter, R., Rocha, K., and Rustichini, A. (2013). Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. *Experimental Economics*, 16 (2), 170-189.

Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602-14.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, 11, 122-133.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer C., and Sugden R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

Bassok, M., Wu, L. L., and Olseth K. L. (1995). Judging a book by its cover: interpretative effects of content on problem solving transfer. *Memory and Cognition*, 23, 354-67.

Bellemare, C. and Kroger, S. (2007). On representative social capital. *European Economic Review*, 51, 183–202.

Belot, M., Duch, R. M., and Miller, L. M. (2010). Who Should be Called to the Lab? A comprehensive comparison of students and non-students in classic experimental games. http://cess-wb.nuff.ox.ac.uk/documents/DP2010/CESS_DP2010_001.pdf

Benson, B. L., and Faminov, M. D., (1988). The impact of experience on prices and profits in experimental duopoly markets. *Journal of Economic Behavior and Organization*, 9(4), 345-365.

Binmore, K. (1994). *Playing Fair: Game Theory and the Social Contract I*. Cambridge MA: MIT Press.

Bohnet, I., and Frey, B. (1999). Social distance and other-regarding behavior in dictator games: comment. *American Economic Review*, 89, 335–339.

Bolton, G. E. (1991). A comparative model of bargaining: theory and evidence. *American Economic Review*, 81(5), 1096-1136.

Bracht, J., and Regner, T. (2011). Moral emotions and partnership. *Jena Economic Research Papers*, 2011-028.

Brañas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior and Organization*, 60, 306-320.

Brosig, J. (2002). Identifying cooperative behavior: some experimental results in a prisoner's dilemma game, *Journal of Economic Behavior and Organization*, (47), 275–290.

Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior and Organization*, 50, 133-144.

Calder, B. J., Phillips, L.W., and Tybout A. M. (1981). Design Research for Application. *Journal of Consumer Research*, 8, 197-207.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton Press.

Camerer, C. F., and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–42.

Cappelen, A. W., Nygaard, K., Sørensen, E. Ø., and Tungodden, B. (2015). Social Preferences in the Lab: A Comparison of Students and a Representative Population. *The Scandinavian Journal of Economics*, 117(4), 1306-1326.

Capraro, V., and Cococcioni, G. (2015). Social setting, intuition, and experience in lab experiments interact to shape cooperative decision-making. *Proceedings of the Royal Society B: Biological Sciences*, 282(1811).

Carbone, E. (2005). Demographics and Behaviour. *Experimental Economics*, 8(3), 217-232.

Carlson, R. (1971). Where Is the Person in Personality Research? *Psychological Bulletin*, 75(3), 203-

219.

Carpenter, J., Burks, S., and Verhoogen, E. (2005). Comparing students to workers: The effects of social framing on behavior in distribution games. In Field Experiments in Economics (Research in Experimental Economics), Carpenter, J., Harrison, G., List, J. (editors), JAI/Elsevier, Greenwich, CT, and London, 261–290.

Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic Behavior in a Representative Dictator Experiment. *Experimental Economics*, 11, 282-298.

Christensen-Szalanski, J. J. J., Beck, D. E., Christensen- Szalanski, C. M., and Koepsell, T. D. (1983). Effects of expertise and experience on risk judgments. *Journal of Applied Psychology*, 68, 278-284.

Churchill, N. C., and Cooper, W. W. (1971). Effects of auditing records: Individual task accomplishment and organizational objectives. In W. M. Evan (Ed.), Organizational experiments: Laboratory and field research. Chicago: Rand McNally.

Cleave, B., Nikiforakis, N., and Slonim, R. (2011). Is there selection bias in laboratory experiments? The Case of Social and Risk Preferences. *Experimental Economics*, 16(3), 372-382.

Conte, A., Levati, M. V., and Montanari, N. (2014). Experience in Public Goods Experiments. http://pubdb.wiwi.uni-jena.de/pdf/wp_2014_010.pdf.

Cooper, D. J., Kagel, J. H., Lo, W., and Gu, Q. L. (1999). Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers. *American Economic Review*, 89(4), 781-804.

Cornelius, E. T., DeNisi, A. S., and Blencoe, A. G. (1984). Expert and naïve raters using the PAQ: Does it matter? *Personnel Psychology*, 37, 453-464.

Danielson, A. J. and Holm, H. J. (2007). Do You Trust Your Brethren? Eliciting Trust Attitudes and Trust Behavior in a Tanzanian Congregation. *Journal of Economic Behavior and Organization*, 62(2), 255–271.

Davis, D. D., and Holt, C. A. (1993). *Experimental economics*. Princeton University Press.

Eckel, C., and Grossman, P. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, 16, 181–191.

Eckel, C., and Grossman P. (2000). Volunteers and Pseudo-Volunteers: The Effect of Recruitment

Method in Dictator Experiments. *Experimental Economics*, 3(2), 107-120.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, 14, 583-610.

Exadaktylos, F., Espín, A. M., and Brañas-Garza, P. (2013). Experimental subjects are not different. *Scientific Reports*, 3, 1213.

Falk, A., and Heckman J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Scienc*e, 326, 535-538.

Falk, A., Meier, S., and Zehnder, C. (2010). Do Lab Experiments Misrepresent Social Preferences? The case of self-selected student samples. *Journal of the European Economic Association*, 11(4), 839-852.

Fehr, E., and List J. A. (2004). The Hidden Costs and Returns of Incentives. Trust and Trustworthiness among CEOs. *Journal of the European Economic Association*, 2(5), 743-771.

Fleming, J. E. (1969). Managers as subjects in business decision research. *Academy of Management Journal*, 12, 59-66.

Fréchette, G. R. (2015). Laboratory Experiments: Professionals versus Students. In Handbook of Experimental Economic Methodology, Guillaume R. Fréchette and Andrew Schotter (editors), Oxford University Press, February, 360-390.

Friedman, M. (1953). *The Methodology of Positive Economics*. In Essays in Positive Economics. Chicago: University of Chicago Press, 3-43.

Friedman, D., and Cassar, A. (2004) *Economics Lab: An intensive course in experimental economics*, Routledge, 2004.

Gintis, H. (2001).The contribution of game theory to experimental design in the behavioral sciences. In Commentary/Hertwig & Ortmann: Experimental practices in economics: A challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383-451.

Gneezy, U., and Potters, J. (1997). An Experiment on Risk Taking and Evaluation Periods. *Quarterly Journal of Economics*, 631- 645.

Gordon, M. E., Slade, L. A., and Schmitt, N. (1986). The Science of the Sophomore Revised: From Conjecture to Empiricism. *Academy of Management Review*, 11(1), 191-207.

Greenberg, J. (1987). The College Sophomore as Guinea Pig: Setting the Record Straight. *Academy of Management Review*, 12, 157-159.

Grether, D. M. (1980). Bayes rule as a description model: The representativeness heuristic. *Quarterly Journal of Economics*, 95, 537–557.

Guala, F. (2005). *The Methodology of Experimental Economics*. Cambridge University Press.

Guala, F., and Mittone, L. (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology*, 12(4), 495-515.

Guala, F., and Mittone, L. (2010). Paradigmatic experiments: The Dictator Game. *The Journal of Socio-Economics*, 39, 578–584.

Guala, F. (2006). *Filosofia dell'economia: modelli, casualità, previsione*. Il Mulino.

Guala, F. (2008). Experimentation in Economics. In Handbook of the Philosophy of Science, Uskali Mäki (editor), Elsevier, 597-640.

Guillén, P., and Veszteg, R. F. (2006). Social Pool Bias in Economic Experiments. http://www.ugr.es/~teoriahe/RePEc/gra/wpaper/thepapers06_03.pdf.

Guillén, P., and Veszteg, R. F. (2012). On "lab rats". *The Journal of Socio-Economics*, (41), 714-720.

Güth, W., Ploner, M., and Regner, T. (2009). Determinants of in-group bias: Is group affiliation mediated by guilt-aversion? *Journal of Economic Psychology*, 30, 814–828.

Güth, W., and Kocher M. (2014). More than thirty years of ultimatum bargaining experiments:Motives, variations, and a survey of the recent literature, *Journal of Economic Behavior and Organization*, (108), 396-409.

Haigh M. S. and List J. A. (2005). Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis. The Journal of Finance, 60(1, 523–534.

Hakel, M. D., Dobmeyer, T. W., and Dunnette, M. D. (1970a). Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. *Journal of Applied Psychology*, 54, 65-71.

Hakel M. D., Hollmann, T. D., and Dunnette, M. D. (1970b). Accuracy of interviewers, certified public accountants, and students in identifying the interests of accountants. *Journal of Applied Psychology*, 54, 115-119.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an

anonymous economic game. *Evolution and Human Behavior*, 26, 245–256.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62, 1251–1289.

Harrison, G. W., McKee, M., and Rutström, E. E. (1987). Experimental evaluations of institutions of monopoly restraints. In Advances in Behavioral Economics, Green, L., Kagel, J. H. (editors), Ablex Press, Norwood, N.J., 54-94.

Harrison, G. W. (1989). Theory and misbehavior of first-price auctions. *American Economic Review* 79, 749–762.

Harrison, G. W. (1992). Theory and misbehavior of first-price auctions: Reply. *American Economic Review*, 82, 1426–1443.

Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., and Tracer, D. (2005). Economic man in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795-815.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83.

Hertwig, R., and Ortmann, A. (2001). Experimental practices in Economics: A challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383-451.

Hey, J. D., 1991. Advances in behavioral economics. *Journal of Economic Psychology*, 12(4), 725-726.

Hey, J. D. (2001). Does Repetition Improve Consistency? *Experimental Economics*, 4, 5–54.

Hochman, G., Ayal, S., and Ariely D. (2014). Keeping your gains close but your money closer: The prepayment effect in riskless choices. *Journal of Economic Behavior & Organization*, 582–594.

Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, 86, 653–660.

Holt, C. A., and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644-1655.

Isaac, M., Walker J. M., and Thomas S. H. (1984). Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations. *Public Choice*, 43, 113-149.

Isaac, M., McCue K., and Plott, C. (1982). Nash equilibrium in public goods provision: Free riding in experimentally controlled markets. California Institute of Technology Social Science Working Paper No. 428.

Jago, A. G., and Vroom, V. H. (1982). Sex differences in the incidence and evaluation of participative leader behavior. *Journal of Applied Psychology*, 67, 776-783.

Jamal, K., and Sunder, S. (1991). Money vs. gaming: Effects of salient monetary payments in double oral auctions. *Organizational Behavior and Human Decision Processes,* 49, 151–166.

Johnson, N.D. and Mislin, A. (2011). Trust games: A meta-analysis, *Journal of Economic Psychology*, (32), 865–889

Kagel, J. H., and Levin, D. (1986). The Winner's Curse and Public Information in Common Value Auctions. *American Economic Review*, 76(5), 894-920.

Kahneman, D., Knetsch, J. L., and Thaler R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98(6), 1325-1348.

Kavanagh, M. J. (1975). Expected supervisory behavior, interpersonal trust and environment preferences: Some relationships based on a dyadic model of leadership. *Organizational Behavior and Human Performance*, 13, 17-30.

Kim, O., and Walker, M. (1984). The free rider problem: Experimental evidence. *Public Choice,* 43(1), 3-24.

Klempt, C., and Pull, K. (2009). Generosity, greed and gambling: What difference does symmetric information in bargaining make? *Jena Economic Research Papers*, 2009-021.

Kraus, S. J. (1995). Attitudes and the Prediction of Behavior: A Meta-analysis of the Empirical Literature. *Personality and Social Psychology Bulletin*, 21, 58–75.

Ledyard, J. (1995). Public goods: a survey of experimental research. In J. Kagel & A. Roth (eds.) The Handbook of Experimental Economics, Princeton: Princeton University Press, 111-194.

Lipsey, R. (1979). *An Introduction to Positive Economics*, 5th edition, London: Weidenfeld and Nicholson.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115, 482-493.

Loewenstein G. (1999). Experimental economics from the viewpoint of behavioural economics. Symposium on 'Experimental economics' in *Economic Journal*, F25-F34.

Marwell, G., and Ames, R. E., (1980). Experiments on the provision of public goods. II. Provision points, stakes, experience and the free rider problem. *American Journal of Sociology*, 85(4), 926-937.

Matthey, A., and Regner, T. (2013). On the independence of history: experience spill-overs between experiments. *Theory and Decision*,75, 403-419.

McNemar, Q. (1946). Opinion-Attitude Methodology. *Psychological Bulletin*, 43, 289-374.

Mill, J. S. (1836). On the Definition of Political Economy and the Method of Investigation Proper to It. In Collected Works of John Stuart Mill, Vol. 4. Toronto: University of Toronto Press, 1967, 120-64.

Moskowitz, H. (1971). Managers as partners in business decision research. *Academy of Management Journal*, 14, 317-325.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal, R.L. Rosnow (Eds.). Artifact in Behavior Research. Academic Press (1969), 147–181.

Pelligra, V., and Stanca, L. (2013). To give or not to give? Equity, efficiency and altruistic behavior in an artefactual field experiment. *The Journal of Socio-Economics*, 46, 1-9.

Peterson, R. A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis. *Journal of Consumer Research*, 28, 450-461.

Plott, C. (1991). Will economics become an experimental science? *Southern Economic Journal*, 57(4), 901-919.

Plott, C. (1996). Rational individual behaviour in markets and social choice processes: the discovered preference hypothesis. In The Rational Foundations of Economic Behaviour (ed. K. J. Arrow, E. Colombatto, M. Perlam, and C. Schmidt). Basingstoke, U.K.: Macmillan.

Regner, T., and Harth, N. S. (2010). Testing belief-dependant models. *Jena Economic Research Papers*, 2010-072.

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. MacMillan and CO., Limited St. Martin's Street, London, 1945.

Rosenthal, R. W., and Rosnow, R. L. (1969). *Artifact in Behavioral Research*. New York: Academic

Press.

Rosnow, R. L. and Rosenthal, R. W. (1997). People studying people: artifacts and ethics in behavioral research. New York: W.H. Freeman.

Roth. A. E. (1986). Laboratory experimentation in economics. *Economics and Philosophy*, 2, 245-273.

Roth. A. E. (1993). On The Early History of Experimental Economics. *Journal of the History of Economic Thought*, 15, 184-209.

Roth. A. E. (1995) Introduction to experimental economics. In: The handbook of experimental economics, ed. J. H. Kagel & A. E. Roth. Princeton University Press.

Schneider, W. G. (1982). The effects of research complexity and tyre of subjects on the results of laboratory research in the field of union management relations. Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12, 225-237.

Schultz, D. P. (1969). The Human Subject in Psychological Research. *Psychological Bulletin*, 72, 214-228.

Sears, D. O. (1986). College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology*, 51(9), 515-530.

Seithe, M. (2012). Introducing the Bonn Experiment System.

http://boxs.uni-bonn.de/boxs_seithe.pdf.

Shapley H. (1964). *Of Stars and Men*. Boston.

Siakantaris, N. (2000). Experimental economics under the microscope. *Cambridge Journal of Economics*, 23, 267-281.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review Proceedings*, 66, 274–279.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72, 923-938.

Smith, V.L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99, 877–897.

Smith, V. L. and Walker, J. M. (1993a). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31, 245–261.

Smith, V. L. and Walker, J. M. (1993b). Rewards, experience and decision costs in first price auctions. *Economic Inquiry*, 31, 237–245.

Starmer, C. (1999). Experiments in economics: should we trust the dismal scientists in white coats? *Journal of Economic Methodology*, 6(1), 1-30.

Stow, B. M., and Ross, J. (1980). Commitment in an experimenting society: A study of the attribution of leadership from administrative scenarios. *Journal of Applied Psychology*, 65, 249-260.

Thaler, R. H., Tversky, A., Kahneman, D., and Schwartz, A. (1997). The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test. *The Quarterly Journal of Economics*, 647-661.

Thurstone, L.L. (1931). The Indifference Function. Journal of Social Psychology, 2, 139-167.

Wallis, W. A. and Friedman M. (1942). The Empirical Derivation of Indifference Functions. In Studies in Mathematical Economics and Econometrics in memory of Henry Schultz, O. Lange, F. McIntyre, and T.O. Yntema, editors, Chicago, University of Chicago Press, 175-189.

Xue, L., Sitzia, S. and Turocy, T. L. (2015). Mathematics self-confidence and the "prepayment effect" in riskless choices. https://www.uea.ac.uk/documents/166500/0/CBESS+15-20.pdf/31995582-7f0a-4ab6-bf72-7b53a9e6cab5

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13(1), 75-9