Doctoral thesis

UNIVERSITY OF TRENTO

School of Social Sciences
PhD program in Economics and Management

Three essays in experimental and labor economics

Imitation, Information aversion, Gender discrimination

A DISSERTATION SUBMITTED TO THE DOCTORAL PROGRAM IN ECONOMICS AND MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DOCTORAL DEGREE (PH.D.) IN ECONOMICS AND MANAGEMENT

Simona
Cicognani

January 2015

Supervisor

Professor Luigi Mittone
Università degli Studi di Trento

Internal Evaluation Commission

Professor Alessandro Rossi
Università degli Studi di Trento

Professor Alessandro Innocenti
Università di Siena

Examination Committee

TBC

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deep gratitude to a number of people, without whom this work would have been impossible.

First of all, I dearly thank my parents, my brother, Laura and Paola, for their understanding, love, patience and faith in me and in my choices. *Grazie cari!*

A very special thank goes to Marco, for his priceless support during these years.

I am grateful to my supervisor, Professor Luigi Mittone, for introducing me to the fascinating field of behavioral and experimental economics, and for constantly providing me guidance over these years.

My deepest gratitude and appreciation goes also to Professor Marie Claire Villeval, who gave me the opportunity of a visiting research period at GATE, Lyon. Working with her during my French adventure has been a great honor: she is a great scholar, as well as a great human being.

I acknowledge important comments and feedback from IRVAPP scholars and researchers, in particular from Daniele Checchi and Enrico Rettore. Thanks also for letting me taking part to excellent seminars and schools. I owe sincere gratitude to Nevena Kulic, coauthor of mine in one of the chapters, for her helpful insights and determination.

I gratefully acknowledge IRVAPP, the University of Trento, the School of Social Sciences of Trento, the Region Rhone-Alpes, ANR, LABEX CORTEX and CNRS for their financial support.

CEEL in Trento and GATE in Lyon have been great and welcoming environments to work in. I am particularly grateful to Matteo, Dominique, Marco, Sandro, Davide, Federico; Virginia, Fabio, Mateus, Nadege, Nicolas, Stephane, Frederic, Laurine, Pauline, Andrea, Damien, Zhixin, M'hamed, Tai, Nelly and Aude. Thank you for your moral and working support. I also wish to thank two colleagues of mine from the Ph.D. in Trento: Michele and in particular Simone, for her precious help and friendship throughout these years.

Abstract

This thesis collects three independent essays. Two of them relate to experimental economics, and one to labor economics. The first essay explores through a laboratory experiment the relationship between cognitive costs and imitation dynamics. The second essay investigates in an experimental setting information aversion towards bad/good news about ones' own condition, and tests the effect of the possibility of exerting effort, which may improve one's own condition, on the willingness to acquire more detailed personal information. The third essay proposes an empirical application aiming at studying gender discrimination in a research context, and in particular whether the gender composition of evaluating commissions affects the hiring of women in research activities.

Keywords. Beliefs; Cognitive Costs; Decision Making; Laboratory Experiments; Gender quotas; Discrimination; Research Recruitment; Connections; Information Aversion; Information Acquisition; Signals

JEL classification: C91; D80; D81; D82; D83; J16; J71; J45;

1 Introduction

This Ph.D. thesis presents three essays that investigate different topics in economics. Each essay focuses on a different subject, therefore each one can be read as an independent contribution to the thesis. Two out of the three essays employ behavioral and experimental economics methodologies to investigate individuals' preferences and deviations from the standard economic paradigm, in two different contexts. One deals with the consideration of cognitive costs, contrary to traditional economic theories of rational behavior that disregard them. The other experimental essay examines an informational framework in which individuals may derive utility from beliefs, as opposed to the standard paradigm of the economics of information, which considers information as valuable only as a means for improving decision making (Stigler, 1961). The third essay lies outside the experimental domain and proposes an empirical investigation in a labor market scenario.

Albeit the thesis encompasses three essays that can be read independently, a few intertwined themes combining the single contributions can be envisaged. These refer to the following keywords: beliefs, signals, costs, and gender.

The underlying motivation of the dissertation is to investigate how individuals undertake decisions in contexts increasingly complex and characterized by uncertainty. From a methodological perspective, the natural choice was to apply the techniques of behavioral economics, which " replaces strong rationality assumptions with more realistic ones and explores their implications"(Camerer, 2003b, p. 1673).

The role of beliefs in every day life is central. Human beings form beliefs about themselves both in terms of skills and qualities, and in terms of future prospects in life, as for instance financial or personal realization. They also form beliefs about the surrounding world, about the behavior of other people in the society and about whether they can trust them. Beliefs shape individuals' decisions and are hence strongly connected to both individual and collective performance. As a consequence, we can think of beliefs as assets that decision makers invest in, protect, value and use (Bénabou and

Tirole, 2011).

Sometimes beliefs are coupled with misbeliefs. Overconfidence and the better-than-average effect are a clear example of a common misbelief. 90% of U.S. drivers deem to be above-average in terms of quality (Svenson, 1981), and 94% of professors at a large U.S. University think to be better than their average fellow colleagues at their own institution (Cross, 1977; Price, 2006). Moving to the health domain, individuals show an optimistic bias as they consider themselves to be less likely than their peers to experience adverse life situations. Weinstein (1982) for instance documents that 100 college students who were asked to compare their chances of experiencing 45 different health and life problems consistently consider their own probabilities to be below average in 34 of the 45 circumstances.
It hence emerges that beliefs and misbeliefs must be considered as important drivers of individual decision making and should be incorporated in the economic modeling.

In the first essay beliefs are a key component of the experimental setting. Subjects face a novel experimental task, whose solution is not the result of computation or other technical skills. They are asked to find a solution to the task, knowing that the better the solution, the higher their final payoff will be. In solving the task they are provided with two kinds of default options that would allow them to skip the task and refer to a decisional short-cut. In one of these default options participants know that they would earn an above-average payoff, but they do not know exactly whether the payoff will be exactly the maximum available or just the average. In the other default option, participants, on top of this above-average information, are informed that the default option has been chosen by the majority of participants in a previous experimental session. It emerges that individuals' decisions will strongly hinge on their beliefs about the default options. We implement an incentivized belief elicitation task in order to retrieve data about beliefs of the quality of the two different default options. We find that the majority default option is considered as the best option when the task to solve is more complex.

Another way in which beliefs enter the first essay is through subjects' under/over-confidence regarding their own absolute and relative performance. By asking subjects at the end of the experiment to indicate how accurate they think their final choice was and to compare it with the majority of other participants, we are able to examine whether participants relying on default options are guided by under/over-confidence about their own skills. We document that subjects are on average overconfident both in relative and in absolute terms, and that subjects who opt for the default options are significantly underconfident and perceive the task as more difficult.

The second essay builds on the idea that people may derive utility from beliefs, and as a consequence they may manipulate sources of available information. It follows that if some information carries a disutility to subjects, an informational trade-off arises between emotional costs deriving from more information and better decisions accruing from more information.

In the experiment participants know they can either be attached to a good or a bad condition, resulting in higher or lower payoffs, respectively. In the first session of the experiment, which develops across a temporal dimension of two weeks, participants receive preliminary signals about their condition and form their initial beliefs. They are given the opportunity of acquiring additional information in order to update their beliefs regarding their chances of being in the good or in the bad condition. If they buy it, the additional information entails lower uncertainty regarding their condition. Beliefs about being in the good/bad condition are computed through the Bayes formula, provided the different signals participants can receive in the first session. However, these theoretical probabilities should be matched with actual beliefs reported by subjects, which will be elicited in future experimental sessions.[1]

We find that subjects are willing to buy more detailed information when their Bayesian theoretical belief of being in the bad condition is higher. This result deviates from the existing literature on information aversion and could be due to the inclusion in the experimental setting of the possibility of improving ones' own condition through the exertion of effort between the first and the second experimental session.

Beliefs are strongly related to signals. Signals occupy the uncertain world we live in, and give rise to beliefs in individuals. It is important to correctly interpret the signals received.

Experimental participants in the first essay receive signals about the quality of default options: in one case they are informed that the default option belongs to the upper half of available options in terms of quality, and in the other default option, on top of this information, they are given a social signal regarding the default option. Moreover, in the first ten seconds of the experiment we provide subjects with uncovered cards, which then become covered and will be inspected in order to solve the experimental task. The rationale of the first ten seconds is to offer subjects a signal of the type of task and of its difficulty.

The role of signals in the second essay is played by balls of different colors that are extracted from

[1]For a detailed review on the various methods to elicit beliefs, see Schlag et al. (2014).

two possible boxes. On the one hand, the box representing the good condition has a higher fraction of white balls, which are connected with a higher payoff in case of final extraction. On the other hand, the bad box encompasses a higher percentage of black balls, which entail a lower payoff than the white ones if they are extracted. Participants first receive a signal deriving from an extraction of three balls from their box, and given this extraction, they can decide to spend 1 Euro for the extraction of two additional balls from their box. On average, roughly 30% of participants chose to receive more accurate signals about their own conditions.

Beliefs and signals are present, from a different perspective, also in the empirical application of the third essay, which examines an Italian labor market case study. Signalling in labor markets has received much attention since the seminal work of Spence (1973). Candidates in the data set used in the third essay apply for research positions and are evaluated by commissions, which consider candidates' signals of quality as indicated in their *curricula*. The signals can encompass applicants' education marks, whether they earned a Ph.D. and the number, level and impact of their scientific productivity, which can be captured by their h-index. Another important signal about candidates' skills could be represented by prior or current work experience at the hiring institution. Commissioners, given the signals of candidates' quality, form their beliefs and choose whether to select or not the candidate.

The perspective from which we consider signals in the third essay is that of a possible discrimination of candidates. Given the scarce representation of females in research positions, especially at top positions, we examine whether female candidates are discriminated in recruitment processes. Therefore, we test whether female and male candidates who provide the same signals (observable characteristics) in the job market, are treated differently or not, according to their gender. Our results seem to support gender discrimination against female applicants in calls for lower research positions, and we also find that this effect could be mitigated by imposing one female researcher in the evaluating commission.

Another common dimension of the essays of this thesis is represented by costs. Many decisions involve costs, of different kinds. In the thesis I mainly consider cognitive and pecuniary costs.

Cognitive costs constitute the core of the first essay. The idea in the first essay is to model cognitive costs through a novel experimental task, and connect them with the availability of a social default option. We examine whether a previous majority choice can be conceived as a low-cost heuristic. Results show that the experimental task is able to model lower and higher cognitive costs. In the task which we hypothesized to be more difficult, subjects' performance rates are actually lower. The rationale of the novel task is to offer subjects an environment in which they do not have a reference point, either computational or logical. The difficulty depends on the different visualization of figures in the two scenarios and does not require mathematical skills.

A pecuniary kind of cost is present in the second essay, in which subjects can subtract 1 Euro from their final payoffs in case they are willing to receive a more precise signal of their own condition in the first session of the experiment. From a rational perspective, the additional extraction of two balls from participants' boxes after the payment of 1 Euro is not connected to any difference in the expected value of the experiment. Each participant is attached to a good or a bad box, and at the end of the experiment one ball is extracted from each box. Having received a more detailed signal about one's own condition does not impact the type of box attached to each individual. However, participants can exert effort in order to improve their condition between the first and the second experimental session, and the exertion of effort is connected to a higher reward in case the participant is in a bad condition. Therefore, paying the additional Euro for a more accurate signal may derive from a strategy of knowing whether effort is worth doing or not.

Finally, another theme of the thesis is gender. Gender is the central issue of the third essay, in which we test for the presence of gender discrimination and whether affirmative actions may reduce it. But gender is also a recurring topic in behavioral economics, as men and women tend to behave differently under certain circumstances.

In the first and the second essay we are able to compare the behavior of females versus that of males in different respects. Males slightly outperform females in the solution of the task of the first essay, and they are significantly more overconfident in relative terms compared to females. This is an interesting result, and corroborates the existing literature of gendered relative overconfidence (see Niederle and Vesterlund (2007)). When asked about their absolute performance, female participants in the first experiment appear more overconfident than males about their skills, but differences are not statistically significant. Conversely, when asked about whether they think to have performed better, equal or worse than the majority of participants, women exhibit lower overconfidence than men.
The gendered behavioral difference in the second essay refers to the completion of effort between the first and the second experimental session. Interestingly, we find that women exert significantly less effort than their male colleagues, even if they all receive the same information about the rewards deriving from effort exertion. Yet, we do not find significant differences across genders in the other

dimensions involved in the experiment, such as for instance risk aversion or procrastination propensity.

In the remaining part of this general introduction to the thesis I briefly introduce the different essays and their contribution to the economic literature.

In the first essay we aim to extend choice theory by allowing for the interaction between cognitive costs and imitative dynamics.[2] The starting point of the study lies in the fact that, in many circumstances, economic actors are boundedly rational and make use of simplifying heuristics, either conscious or unconscious, when they process information that carries cognitive costs (Gigerenzer and Gaissmaier, 2011). Traditional economic theories of rational behavior disregard cognitive costs and assume that economic agents process costly information fairly easily, since they are always able to select the utility-maximizing option among different ones. In contrast, evidence gathered by some psychologists and economists supports the idea that decision makers systematically violate the assumptions of rational choice theory (Tversky and Kahneman, 1974; Camerer, 2003a; Bicchieri, 2006). In this essay we investigate as a potential cognitive shortcut faced by decision makers the influence of other agents in their reference community. Imitation is a very frequent pattern of real world behaviors but its sources have been limitedly examined by economic theory. Imitation is crucial in the transmission of knowledge and represents one of the main sources of learning.

We experimentally examine the role of imitation when participants face a task which is costly in cognitive terms. To do this, we devise a novel experimental task in which we model the choice of different alternatives through high or low cognitive costs. The task is not purely mechanical and its rationale is to offer subjects an environment in which, while searching the solution, they do not have a reference point, either computational or logical: the assignment is not related to an univocally clear solution, characterized by normative determinants. Our approach aims indeed at resembling a dynamic social environment, where available options must be discovered, but it is not straightforward which is the best one among them. We show participants a set of abstract figures, and we model the difficulty of such figures into high or low. In these two different scenarios, we test the impact of information available to participants: whether they know or not the choices of a majority of subjects who took part in a previous experimental session.

Our results show that the experimental task implemented allows us to carry out a scenario with

[2]This chapter is co-authored with Professor Luigi Mittone. We jointly devised the experimental design, whereas the rest of the work derives from my own contribution.

low cognitive costs and one with high cognitive costs, with the latter being characterized by lower performance levels. On average, participants deem that the majority choice is a winning option when the task to perform presents higher cognitive costs. Therefore, the imitative component is driven by the beliefs about others' performance. We also study the extent to which participants under or over-estimate their skills in completing the experiment, and we find that those who rely on a previous majority choice or on a default option are significantly underconfident, whereas the majority of participants is clearly overconfident, especially when the task to be solved is more cognitively demanding. We thus provide evidence for imitative behavior driven by underconfidence about own skills and by the beliefs regarding others' performance. We also find evidence for an additional heuristic related to the timing of decision making, which lends support to the saliency of more recent memories in cognitively demanding scenarios.

In the second essay we investigate a particular issue in the economics of information: information aversion.[3] More specifically, we are interested in exploring the determinants of information acquisition or aversion in a general context in which individuals first receive a preliminary signal about their own condition, after which they can decide to acquire more detailed information. Evidence from recent studies in Economics, Psychology and Medicine suggests that decision makers assign a value to information that goes beyond the instrumental role considered by traditional economic theories. In many circumstances individuals avoid conclusive information when they fear to receive bad news, whereas they search for information when they expect good news (Lerman et al., 1998; Karlsson et al., 2009).

This informational issue, also known as ostrich effect, can be relevant if we consider for instance health contexts, in which individuals may shy away from additional information when they fear bad news, even if additional information about their health status may be relevant in terms of prevention or treatment (Lyter et al., 1987; Facione, 1993; Caplan, 1994; Lerman et al., 1999; Nosarti et al., 2000; Richard et al., 2000; Meechan et al., 2002). If people derive utility from beliefs, they may manipulate sources of available information and obtain sub-optimal outcomes if their choices were different from the ones undertaken under a full information set. It follows that the amount of information acquired is not trivial and gives rise to an important trade-off between emotional costs from information and better decisions deriving from more information.

[3]This chapter is co-authored with Professor Marie Claire Villeval. We jointly devised the experimental design, whereas the rest of the work derives from my own contribution. Therefore the writing of the essay, the analysis of the existing literature, the data analysis and discussion follows from my own work.

To investigate this informational trade-off, we devise a novel experiment in which it is possible to exert an effort that increases the final expected payoff. The experimental design is characterized by two experimental sessions in a longitudinal framework, with the second session taking place two weeks after the first one. In the first session participants receive a fuzzy signal about their condition, which may become more informative if they decide to buy additional information in week 1. Between week 1 and 3 they are given the opportunity of completing an online task (effort), on three nonconsecutive days, which increases their expected payoff. The completion of the tasks has a higher impact on participants' expected payoff if their condition is bad rather than good. The main innovative component of the essay is the introduction of an interrelation between information acquisition and the possibility of exerting an effort that increases final expected payoffs. With this relationship we aim to capture and investigate the informational trade-off at the basis of the ostrich effect.

We find that participants significantly buy more information after a preliminary bad signal than after a good signal, revealing an opposite result with respect to the previous literature. Accuracy of signals fosters the exertion of effort only when the signal is bad. Participants understand the incentives of completing effort when they receive a bad signal about their condition only when they decided to be more informed. Providing participants with the possibility of acquiring additional information lowers their effort exertion. Moreover, the possibility of acquiring additional information points out a duality of individuals: those who decide to be more informed and thus follow a rational behavior in the exertion of effort, and those who decide to remain with a limited information set and behave irrationally in their effort exertion.

The third essay investigates whether the gender composition of evaluating commissions affects the presence of women in research activities.[4] The topic is currently highly debated and follows from a stylized fact: the great gender imbalance in research. In fact, women outnumber men by 18% when they finish university, but they represent 37% of Associate Professors and just 20% of Full Professors (Meulders and O'Dorchai, 2013). Several policy solutions have been proposed to reduce gender disparities, such as gender quotas for employees and, lately, gender quotas in evaluating commissions for research positions. The latter is motivated by possibly gender-biased preferences towards

[4]This chapter is co-authored with Professor Daniele Checchi and Dr. Nevena Kulic. I contributed to the essay in the collection of around one third of the data set, involving any bibliometric measure of candidates from the Scopus database. I wrote almost the entirety of the essay, except the parts in which the research Institute is described. I contributed extensively in the analysis of the existing literature and did the data analysis, both in terms of models and regressions and in terms of post-competition analysis.

same-sex candidates in evaluating commissions, which are generally male-dominated. Gender quotas in commissions have already been introduced in a few European countries and have recently been advocated by the European Commission.

However, imposing them does not represent a zero-cost policy since it requires a disproportionate share of senior academic women's time for attending evaluating commissions, a "non-productive" activity from a research perspective. This policy could be detrimental to female academics research activities and could thereby foster the vicious cycle of women being trapped in lower academic positions. This essay aims to examine the potential effectiveness of such a policy by using data from an Italian research community. We complement and extend the already existing investigation on the Italian research environment by studying the determinants of entrance to a non-academic research context.

We exploit a novel data set on recruitment processes in a leading Italian research center that mainly operates in hard science. Unlike previous studies that focus on selections into professorships (De Paola and Scoppa, 2011; Zinovyeva and Bagues, 2011; Bagues et al., 2014), this essay mainly examines entry level research positions, where gender imbalance usually starts taking place. We aim at identifying the main drivers for winning a research competition, and in particular we are interested in knowing whether a higher proportion of women in evaluating commissions increases the proportion of women selected for research jobs. As in almost all calls in the data set only one candidate is selected, the competition among candidates is more evident and explicit than in previous qualification data sets. This allows us to use different econometric techniques that take into account the interdependent success probability of candidates applying for the same call. The data set also enables us to control for pre-existing ties between candidates and commissioners or with the research center.

We find some evidence of discrimination against women at entry research levels, while the presence of women in commissions increases the chances for female researchers to be selected for these posts. Gender discrimination does not seem to be a concern for hierarchically higher positions. Higher quality commissions tend to choose more productive candidates and their decisions are not influenced by pre-existing ties with the institution. Prior ties are however the most important determinant of success in all other cases. The analysis of the post-competition productivity of candidates shows that applicants with prior ties are significantly more productive, suggesting a potentially positive role of prior ties in reducing the information gap about candidates' research quality.

2 Over-confidence and low-cost heuristics: an experimental investigation of choice behavior[1]

2.1 Introduction

In every day life decision makers frequently face cognitive costs when choosing among different options. Traditional economic theories of rational behavior disregard cognitive costs and assume that economic agents process costly information fairly easily, since they are always able to select the utility-maximizing option among different ones. In contrast, evidence gathered by some psychologists and economists supports the idea that decision makers systematically violate the assumptions of rational choice theory (Tversky and Kahneman, 1974; Camerer, 2003a; Bicchieri, 2006). In particular, the axiom of complete preferences, which entails that an individual is able to compare and express a preference relation between any two objects, requires conditions such as extraordinary computational and cognitive skills that are rarely met in practice. It follows that, in many circumstances, economic actors are boundedly rational and make use of simplifying heuristics, either conscious or unconscious, when they process information that carries cognitive costs (Gigerenzer and Gaissmaier, 2011).

In this paper we investigate a potential cognitive shortcut faced by decision makers: the influence of other agents in their reference community. A growing strand of literature argues that many decisions are affected by social interactions, also known as "peer effects"(Bernheim, 1994; Glaeser and Scheinkman, 2001). Peer effects represent the channel through which individuals belonging to a reference group influence the behavior of individuals in the same reference group, and they can be

[1] This chapter is a joint work with Professor Luigi Mittone, published as Cicognani and Mittone (2014), "Over-confidence and low-cost heuristics: an experimental investigation of choice behavior" on *Economics, the Open-Access, Open-Assessment E-Journal*. Financial support from the University of Trento and the School of Social Sciences of Trento is gratefully acknowledged.

conceived as "an average intra-group externality that affects identically all the members of a given group"(Calvo-Armengol et al., 2009, p. 1239).

The perspective from which we consider peer effects in this work is that according to which the behavior of other individuals in a reference community may represent a simplifying heuristic, that allows to economize on decision costs, especially when choices carry high cognitive costs. A potential channel through which this heuristic operates is the belief system, both about the quality of others' choices and about the quality of one's own choices. The latter refers to the definition of overconfidence, which in this essay we consider in both absolute and relative terms. The higher the overconfidence regarding own skills, the more we should expect a subject not to rely on other people's choices.

This paper builds on contributions such as Fortin et al. (2007) for the way in which information about others' behavior is put forward in the experiment: we provide subjects with information about the choice of a majority of participants in a previous experimental session. Our departure from existing research is twofold. First, we incorporate cognitive costs in an experimental setting of imitation choices. To the best of our knowledge, this is the first study that investigates experimentally social interactions from a cognitive perspective: we estimate the impact of the reliance on choices made by a majority of subjects in two scenarios that differ in terms of cognitive costs involved in the experimental task.

Second, we implement an original experimental task whose solution is not the result of computations or counting[2], but which entails a more complex kind of reasoning that is more similar to modern and dynamic decisional contexts. We show participants a set of abstract figures, and we model the difficulty of such figures into high or low. In these two different scenarios, we test the impact of information available to participants: whether they know or not the choices of a majority of subjects who took part in a previous experimental session.

We do not find strong evidence in favor of imitative behavior, but on average participants deem that the majority choice is a winning option when the task to perform presents higher cognitive costs. Therefore, the imitative component is driven by the beliefs about others' performance. We also study the extent to which participants under or over-estimate their skills in completing the experiment, and our results show that those who rely on a previous majority choice or on a default option are significantly underconfident, whereas the majority of participants is clearly overconfident, especially

[2]For instance, see Pokorny (2008).

when the task to be solved is more cognitively demanding. Our results in terms of under/over-confidence also allow us to shed more light on the gender differences related to this topic: we don't find gender differences in overconfidence if this is measured in absolute terms, but we do find that men are statistically more overconfident than women in relative terms.

In addition to investigating the connection between cognitive costs and imitative dynamics, our experimental setting also allows us to examine very precisely the temporal pattern of subjects' decisions. This enables us to provide evidence for an additional heuristic related to the timing of decision making, which lends support to the saliency of more recent memories in cognitively demanding scenarios.

The chapter is organized as follows: in Section 2.2 we present the related literature; Section 2.3 outlines the experimental design, the procedures followed during the experiment and the behavioral hypotheses; Section 2.4 presents the data and results from the experiment; Section 2.5 discusses and draws conclusions.

2.2 Literature review

Decision makers are characterized by limited information processing systems. Not only their computational and cognitive capacity is limited, but also their memory is confined in both capacity and duration. Accordingly, they frequently employ heuristics in order to reduce the information processing requirements of the tasks they face, which differ in terms of complexity (Payne, 1976). In any decision there is an underlying trade-off between the precision of the choice and the efficiency in the decisional process. When using heuristics, individuals give up to the former, in favor of the efficiency of the process. According to Payne (1976), we use heuristics in contexts of decision making under risk, when we face some probabilistic information processing. Examples of heuristics include representativeness, when individuals anticipate the most representative outcome of the existing evidence, as well as anchoring and adjustment, according to which we tend to base decisions on familiar positions, with an adjustment relative to this anchor. As the task to be solved increases in complexity, "individuals will employ decision strategies resulting in a restricted pattern of information search when the choice task becomes sufficiently complex. These decision strategies can be characterized as heuristic processes similar to those found in studies of human problem solving" (Payne, 1976, p. 324).

When facing a new task, which is neither mechanical nor computational, subjects might conform to the norms deriving from the majority of members of their community. According to Bicchieri (2006, p. 5), "to efficiently search our memory and group a new event with previously encountered ones, we use cognitive shortcuts. Cognitive shortcuts play a crucial role in categorization and the subsequent activation of scripts and schemata.(..) In the *heuristic* route, behavior is guided by *default rules* stored in memory that are cued by contextual stimuli. Norms are one class of default rules.". Referring to what has already been chosen by other subjects in the same decisional environment may represent a channel through which it is possible to save on cognitive costs. This type of social shortcut may be increasingly relevant when decisions are more demanding in cognitive terms. An underlying behavioral pattern could be driven by the beliefs regarding the behavior of others, and especially of the majority that one decides to conform to: this majority can be considered as acting better than the individual subjects, and hence could be chosen to be a reference point in decisional dynamics.

Economists' interest in the connection between social interactions and cognitive costs is very limited and relatively recent. An attempt in this direction is represented by the theoretical axiomatization of Hayakawa (2000), who supports the idea of social capital as a source of low-cost heuristics: choice decisions made by an individual depend crucially on the organization of her perceptions of available choices. To the extent that such perceptions are affected by social elements, they cannot be considered independent of the particular social environment where decisions take place, and, in such a context, bounded rationality is a tool for the formation of endogenous preferences.

Along these lines, Earl and Potts (2004), in investigating the social side of choice in complicated situations make a sharp distinction between high-level preferences and low-level ones. While the former relate to innate behavior, the latter are the result of a learning and specialization process, and they can be acquired in the market for preferences, by simply using and buying other people's learning. Economic agents are indeed cognitively normal individuals who may access and acquire specialized consumer knowledge, regarded as recommendation of someone having relevant expertise on some specific problems. This is what happens in highly uncertain speculative markets, as highlighted by Earl et al. (2007), in which economic agents make use of decision rules picked up from others in an attempt of interpreting and gathering new relevant information. Heuristics are hence socially transmitted.

Conversely, very large and heterogeneous is the extent of contributions that investigate the ways and motives for which individuals are influenced by others in their decisions making processes, irrespective of the cognitive costs they face. These works focus on issues ranging from peer effects, to informational cascades, conformism and social learning.

Peer effects investigations aim at measuring and disentangling endogenous effects and at solving the identification problem: individuals in a group influence each other and it is not easy to state who influences whom.[3] While this represents quite a complex problem to overcome empirically, through field data, the experimental research allows to avoid it, due to a complete control of the information provided to experimental subjects and to the absence of self-selection of subjects into groups. More specifically, in a laboratory setting the experimenter is able to control for different information scenarios, and hence to regulate the direction of influence of knowledge.

A significant amount of experimental work has modeled peer effects by conveying information to participants about the behavior of experimental subjects in previous sessions. Fortin et al. (2007) investigate a tax evasion scenario and divide participants into groups: at the beginning of each period subjects receive feedback about the amount evaded and the number of evaders among the members of their group in the previous period. Beugnot et al. (2013) devise both a recursive and a simultaneous source of influence from others' behavior. In the former, each participant has one or two peers (belonging to the previous baseline treatment) and receives information at the beginning of the session about peers' personal information, average piece-rate and performance in the task[4]; in the latter, interactions occur in real time among players, in order to recreate real-like working situations. Overall, individuals are found to be significantly influenced by others' performance.

A more sophisticated network structure, which also entails network formation during the experiment, is proposed by Conte et al. (2009). The authors investigate which are the factors underlying network formation in an experiment in which, over a minimum of 15 rounds, connections between two players can be set up during the game if both players mutually agree. Network formation involves positive externalities to the extent that both direct and indirect links accrue benefits, but direct links involve costs. Given these payoff rules, players aim at maximizing their own earnings by forming indirect and direct links. They find empirical ground for best-response behavior, besides reciprocator and opportunistic patterns of playing the network formation game.

Despite these network studies, existing experimental literature does not seem to examine and disentangle the underlying factors that drive individuals to conform to others' choices, which is the focus of this work.

[3] This issue represents the "reflection problem" introduced by Manski (1993).
[4] Participants must perform a multiplication task without calculators and pens

A potential motivation for being influenced by others' behavior may act through the beliefs regarding others' outcome: we may imitate those individuals that we consider as successful. This is what Apesteguia et al. (2007) and Offerman and Sonnemans (1998) indicate among their results and it is also investigated in our experiment through an incentivized elicitation of beliefs regarding others' performance. Apesteguia et al. (2007) in a symmetric Cournot game distinguish between participants and roles played by participants: participants can play in three roles and are not told with whom they are matched within groups of nine subjects each. The authors find evidence in favor of imitation of more successful individuals: the degree of imitation they find is increasing in the difference between own and others' payoff. Moreover, higher imitating patterns arise with respect to the individuals with whom one interacts, rather than with respect to individuals who play in the same role but in different groups. A similar result is found by Offerman and Sonnemans (1998), with an experiment about investment decisions that depend on different states of the world. Rather than imitation of choices, they test for the presence of imitation of incentivized beliefs: least-performing individuals are found to imitate more the judgements and beliefs of more successful subjects.

Another behavioral motivation underlying imitative dynamics can rely on the degree of under/overconfidence of subjects. To the extent that one considers herself to perform better than a majority, referring to what others have already chosen may represent an unattractive strategy, although it would allow to economize on decision costs. The better-than-average effect is a widespread cognitive bias in the population. According to Svenson (1981), 90% of the U.S. population considers herself to be above-average drivers. Overconfidence about one's relative skills can have many repercussions in real life situations. For instance, it has been put forward as an explanation for the high rate of business failures (Camerer and Lovallo, 1999), or for disproportionate job market search and unemployment in labor economics types of contexts (Dubra, 2004). Also in financial markets overconfidence in relative ability has been found and connected with excessive trading, which entailed a consequent reduction in net returns (Barber and Odean, 2001). Contrary to these findings, in an experiment aiming at testing overconfidence in individuals' forecasts of their absolute or relative performance in two unfamiliar tasks, Clark and Friesen (2009) report a general tendency to underconfidence, although "underconfidence is greatest in forecasts of absolute rather than relative performance"(Clark and Friesen, 2009, p. 229).

Interesting evidence in the overconfidence literature is represented by differences across genders: it seems to be quite established by now that men are more overconfident than women (Frechette and Schotter, 2015). For instance, in an empirical study on a common stock investments data set, Barber and Odean (2001, p. 261) find that "men trade 45 percent more than women". This gendered overconfidence result holds true also when we look at relative, rather absolute overconfidence. As Niederle and Vesterlund (2007, p. 1069) document, "men are substantially more overconfident about their relative performance than women". Therefore, it seems that women have on average lower expectations regarding their relative ability. In our study we are able to track both participants' absolute under/over-confidence and their relative one, through ad-hoc questions posed in the ex-post questionnaire at the end of the experimental sessions.

This essay is also related to the theoretical and applied literature on crowd-following or herd behavior, which, starting from the seminal contribution of Banerjee (1992), considers economic agents who have incomplete information about the quality of a good or service and hence let the choices or evaluations made by others to guide their own actions.[5]

Imitation can also be conceived as the desire to conform. Yet, in our experimental setting participants are not provided with information about who chose the majority card. Therefore, it is difficult to conceive that participants aim to follow the majority for mere conformist reasons, since they do not know anything about which kind of majority it is.

Jones (1984) examines different explanations for conformism: these include economies of scale, heuristics as solutions to complex coordination problems, and conforming as a way to choose one of the many existing equilibria.

Imitation is also a source of social learning. For instance, Bala and Goyal (1998) develop a theoretical model that explains how social learning from neighbors guides individuals' decisions when payoffs from different actions are unknown.
From a different and experimental perspective, based on a repeated ultimatum game and a repeated best-shot game, Duffy and Feltovich (1999) test whether observing other players' actions and payoffs leads to an evolution of the way the game is played. Evidence is mixed and strictly hinges on the type of game played.

A different line of argument characterizing individuals' decisions and possible shortcuts refers to the different timing of experiences faced by subjects. Indeed, cognitive processing could be related to memory capacity, so that more recent events become more salient when subjects experience cognitively demanding tasks (see, for instance, Hastie and Dawes (2010)). When facing alternatives,

[5]See, for an example, Çelen and Kariv (2004) and Sorensen (2007).

decision makers experience utilities, and they attach such utilities to the alternatives present in their memory. Hence, preferences arise from memory representations (Weber and Johnson, 2006). In this context, a central role is played by the timing of the experience, which leads to different memories of experiences (Miron-Shatz et al., 2009): preferences depend on the memory of past experiences. In our experiment, when confronted with cognitive costs, we expect that individuals attach more relevance to memories of experiences occurred later in time.

2.3 Design, procedures and behavioral predictions

2.3.1 Design

The novel feature of our experimental design is that participants do not face a typical experimental game, nor a mechanical counting or computational task.[6] Rather, on their computer screens, they face covered cards which contain on their inner side abstract figures composed of black squares.[7] Participants' aim is to select the card containing the figure which most closely resembles the figure appearing on the top left of the screen. In doing so, they can either uncover and inspect the cards that appear covered on the screen, or they can choose a covered card representing a default option, which is placed on the bottom right of the screen. The default card is accompanied by a different amount of information according to treatments. Indeed, in the DEFAULT treatments, participants are informed that the default card is one of the best 8 cards (among 16) appearing covered on the screen, while in the MAJORITY treatments they are told that the default card (in this case called "majority card") has been chosen in a previous experimental session by the majority of participants, and that, moreover, it represents one of the 8 best cards appearing covered on the screen.

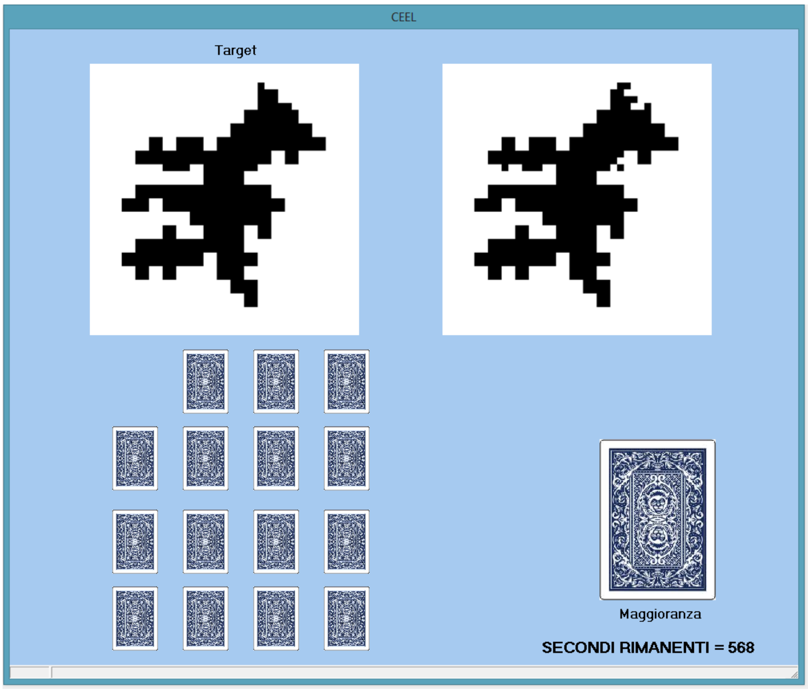
The task is implemented both under a high-cost and a low-cost scenarios, both considered in cognitive terms: in the high-cost condition, the visual resolution of the figures (number of "pixels") is higher than in the low-cost condition. The rationale of the task is to offer subjects an environment in which, while searching the solution, they do not have a reference point, either computational or logical: the assignment is not related to an univocally clear solution, characterized by normative determinants. Our approach aims indeed at resembling a dynamic social environment, where available options must be discovered, but it is not straightforward which is the best one among them.

The implementation of the design entails participants to face a screen with one card on the top

[6]For an overview of real-effort tasks, see Gill and Prowse (2013).
[7]We are deeply indebted to Paolo Crosetto for the programming of figures through Python.

left (also known as "target card"), and 16 covered cards below, divided into rows of 4 cards each.[8] Cards are randomly distributed on the screen for each subject and for each treatment, so that the distribution of the 16 cards is different for each participant. Each card carries an abstract figure on its covered side. For the initial 10 seconds all cards are uncovered, and hence the figures portrayed on them are visible. As previously mentioned, the aim of the game is to choose the card which is the most similar to the target card, within a time constraint of 10 minutes. After the 10 initial seconds, all cards get covered: comparisons between any of the 16 cards and the target card can be carried out by participants by clicking on each card which is chosen for the comparison. After clicking on one of the 16 cards, the selected card is shifted next to the target card and both become uncovered. Participants can keep the selected card beside the target card, for a visual comparison, for as much time as they wish, within the 10 minutes of time constraint of the whole experiment. In order to place the card back to its initial position, it is necessary to click again on the card (see Figure 2.1).

Figure 2.1: Example of a screenshot during the comparison



Notes: The Figure corresponds to the high-majority treatment.

Moreover, after the 10 initial seconds, a default card appears, covered, on the bottom right of the screen, according to the type of treatment.[9] If one of the 16 cards is inspected and then placed back in

[8]See Appendix 2.A.1 for the instructions provided to participants.

[9]In the MAJORITY treatments, the default card is called majority card, since it represents the choice of a majority of participants in a previous experimental session.

its original position among the other 15 cards, its boundaries get red. Participants are informed that within the time limit of 10 minutes they have to select either one of the 16 cards or the default/majority card. The 16 cards provide 16 different payoffs, ranging from 0.25 to 4 Euros, depending on the fitness of the figure represented on the chosen card with respect to the figure in the target card. For "fitness" we indicate the similarity between the chosen and the target card in terms of black squares constituting the figures portrayed on the cards.

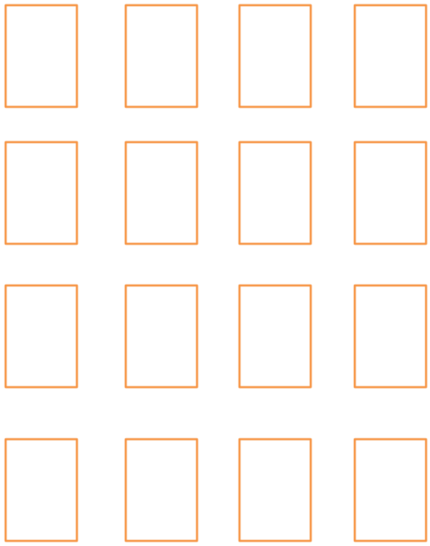
Participants also know that the default/majority card represents one of the 8 best cards in terms of fitness among the 16 cards (hence leading to an earning from 2 to 4 Euros). If the default/majority card is selected, it represents the final choice: in this case participants can no longer explore the other cards appearing on the screen and they will receive the payoff corresponding to the chosen card. Since participants cannot inspect the default/majority card, they can rely only on the information about this card which is provided in the experimental instructions. The rationale for this is to test subjects' beliefs about the goodness of the default/majority card, without subjects being influenced by what they see on this specific card.

In case no card is selected within 10 minutes, participants simply receive the show-up fee of 2.5 Euros. The countdown of seconds available is displayed at the bottom right of the screen. At the beginning of each session, participants are endowed with a pen and a printed page consisting of the contours of the 16 cards divided into rows of four, as they appear on each computer screen. The idea is to facilitate subjects in remembering the fitness of the cards as they work through them, by allowing participants to take notes related to the cards inspected. An example of the page is provided in Figure 2.2.

After the completion of the choice, an incentivized belief elicitation task regarding the "fitness" of the default/majority card takes place.
What we mean by "fitness" is the number of black squares composing each figure that differ in terms of position between the selected and the target card. This measure changes across the high-cost and the low-cost treatments. In the low-cost scenario, the target figure is made up of 100 black squares. Each of the 16 cards differs from the target not for the total number of squares composing the figure, but for 5 to 20 squares which are placed in a different position with respect to the target figure.

The idea for the high-cost scenario is that of increasing the number of "pixels" composing the figures, that is the resolution: the target figure in this case consists of 400 black squares and each of the 16 cards differs from the target card for 20 to 80 squares which are placed in a different position

Figure 2.2: Example of the annotation page



with respect to the target figure, while the total amount of squares remains the same. In practice, each black square of a low-cost figure corresponds to 4 black squares in a figure in the high-cost scenario. Therefore, in both scenarios, there isn't the perfect match card among the 16 cards on the screen: the best cards in terms of "fitness" are those with 5 and 20 squares in a different position with respect to the target figure, in the low-cost and high-cost scenarios respectively. Panel (a) and (b) in Figure 2.3 provide examples of figures in the low-cost and high-cost treatments, respectively. We expect that the increase in the number of squares composing the figures leads to an increased complexity in the task, and hence to higher cognitive costs.

It follows that in the belief elicitation task participants are asked to choose the number of squares they believe that differ between the target and the default/majority card.[10] All options, according to the treatment, are presented and participants are asked to tick on one of them in order to end the experiment: in the low-cost treatments, numbers from 5 to 16 were displayed, whereas in the high-cost treatments participants could tick on numbers ranging from 20 to 80, with a distance of 4 to each other (20, 24, 28, 32, and so on, until 80). Hence, subjects were not requested to write the amounts on their own, without any prior information of the possible numbers, but were asked to select one among the numbers displayed.

If the number selected is correct, one additional Euro is earned at the end of the experiment by the participant who made a right guess about the fitness of the default/majority card.

[10]Once the experimental task has been completed, we ask subjects the following question "By what amount of squares do you think that the default (or majority) card differs from the target?"

Figure 2.3: Example of figures

(a) Low-cost treatments

(b) High-cost treatments

Notes: In red it is represented the dimension of a square.

In addition, at the end of each session, participants were asked to fill out an anonymous ex-post questionnaire. This questionnaire aimed at gathering information about the perceived difficulty of the task, impressions about what the figure represented, beliefs about their absolute and relative performance as well as a series of demographics data such as age, gender and college major (see Appendix 2.A.2 for details).

The experimental design develops across two dimensions of treatments, that allow us to disentangle the role played by the availability of social influence in a high-demanding scenario in cognitive terms versus a low-demanding one. Moreover, the experimental setting also permits us to control for the effect of imitation and the temporal saliency of choices, in conditions of low and high cognitive costs.

A first dimension of the 2 x 2 experimental design is represented by cognitive costs: in the low-cost treatment, the figures are made up of 100 black squares, while in the high-cost treatment figures have a higher precision, since they include 400 squares. The second dimension of treatments involves the information given to participants about the default/majority card. In the default treatment, the card on the bottom right is called DEFAULT CARD and participants are informed that it is one of the 8 best cards among the 16, in terms of fitness with the target card. Conversely, in the majority treatment, the card on the bottom right is denominated MAJORITY CARD: participants are notified that it is one of the 8 best cards, and, in addition, that it was chosen by the majority of participants from a previous experimental session. The rationale of providing information also in the default treatment is to preserve comparability between the default and the majority treatments. In this way, the communication of the fitness of the cards, in other words their expected value, remains invariant across treatments. What changes is the additional information of the choice implemented by a previous majority, that in our approach aims at representing a social interaction component which participants may decide to follow when facing a cognitive cost.

Table 2.1 provides an outline of the four treatments.

Table 2.1: Treatments

		Cognitive costs	
		Low-Cost	High-Cost
Information on the majority	No	*low-default*	*high-default*
	Yes	*low-majority*	*high-majority*

2.3.2 Procedures

A baseline session without the availability of a default/majority card was previously run with 20 participants in order to gather data on the majority card used in the following majority treatments. The low-cost treatment was administered to 10 subjects, and the high-cost treatment was conducted on the remaining 10 subjects. The best card in terms of fitness with the target card was chosen by the majority of participants in both treatments.[11]

Following this baseline initial session, 154 subjects took part in the experiment, divided into 12 sessions.[12] 40 participants took part in each of the four treatments, except in the low-majority treatment in which the number of participants was 34. The experiment was fully computerized and it was performed through a between-subject design. Participants were volunteer undergraduate and graduate students from different humanities and technical faculties from the University of Trento. Each session lasted around 30 minutes and it was conducted at CEEL (Cognitive and Experimental Economics Laboratory, Trento) between April and June 2013. A show-up fee of 2.5 Euros was given. The average earning per participant was of 5.7 Euros, and it was paid privately and anonymously in a separate room at the end of the experiment. Subjects could not participate in more than one session.

[11] Appendix 2.A.3 reports descriptive statistics of the baseline session, used to retrieve data about the majority choices in the low-cost and in the high-cost setting.

[12] Among a total of 166 participants, 154 selected a final choice, while 12 run out of time. For the analysis, we decided to consider only data about the 154 subjects who finalized a choice.

Upon their arrival at the laboratory, participants were randomly allocated to cubicles inhibiting interaction with other participants. Each participant could read the instructions for the experiment on the computer screen placed in the cubicle, and she was invited to read them privately. Later, a member of staff read the instructions aloud and participants were given the opportunity to privately ask staff members for clarifications. A set of computerized control questions was administered to subjects in all treatments, in order to verify their understanding of the task.

The experiment consisted of four treatments, as depicted in Table 2.1. Three sessions per treatment were run.

2.3.3 Behavioral predictions

A first conjecture arises related to the task administered to subjects. We believe that our approach in designing the figures that participants have to visually compare entails a higher cognitive cost in case of a higher amount of resolution of the figure in the card, which consequently involves a lower performance rate. Therefore, we maintain that increasing the cognitive cost implies lower performance choice rates.

In this study we are interested in analyzing what kind of behavioral shortcuts do individuals put into practice when they face decisions which are cognitively demanding and do not present a clear cut solution. We aim at testing two possible heuristics: a social one, related to the influence of a previous majority of participants, and a temporal one, which may guide the timing in which decisions are taken.

Following the arguments introduced in Section 2.2, information on others' choices can be conceived as a way to economize on decision costs. Therefore, this leads us to formulate the following testable prediction:

Hypothesis 1 People imitate more in contexts highly demanding in cognitive terms

If Hypothesis 1 is verified, we should observe more participants opting for the default choice in the high-majority treatment. At the basis of the hypothesis, our experimental setting allows us to shed light on participants' motivations underlying their choices. This is possible by eliciting their beliefs regarding the quality of the default/majority card, and also by eliciting their beliefs related to their own relative and absolute performance. On the one hand, the better they think the default/majority card is, with respect to the target card, the more frequently they should choose it. On the other hand, the perceptions on their own skills in solving the task can also play a role: we expect that more overconfident subjects rely less on what has been previously chosen by other participants.

The second heuristic we aim at examining refers to the timing of choices, and to whether this differs in case participants face a more complex kind of task. More specifically, we expect that:

Hypothesis 2 More recent memories have a greater impact on choices

Our experimental design enables us to follow with much detail participants choices in terms of cards uncovered and time spent in solving the task. If Hypothesis 2 is confirmed, we should observe a higher frequency of participants choosing their final card among the more recently uncovered cards. This behavioral pattern could be even more important in the case of high-cost treatments.

2.4 Results

In this section, we first present the descriptive statistics of the sample of participants. We then discuss econometric results related to the research hypotheses we aim to test. A list of the variables used in the analyses is included as Table 2.11 in Appendix 2.A.4.

2.4.1 Descriptive statistics and prima facie evidence

From the pages for notes collected after each experimental session, it is clear that for most participants they were useful tools for recording their thoughts as they inspected the cards, and hence remembering the goodness of the cards visualized. Overall, 27% of such pages were left blank, 12% presented draws or scrawls, while the remaining 61% exhibited some measures of quality about the cards, as for instance arrows, percentages, pluses and minuses or different numbers of stars written within the contours of some cards. Across treatments we do not observe statistical differences in the types of writings left on the pages for notes, or in the amount of cards left blank.

Table 2.2 reports means, standard deviations, minimum and maximum values of the main variables deriving from the experimental sessions.

Table 2.2: Descriptive statistics

	Mean	Std. Dev.	Min	Max
Payoff	5.71	1.00	2.75	7.5
Net payoff	5.42	0.95	2.75	6.5
Seconds	417231.6	162956.2	1	581673
N. uncovered cards	21.80	8.74	1	49

The variable "payoff" included in Table 2.2 represents the total payoff from the experiment for each subject, while in the variable "net payoff" it is subtracted the additional Euro gained in case of a correct guess of the belief about the default/majority card (out of 154 participants, 45 guessed the fitness of the default/majority card correctly). The average final payment consists of almost 6 Euros. On average, subjects complete the task in almost 7 minutes and they uncover 22 cards.

In Table 2.3 are included the descriptive statistics of the answers provided by participants in the belief elicitation task and in ex-post questionnaire administered at the end of the experiment, differentiated between men and women.

The variable "belief default" of Table 2.3 is the ratio between the real difference in black squares between the target and the default/majority card, and the believed difference in black squares between the target and the default/majority card declared by each subject at the end of the experimental sessions. As indicated in Section 2.3.1, at the end of the experiment participants face a belief elicitation task: they are asked to answer to the following question "By what amount of squares do you think that the default (or majority) card differs from the target?". This elicitation is incentivized, as subjects earn one additional Euro if they answer correctly. Since in both low-cost and high-cost treatments the default/majority card is the one with the smallest amount of different squares with respect to the target card (5 over 20 in the low-cost treatment vs 20 over 80 in the high-cost treatment), it follows that the "belief default" values range between 0.25 and 1, in case of the worst and best belief about the default card, respectively. Therefore, a higher value for "belief default" represents higher expectations about the goodness of fit of the default/majority card. On average women have worse beliefs regarding the default/majority card than men, but the difference is not statistically significant (Wilcoxon rank-sum test, $z = -1.564$, $p = 0.1177$).

At question 1) of the ex-post questionnaire, participants were asked to indicate how accurate they think their final choice is. The answer is in terms of the number of squares of difference between their chosen card and the target one. Answers can take values between 5 and 20 in the low-cost treatments, and values between 20 and 80, with intervals of four, in the case of high-cost treatments. The answer to question 1), which we call "believed accuracy", is normalized on a scale between 1 and 16, where 1 indicates that the participant believes to have chosen the best card among the 16, and 16 indicates the worst expectation about one's own performance in the task. The normalization on a scale 1-16 is implemented in order to compare the answers in the low-cost and high-cost treatments, since the number of squares differ in these two conditions.

Table 2.3: Descriptive statistics - Belief elicitation and ex-post questionnaire

	Females		Males	
	Frequency	Percentage	Frequency	Percentage
Belief default				
Mean	0.69		0.75	
Believed accuracy				
Mean	2.85		2.63	
Absolute under/over-confidence				
Mean	-2.81		-2.31	
Difficulty				
Mean	5.23		5.54	
Relative under/over-confidence				
Better	31	37.80	40	55.56
Equal	42	51.22	29	40.28
Worse	9	10.98	3	4.17
Choice majority				
Yes	49	59.76	52	72.22
No	33	40.24	20	27.78
	Frequency		Percentage	
Faculty				
Economics & Management	73		47.40	
Engineering	10		6.49	
Law	23		14.94	
Literature	14		9.09	
Math	11		7.14	
Not a student	5		3.25	
Sociology	18		11.69	
Degree level				
Out of due date student	17		11.04	
Master student	33		21.43	
Not a student	7		4.55	
1st year student	21		13.64	
2nd year student	39		25.32	
3rd year student	37		24.03	
Gender				
Female	82		53.25	
Male	72		46.75	
Age				
Mean	23.37		23.49	

At the end of the experiment participants finalize their choice. We can measure the accuracy of their final choice and we call it "actual accuracy". The cards rank from the best one -number 1- to the worst one -number 16- in terms of similarity with respect to the target card. Therefore, also the actual accuracy measure of participants' choices ranges between 1 and 16, with 1 indicating that the participant was able to select the best card, and 16 indicating that the participant selected the worst card.

Believed and actual accuracy can be matched together, allowing us to construct a variable representing absolute under/over-confidence. More specifically, the variable is obtained by subtracting actual accuracy from believed accuracy, and ranges between - 15 and + 15, with -15 indicating the maximum level of absolute overconfidence and +15 representing the maximum level of absolute underconfidence a subject can report. As can be noted from Table 2.3, both women and men are on average overconfident in absolute terms, with females being slightly more overconfident than men, although the difference is not statistically significant (Wilcoxon rank-sum test, z = -0.345, p = 0.7305).

At question 2) of the ex-post questionnaire, participants were asked to indicate the level of difficulty of the task, on a scale between 1 (very easy) to 10 (very difficult). In Table 2.3 the variable "difficulty" depicts the average of the answers to this question. Women perceive the task as slightly easier than men, but the difference between the averages is not statistically significant (Wilcoxon rank-sum test, z = -0.899, p = 0.3688).

We also asked participants to compare their performance in the task with that of the majority of participants. More specifically, in question 3) of the ex-post questionnaire we ask whether they think to have performed better, equally or worse than the majority of the participants. Answers to this question provide us with a measure of relative under/over-confidence. In the male answers we have a clear better-than-average effect, since more than 50% of male participants believe to have performed better than the others. Women appear much more overconfident than men in relative terms. If we join the answers "better" and "equal", and we compare them to the "worse" answers, we find that women are significantly less overconfident than men (Wilcoxon rank-sum test, z = -2.2049, p = 0.0275).

Taking into account some of these variables across the four treatments, we can see a more detailed picture of the underlying patterns of the experiment in Table 2.4.

We find a statistically significant lower payoff in the high-cost treatments compared to the low-cost treatments, both in case of the payoff and the net payoff variables (Wilcoxon rank-sum test, z =

Table 2.4: Means across treatments

	Low-default	High-default	Low-majority	High-majority
Payoff	6.12	5.09	6.14	5.56
Net payoff	5.87	4.84	5.93	5.11
Seconds	453786.10	411323.20	405916.30	396203.50
N. uncovered cards	22.75	22.53	20.26	21.43
Difficulty declared	5.33	5.83	5.59	4.80
Belief default	.70	.71	.65	.79

Table 2.5: Means - Default versus majority treatments

	Default treatments	Majority treatments
Payoff	5.60	5.83
Net payoff	5.35	5.49
Seconds	432554.7	400666.2
N. uncovered cards	22.64	20.89
Difficulty declared	5.58	5.16
Belief default	.71	.73

5.492, p = 0.00; z = 6.859, p = 0.00, respectively[13]). This is corroborated by looking at Figure 2.4, in which we can notice that the median payoff values for the high-cost treatments are always below the median payoff values for the low-cost treatments.

This result supports our experimental approach and sheds light in favor of the presence of a higher cognitive difficulty in case of figures with a higher resolution, since the net payoff is a clear measure of performance. This preliminary evidence is verified in the ordinary least-squares regression analysis which investigates the determinants of the performance reached by the 154 participants in the experiment (see Table 2.6).

Table 2.6: Regression model of the determinants of performance

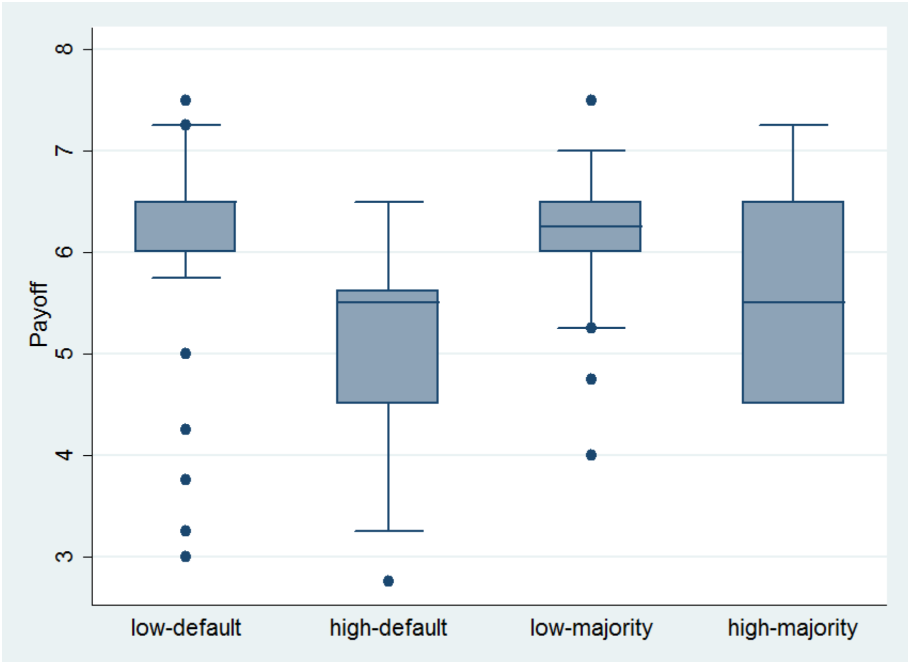
Payoff ∼	Coeff.	Std. Errors
(Intercept)	5.776	(.290)***
info	.290	(.147)°
cost	-.788	(.146)***
male	.250	(.147)°
difficulty	-.051	(.033)
seconds	8.73e-07	(4.55e-07)°
Number of Observations: 154		
Significance levels: *** 0.001; ** 0.01; * 0.05; ° 0.1		

Notes: The dependent variable in the model is the payoff in the experiment.

The dependent variable in the model - the payoff in the experiment- is regressed on the following

[13]Throughout the paper, treatment means are used for statistical two tailed tests, unless otherwise specified.

Figure 2.4: Payoff per each treatment



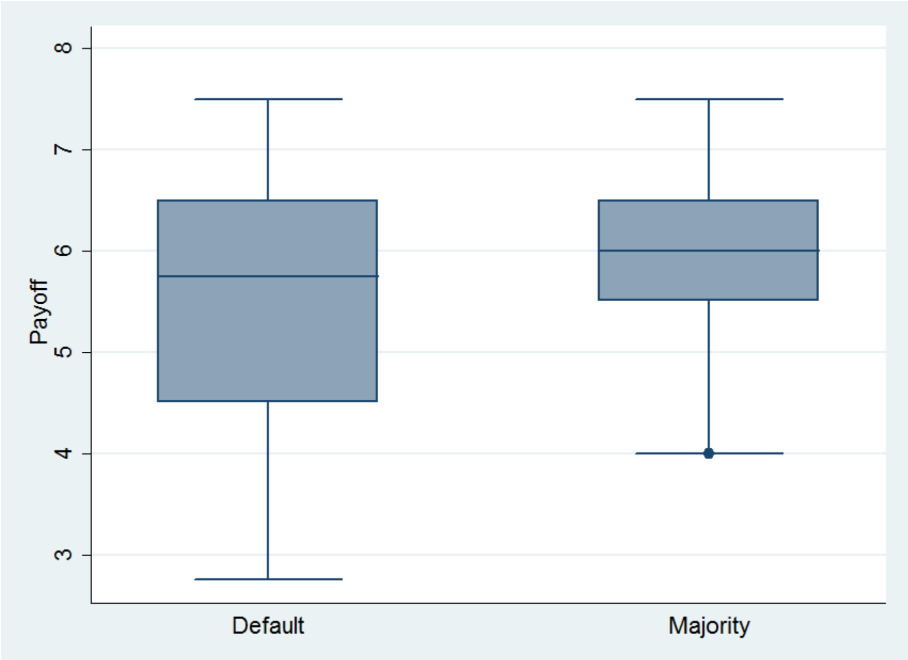
Notes: Middle bars correspond to median values, the edges of the boxes comprise the inter-quartile range between the 25^{th} and the 75^{th} percentiles, whiskers correspond to 1.5 times this range and circles characterize any other observation.

explanatory variables: *info* is a dummy variable equal to 1 if the treatment provides information about the majority (low-majority and high-majority treatments), and it is equal to 0 otherwise (low-default and high-default treatments); *cost* is equal to 1 if the treatment is highly costly in cognitive terms, and it is equal to 0 otherwise; *male* is equal to 1 if the participant is a male, and to 0 in case of females; *difficulty* is the perceived difficulty declared by subjects in the ex-post questionnaire and *seconds* is the amount of milliseconds employed for the completion of the task.

The regression output reported in Table 2.6 confirms what previously discussed, as can be noticed from the negative and highly significant coefficient of *cost*. The positive and statistically significant coefficient of the dummy information might be due to the higher frequency of default choices in the majority treatments: the default choice was attached to the highest payoff, even if participants were not aware of this. Males, *ceteris paribus*, seem to perform slightly better than females. Furthermore, the perceived difficulty declared at the end of the experiment does not seem to predict the performance of participants, while the more time spent completing the task is significantly associated to a higher payoff. This evidence leads to the following result, which confirms the validity of our experimental task:

Result 2.1. *Figures with a higher resolution are associated to higher cognitive costs and hence lower*

Figure 2.5: Payoff - Default versus majority treatments



Notes: Middle bars correspond to median values, the edges of the boxes comprise the inter-quartile range between the 25th and the 75th percentiles, whiskers correspond to 1.5 times this range and circles characterize any other observation.

performance rates.

Following the descriptive statistics of Table 2.4, in the majority treatments participants employ the smallest amount of seconds, especially in the high-majority one. Moreover, participants finalize their choice faster if the task involves high cognitive costs, seeming hence to give up before if they face a difficult task.

Conversely, the maximum number of uncovered cards is registered in the low-default treatment. This might be due to the low difficulty of the task, so that participants are more willing to inspect the 16 cards on the screen. Nevertheless, and quite surprisingly, the lowest difficulty declared for the task is related to the high-majority treatment.

When we consider the beliefs about the goodness of fit of the default/majority card, they are better in the case of the high-cost treatments with respect to the low-cost ones, while the information setting does not seem to influence subjects' beliefs (Wilcoxon rank-sum test, z = -2.261, p = 0.0238; z = -0.618, p = 0.5369, respectively). This enables us to derive the following result:

Result 2.2. *High-cost treatments present better beliefs about the default/majority card.*

Table 2.5 reports the descriptive statistics of default and majority treatments, pooled together.

From a first visual comparison no sharp differences emerge when comparing the column of the default treatments and the column of the majority treatments. This is confirmed after performing a series of Wilcoxon rank-sum tests on the means of the variables of Table 2.5: the tests do not show statistical differences between the default and majority treatments.[14] Absence of significant differences between the default and majority treatments can also be noted in Figure 2.5, which shows that the only discrepancy in the payoff between default and majority treatments is in the variance of payoffs, which is higher in the default treatments.

2.4.2 Imitative dynamics and over-confidence

Table 2.7 reports the frequencies of subjects who chose either the default or the majority card.

Table 2.7: Frequencies of default options

	Low-default	High-default	Low-majority	High-majority
Frequencies default	1	2	2	3
N. of subjects	40	40	34	40

As we can note, the frequencies of subjects choosing the default/majority option are very low: out of 154 individuals, only 8 adopted this decisional shortcut. The highest frequency of default choices is in line with our imitation hypothesis: people decide for a shortcut in case the task is more costly in cognitive terms and a majority outside option is available, but differences across treatments are not statistically significant (two-sample test of proportions with respect to cost and information treatments, respectively: $z = -0.6135$, $p = 0.5395$; $z = -0.8400$ $p = 0.4009$).

We do not find evidence for imitative behavior in choices, but we consider to investigate more deeply the beliefs about the default/majority cards in order to have a better understanding of imitative dynamics.

A closer examination of the elicited beliefs across treatments requires us to undertake pairwise comparisons. We first compare average beliefs about the goodness of the default/majority cards between low-default and high-default treatments. By doing so we aim to test whether, while keeping the same default informational setting and changing the difficulty of the task, participants have different expectations regarding the goodness of the default/majority card. We do not find any statistically significant difference (Wilcoxon rank-sum test, $z = -0.483$, $p = 0.6290$).

[14]The results of the Wilcoxon rank-sum tests, considering the order of the variables of Table 2.5, are: $z = -1.157$, $p = 0.2474$; $z = -0.344$, $p = 0.7307$; $z = 1.349$, $p = 0.1774$; $z = 0.964$, $p = 0.3349$; $z = 0.994$, $p = 0.3201$; $z = -0.618$, $p = 0.5369$.

The same result holds when we compare average beliefs between the low-default and the low-majority treatments (Wilcoxon rank-sum test, z = 1.030, p = 0.3029). This means that, while keeping the same (low) difficulty of the task, participants do not report differences in the expectations regarding the default/majority card.

A different pattern arises concerning beliefs in the high-majority treatment. Participants in this condition have higher beliefs related to the default card both with respect to the low-majority treatment and the high-default treatment (Wilcoxon rank-sum test, z = -2.636, p = 0.0084; z = -1.748, p = 0.0804, respectively). This suggests that a majority, considered as social component in our experiment, is considered to implement better decisions in presence of high cognitive costs. Hence, by investigating participants' beliefs about the quality of the default/majority card, we can argue that the decisions of the majority represent a focal point in the presence of higher cognitive costs. We can therefore conclude that:

Result 2.3. *Beliefs about the choice of a majority of individuals are better with respect to a setting with lower cognitive costs and with respect to a setting in which there is no information about the choice of a majority.*

Imitative dynamics are also investigated in connection with participants' under/over-confidence regarding the experimental task. It might indeed be the case that subjects decide to conform to what a majority has already decided because they feel underconfident about their skills in solving the task. We first compare the perceived difficulty of the task between those participants who opted for the default/majority card and those who chose as final card one inspected by themselves. For the former group the average reported difficulty is 6.75, whereas for the latter group it is on average 5.3. This represents a first clue about referring to the default/majority choice when the task to be completed is perceived as complicated, although the Wilcoxon rank-sum test for the comparison of average difficulty between the two samples is not statistically significant (z = -1.438, p = 0.1506).

When we consider the mean value of the absolute under/overconfidence variable, we find that it is -.364 in the low-cost treatments and -4.625 in the high-cost treatments (the differences in this case are statistically significant: Wilcoxon rank-sum test, z = 6.477, p = 0.0000). There are no differences in terms of under/over-confidence when information treatments are compared (the mean value of the under/over-confidence variable is equal to -2.81 and -2.32 in the default treatments and the majority treatments, respectively).

Comparing the subjects who chose the default/majority card and the subjects who did not, we find a statistically significant difference in the absolute under/over-confidence variable: more specifically, the former subjects under-evaluate their skills by 1, while the latter over-evaluate their skills on average by 2.77 (Wilcoxon rank-sum test, z = -2.828, p = 0.0047). That is to say that participants are more overconfident about their capacities in absolute terms when the task is more difficult, and, as expected, they are more overconfident when they don't rely on the default/majority card as their final choice. This result is in line with the values declared about the difficulty of the task. The above analysis leads to the following result:

Result 2.4. *In the high-cost treatments participants are more overconfident about their skills in absolute terms. The same applies to participants who don't rely on a default/majority card as their final choice.*

2.4.3 Temporal patterns of decision

As regards the time distribution of choices, our data set enables us to track the sequence of cards uncovered by subjects, as well as the amount of seconds they spend looking at each card inspected. Descriptive statistics across the four treatments of the variables used in this section are detailed in Table 2.8.

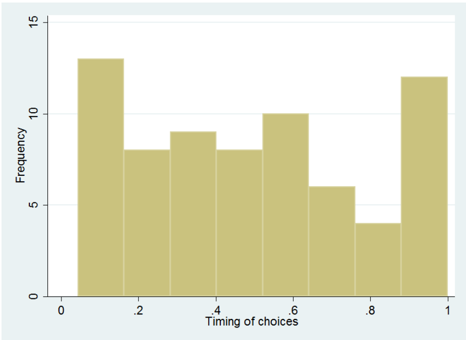
On average, subjects uncovered 22 cards each: this means, since the available cards for inspection on their screen are 16, that they uncover the same card more than once. Therefore, whenever this was the case, we decided to consider as timing of choice the last moment the chosen card was uncovered. For instance, if a subject uncovers the chosen card as the third card in his sequence of uncovered cards, but also as tenth card, we consider that the subject selects the tenth uncovered card as her final choice. Of course, this criterion pushes the distribution of choices towards the right tail, but we believe it is more realistic than other criteria (for example, considering the first time in which the chosen card was uncovered or a point in between the first and the last time in which it was uncovered).

Given this assumption, we construct the variable "timing of choices" by normalizing the order in which the chosen card was uncovered by the total number of uncovered cards for each subject. Panel (a) and (b) in Figure 2.6 display the frequencies for the timing of choices of participants in the high-cost and low-cost treatments, respectively.

From a first visual comparison, we can note that in the high-cost treatment there is a higher frequency of participants choosing among the recently uncovered cards. This is confirmed, although

Figure 2.6: Time distribution of choices

(a) Low-cost treatments



(b) High-cost treatments

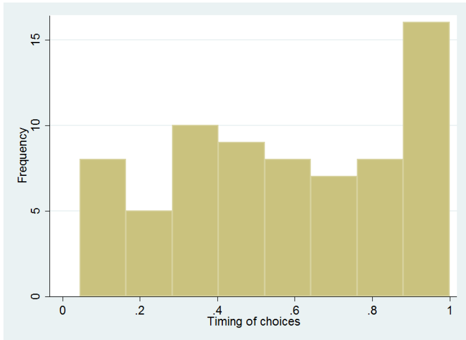


Table 2.8: Descriptive statistics - Temporal patterns

	Mean	Std. Dev.	Min	Max
Timing of choices				
Low-default	0.50	0.32	0.048	1
High-default	0.48	0.29	0.043	1
Low-majority	0.59	0.28	0.056	1
High-majority	0.55	0.30	0.045	1
Distance				
Low-default	0.18	0.24	0	0.791
High-default	0.23	0.24	0	0.846
Low-majority	0.30	0.24	0	0.875
High-majority	0.22	0.19	0	0.727

by a marginal significance, when we perform a Wilcoxon rank-sum test on the average timing of decision, comparing low-cost and high-cost scenarios (z = -1.662, p = 0.0965). When looking at the patterns in which the cards were uncovered and then chosen, we notice that all subjects, except one, uncovered the best card among the 16 present in the screen. That is to say, virtually all participants knew where the best card was. We hence analyze the distance, in terms of number of cards uncovered, between the the best card and the card chosen and we normalize it for the total amount of uncovered cards by each subject. We compare this distance across the high-cost and the low-cost treatments and we find that it is significantly higher in case of high-cost treatments (Wilcoxon rank-sum test, z = -2.437, p = 0.0148). Also this finding corroborates our Hypothesis 2:

Result 2.5. *More recent memories have a greater impact on choices, especially in scenarios with higher cognitive costs.*

2.5 Discussion and Conclusion

Cognitive costs characterize many decisions we frequently face in real life situations. This paper investigates the role played by cognitive costs in contexts involving social interactions and sheds more light on the determinants of imitative behavior.

Imitation is a very frequent pattern of real world behaviors but its sources have been limitedly examined by economic theory. Imitation is crucial in the transmission of knowledge and represents one of the main sources of learning.

We analyze the role of imitation in decision making through an experiment on choice behavior, in which we disentangle the potential drivers of imitative dynamics through the modeling of different levels of cognitive costs. For this purpose, we expressly devise a novel experimental task which is not purely mechanical like for instance counting tasks. Our results show that the experimental task implemented allows us to carry out a scenario with low cognitive costs and one with high cognitive costs, with the latter being characterized by lower performance levels. Moreover, we analyze temporal decisional patterns, since these might uncover relevant heuristics in decision making, especially in scenarios with high cognitive costs.

In our experiment, imitation, conceived as a cognitive shortcut, does not seem to be present in participants' choices, but we find higher expectations concerning the goodness of choice of a majority of subjects from a previous experimental session, if the task to be performed carries high cognitive costs. There is therefore some evidence of imitation of a majority as concerns beliefs. Moreover, we find that participants are statistically more overconfident in the high-cost treatments, and when they do not rely on choosing the default/majority card as their final decision. Therefore, choosing the short-cut of the default/majority card is correlated with underconfidence about one's own skills in solving the experimental task.

A possible explanation of this finding could be identified in the way in which the imitative component has been introduced in the experiment (i.e., by providing participants information about what a majority did in a previous experimental session): this indeed might have been too weak to generate some imitative behavior. For this reason, further research could be devoted towards an endogenous imitation pattern within the experiment: the majority could be formed during the experiment and feedback about it would be provided during the game. Another stream for future research might be represented by modeling differently the social component, and by strengthening it. This would be achievable by incorporating an in-group/ out-group framework in the experimental design.

Moreover, from the heuristics and biases literature introduced by the seminal work of Tversky and Kahneman (1974), we know that people tend to presume they outperform the average, as it is the case of participants in our experiment.[15] With this in mind, following what a majority previously did, without knowing the details of what kind of individuals the majority consisted of, might have been too risky for such kind of overconfident agents. For instance, the source credibility literature from marketing[16] highlights the importance of the source of an advertising message, in order to strengthen the validity of the message sent. This is the case of celebrity endorsements, in which the celebrity does not sponsor questionable items because of possible negative reputation effects. It is hence conceivable to design future follow-up studies of the current one with different information, across the treatments, about the composition of who has previously chosen the majority card. Telling this kind of information to participants could help increasing the attractiveness of this card and shed more light on the distinction between social and personal heuristics.

In this paper, we also provide empirical evidence that in the presence of high cognitive costs subjects implement decisional heuristics in order to find a solution to the task. More specifically, they attach more relevance to more recent memories.

The experiment studies the situation of a new decisional process, which does not have room for

[15]To the question "Do you think you chose a final card which is better, equal or worse than the card chosen by the majority?", included in the ex-post questionnaire, out of 154 individuals 71 replied they performed better than the majority, 71 equal and just 12 worse, thus confirming the overconfident pattern first proposed by Kahneman.

[16]For a review, see Pornpitakpan (2004).

learning dynamics since it is played only once by subjects. Nevertheless, for future research, it would be interesting to investigate how participants behave when learning dynamics are taken into account, and hence to see whether practice and repetition have an impact on participants' behavior.

We are aware of the concerns related to the external validity of laboratory experiments, and of the limitations of letting individuals interact with reference groups exogenously imposed. Nevertheless, our approach might mimic those situations in which there is a decisional default option such as in Internet contexts, in which frequently the most popular item is signalled and individuals might choose to opt for it instead of individually looking at its characteristics, which may be very costly in cognitive terms, regarding both effort and knowledge required.

2.A Appendix

2.A.1 Instructions for the "high cost - default" treatment (translated from Italian)

Welcome to this experiment![17]

You will receive 2.5 Euros as a showing-up fee for coming on time to the experiment. We kindly ask you to read the instructions carefully. During the experiment it is forbidden the communication with other participants. If you have any doubts and you want to ask a question, please raise your hand, an experimenter will come to you and will answer your question. If you don't respect these rules we will have to exclude you from the experiment and you will not be paid.

You are taking part to an economic experiment on decision-making. You can earn an amount of Euros depending upon your decisions during the experiment. Both your choices and other participants' choices will remain anonymous and will never be attached to your name.
Initially, for 15 seconds, it will appear on your screen:

- a card, on the top left. This card is called "TARGET" card. It contains a figure composed of many squares. As a square we mean the part represented in red in the figure below:

- below the TARGET card you will find 16 cards containing figures. Each figure differs from the figure contained in the TARGET card according to a different amount of squares.

Your aim is to find the card which contains the figure more similar to the figure contained in the TARGET card, **that is with the smallest amount of squares placed in a different position with**

[17]Instructions for the other treatments are available upon request. In Italics are the parts related to the majority treatments, for which DEFAULT card should always be replaced by textscMAJORITY card.

respect to the figure in the TARGET card.

Your final earning will be as greater as smaller is the amount of squares that differ from the figure of the card you chose and the figure on the TARGET card.

Your earning will be equal to the show-up fee plus:

- 4 Euros, in the case in which the card you chose is the best one, that is the most similar to the TARGET figure;
- 0.25 Euros, in the case in which you chose is the worst one, that is the least similar to the TARGET figure;
- the intermediate cards, between the best and the worst one, will provide you intermediate earnings between 0.25 and 4 Euros, with increases of 0.25 per card.

At the end of the initial 15 seconds, during which both the TARGET card and the 16 cards underneath are visible, all cards will be covered. On the bottom right of the screen will appear a covered card, called DEFAULT card. This card represents one of the best 8 cards (more similar to the TARGET card among the 16 you find covered. *[Majority treatments only] On the bottom right of the screen will appear a covered card, called* MAJORITY *card. This card has been chosen by the majority of participants in a preceding session of the experiment and moreover it represents one of the best 8 cards (more similar to the* TARGET *card among the 16 you find covered.* You can choose whether to select this card as your final choice, or to uncover and inspect the other 16 cards that appear on the left of the DEFAULT card.

In case you choose the DEFAULT card, this represents your conclusive choice and you will not be able to uncover it and then inspect the other 16 cards. Your earning in case you choose the DEFAULT card depends on the similarity of the DEFAULT card with respect to the TARGET card. About the DEFAULT card you are only informed that it is one of the best 8 cards among the 16 present on its left. *[Majority treatments only] About the* MAJORITY *card you are only informed that it has been chosen by the majority of participants in a preceding session of the experiment and moreover that it represents one of the best 8 cards among the 16 you find covered.* **If you choose it, this represents your conclusive choice.**

If you choose to inspect one of the 16 cards, you have to click on it. The card will be enlarged and displayed beside the TARGET card. Once you decide to conclude the visual comparison between the selected card and the TARGET card, in order to bring back to the initial position the selected card you must re-click on it. In this way, the card you have just compared will return to its initial position and and its border will become red, to indicate that the card has already been inspected.

You can decide whether to choose this card as you definitive card, and in this case you must re-click on it, or to proceed with the visualization of another card. In case you want to visualize again a card which has already been inspected (therefore with the red border), you can re-click on it. The computer will ask you whether you want to select this card as your final choice or if you simply want to visualize it again.

If you choose to inspect one or more of the 16 cards, at any moment you will be able to interrupt your search and to choose the DEFAULT card. In this case the experiment ends and your earning will depend on the similarity between the DEFAULT card and the TARGET card.

The time available to you is 10 minutes, starting after the reading of these instructions. **When the time expires, if you haven't selected any card yet, your final earning will be equal to the 2.5 Euros of the show-up fee.**

Once you have selected your conclusive card, we ask you to remain seated and silent, while waiting for the other participants to complete the experiment. After you made your choice (either one of the 16 cards or the DEFAULT card), the computer will ask you to specify the amount of squares for which you think the DEFAULT card) (that you cannot visualize) differs with respect to the TARGET card. In case your answer is correct, you will receive an additional 1 Euro to your final earning.

Your final earning will paid to you in cash and privately, so that the other participants to the experiment will not know your earning.

2.A.2 Ex-post questionnaire

1) How accurate do you think your final choice was? Indicate the number of squares for which you think your final card differed compared to the TARGET card [18]

(5/ 6/ 7/ 8/ 9/ 10/ 11/ 12/ 13/ 14/ 15/ 16/ 17/ 18/ 19/ 20)

2) On a scale of 1-10, with 1 being the lowest grade of difficulty and 10 the highest, how difficult did you consider the experimental task?

(1/ 2/ 3/ 4/ 5/ 6/ 7/ 8/ 9/ 10)

3) Do you think you chose a final card which is better, equal or worse than the card chosen by the majority?

(better/ equal/ worse)

4) What do you think the TARGET card represented?

5) Among the cards, do you think there was one that will most probably be chosen by the majority? Why?

(yes/ no)

6) In which faculty are you enrolled?

(Economics/ Humanities/ Engineering/ Law/ Physics and Mathematics/ Sociology or Psychology/ Other Natural Sciences/ Other Social Sciences/ not a student)

7) In which year of studies are you enrolled?

(first year/ second/ third/ out-of-due date/ master student/ not a student)

8) Which is your sex?

[18]In the high-cost treatments, due to the different resolution of figures, the options available to subjects for this question were: (20/ 24/ 28/ 32/ 36/ 40/ 44/ 48/ 52/ 56/ 60/ 64/ 68/ 72/ 76/ 80). The rest of the questionnaire did not present any differences across treatments.

(female/ male)

9) In which year were you born?

2.A.3 Descriptive statistics - Baseline session

Table 2.9 reports participants' choices of cards in the low-cost and high-cost treatments of the baseline session, respectively.

Table 2.9: Chosen cards - Baseline session

	Low-cost treatment		High-cost treatment	
	Frequency	Percentage	Frequency	Percentage
Fitness chosen card				
1	3	30	2	20
2	1	10	-	-
3	-	-	1	10
4	-	-	-	-
5	1	10	-	-
6	1	10	-	-
7	1	10	1	10
8	-	-	1	10
9	-	-	-	-
10	2	20	1	10
11	1	10	1	10
12	-	-	1	10
13	-	-	1	10
14	-	-	1	10
15	-	-	-	-
16	-	-	-	-
Mean	5.4		8	
St. Dev.	4.03		4.87	
Min	1		1	
Max	11		14	

Table 2.10: Descriptive statistics - Baseline session

	Low-cost treatment		High-cost treatment	
	Frequency	Percentage	Frequency	Percentage
Gender				
Female	5	50	6	60
Male	5	50	4	40
Faculty				
Economics & Management	3	30	5	50
Engineering	1	10	1	10
Law	3	30	1	10
Literature	1	10	1	10
Math	1	10	-	-
Not a student	-	-	1	10
Sociology	1	10	1	10
Degree level				
Out of due date student	2	20	1	10
Master student	2	20	2	20
Not a student	-	-	1	10
1st year student	2	20	2	20
2nd year student	3	30	2	20
3rd year student	1	10	2	20
Age				
Mean	22.4		22.6	
St. Dev.	1.955		2.22	
Min	20		20	
Max	25		27	
Payoff				
Mean	5.4		4.75	
St. Dev.	1.008		1.219	
Min	4		3.25	
Max	6.5		6.5	

In Table 2.10 we report other descriptive statistics from the baseline session.

2.A.4 List of variables

Table 2.11 includes a list of the variables used in the analyses.

Table 2.11: List of variables

Variable	Description
Male	Dummy variable assuming value 1 if the participant is male and value 0 otherwise

Continued on next page

Table 2.11 – *Continued from previous page*

Variable	Description
Age	Age in years
Payoff	Total payoff earned by the participant at the end of the experiment
Net payoff	Payoff earned by the participant excluding the 1 Euro gained in case of correct guess of the fitness of the default/ majority card
Seconds	Amount of milliseconds spent in choosing one card on the screen
N. uncovered cards	Number of cards uncovered by the participant in solving the task
Low-default	Indicates a low-cost treatment in which the card at the bottom right of the screen is a default card
High-default	Indicates a high-cost treatment in which the card at the bottom right of the screen is a default card
Low-majority	Indicates a low-cost treatment in which the card at the bottom right of the screen is a majority card
High-majority	Indicates a high-cost treatment in which the card at the bottom right of the screen is a majority card
Info	Dummy variable assuming value 1 if the participant takes part in a majority treatment (either low-majority or high-majority) and value 0 otherwise
Cost	Dummy variable assuming value 1 if the participant takes part in a high-cost treatment (either high-default or high-majority) and value 0 otherwise
Belief default	The ratio between the real difference in black squares between the target and the default/majority card, and the believed corresponding difference in black squares declared at the end of each session by the participant (answer to the question included in footnote 10 of Chapter 2). The variable ranges between 0.25, if the participant believes that the default/majority card is the worst, in terms of similarity with the target card, and 1, in case she believes it is the most similar card with respect to the target
Believed accuracy	Number of squares the participant believes differ between her chosen card and the target card. It has been normalized to values from 1 to 16, where 1 indicates that the participants believes she has chosen the best card possible and 16 indicates the worst card (participant's answer to Question 1 of the ex-post questionnaire, normalized to 1-16)

Continued on next page

Table 2.11 – *Continued from previous page*

Variable	Description
Actual accuracy	Number of squares of difference between the participant's chosen card and the target card. It has been normalized to values from 1 to 16, where 1 indicates that the participant has chosen the best card and 16 indicates that she has chosen the worst card
Absolute under/over-confidence	Believed accuracy - actual accuracy. It ranges from -15 (maximum over-confidence) to +15 (maximum underconfidence)
Difficulty	Difficulty declared about the task (participant's answer to Question 2 of the ex-post questionnaire)
Relative under/over-confidence	Believed performance of the participant compared to the majority: better/ equal / worse. (participant's answer to Question 3 of the ex-post questionnaire)
Choice majority	Dummy variable assuming value 1 if the participant believed that one card will most probably be chosen by the majority of participants (participant's answer to Question 5 of the ex-post questionnaire)
Timing of choices	Order in which the chosen card was uncovered, divided by the number of cards uncovered by the participant (it ranges from 0 to 1: the smaller the value of the variable, the sooner the participant uncovered the card which she finally chooses; the higher the value of the variable, the later the participant inspected the card which she finally chooses)
Distance	Distance in terms of number of uncovered cards between the best card in absolute and the card chosen, divided by the number of cards uncovered by the participant (it ranges from 0 to 1)

3 Exploring information aversion: an experimental analysis[1]

3.1 Introduction

In neoclassical economic theory information is deemed as valuable only as a means for improving decision making. Accordingly, individuals should collect as much information as possible as it will allow them to fine-tune their choices (Stigler, 1961).

Yet, a growing number of studies in various fields, such as Economics, Psychology and Medicine, point in the opposite direction and suggest that decision makers assign a value to information that goes beyond the instrumental role considered by traditional economic theories. Every day life offers many instances of such anomalous attitudes towards information: we avoid information when we fear to receive bad news and we look for information when we expect good news or we want to confirm our opinions. This is especially true in health contexts, in which emotions such as anxiety and fear play an important role in driving decisions about medical treatments and preventive behavior in general. Consider for instance a patient who neglects medical tests when she has some clues of being ill, while she performs more accurate tests when she is almost certain of being healthy. In line with this, Lerman et al. (1998) show that 46% of subjects with an hereditary history of breast and ovarian cancer who perform a blood test to check for genetic mutations associated to these illnesses refuse to receive test results. Likewise, but with a social interaction component, Zanella and Banerjee (2014) find in a U.S. workplace environment that if a co-worker is diagnosed with breast cancer, the probability of subjects to perform a mammography decreases.[2]

[1]This chapter is based on a joint work with Marie Claire Villeval. This research has been supported by a grant from the French National Research Agency (ANR, EMCO program, HEIDI grant, ANR-11-EMCO-011-01) and was performed within the framework of the LABEX CORTEX (ANR-11 LABX-0042) of Universit de Lyon, within the program Investissements dAvenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). Simona Cicognani gratefully acknowledges financial support from PALSE and the School of Social Sciences of Trento.

[2]Other empirical studies indicating information avoidance in health related issues and procrastination in going to a doctor if symptoms of illness are evident include Lyter et al. (1987), Facione (1993), Caplan (1994) Lerman et al. (1999),

The asymmetric impact of good and bad news on the acquisition of complementary information about own condition has been recently theorized and investigated as the so-called "ostrich effect".[3] Karlsson et al. (2009) find clear evidence of this selective attention phenomenon in a financial context, analyzing data on investment decisions: after a preliminary good signal about portfolio returns, as represented by generic financial news on media, private investors are more willing to check their online bank accounts in order to obtain conclusive information that reassures them. The contrary occurs in case of preliminary bad news, after which subjects are less prone to look for precise and definitive information in their online financial prospects.[4]

Selective attention towards information may originate from the inclusion of beliefs in subjects' utility functions, as opposed to the traditional von Neumann-Morgenstern (vNM) utility function. Information would be a means for decision makers to nurture desired beliefs and to overcome unpleasant beliefs (Eliaz and Spiegler, 2006). Hence, anticipatory utility should also be taken into account when considering individuals' welfare and to predict policy implications. If people derive utility from beliefs, they may manipulate sources of available information and obtain sub-optimal outcomes if their choices were different from the ones undertaken under a full information set. It follows that the amount of information acquired is not trivial and gives rise to an important trade-off between emotional costs from information and better decisions deriving from more information.

We investigate this informational trade-off by devising a novel experiment composed of two experimental sessions, with the second session taking place two weeks after the first one. In the first session participants receive a fuzzy signal about their condition, which may become more informative if they decide to buy additional information in week 1. Between week 1 and 3 they are given the opportunity of completing an online task, on three nonconsecutive days, which increases their expected payoff. The completion of effort has a higher impact on their expected payoff if participants' condition is bad rather than good. By interacting the possibility of acquiring additional information with the possibility of exerting effort, we aim to capture and investigate the informational trade-off at the basis of the ostrich effect.

Our experiment departs from the existing research in two dimensions. First, it enables us to study the main characteristics of the ostrich effect without providing a specific framing. Second, we intro-

Nosarti et al. (2000), Richard et al. (2000) and Meechan et al. (2002).

[3]The term was coined by Galai and Sade (2006) and theorized by Karlsson et al. (2009).

[4]Golman and Loewenstein (2013)) present a modified version of Karlsson et al. (2009), which includes a unified theoretical framework that aims to explain a wide array of anomalous attitudes towards information, such as curiosity or the desire for wisdom.

duce the possibility of exerting an effort that increases expected payoffs. This may contribute to shed some light on policy implications of the ostrich effect, as we are interested in knowing whether the opportunity to acquire more specific information boosts subjects to exert effort.

Our findings show that participants significantly buy more information after a preliminary bad signal than after a good signal. Accuracy of signals has an impact on the exertion of effort only when the signal is bad: in this case participants exert significantly more effort than after receiving a good signal, which lends support to the importance of additional information acquisition if someone deems to be in a bad situation. Hence, participants understand the incentives of completing effort after receiving a bad signal about their condition only when they decided to be more informed. Effort exertion and information acquisition seem to be substitutable rather than complementary: providing participants with the possibility of acquiring additional information lowers their effort exertion. Moreover, the possibility of acquiring additional information about own condition points out a duality of individuals: those who decide to be more informed and thus follow a rational behavior in the exertion of effort, and those who decide to remain with a limited information set and behave irrationally in the exertion of effort.

The remainder of the paper is organized as follows. Section 3.2 presents the related literature and Section 3.3 details the experimental design and the procedures followed to conduct the experiment. In Section 3.4 we present the behavioral predictions we aim to test. Section 3.5 provides the experimental results and Section 3.6 discusses the results and concludes.

3.2 Literature review

A few psychological theories are closely connected to the ostrich effect and envisage that individuals may influence information acquisition in order to seek consistency in their beliefs. According to the cognitive dissonance theory, theorized in the seminal work of Festinger (1962), individuals would naturally tend towards an harmony of attitudes and beliefs and would hence face discomfort whenever they cope with new information that contrasts existing beliefs or values. Inconsistent or conflicting beliefs lead to a discomfort that individuals would seek to avoid, but in trying to eliminate the cognitive discrepancy they may act irrationally. Different ways to reduce the dissonance in beliefs encompass: a change in the attitudes causing the dissonance, the acquisition of new information that compensates the dissonant belief or the reduction in the relevance of the conflicting belief. For instance, a smoker may enjoy smoking but at the same time she may feel uncomfortable for the negative health effects driven by her behavior. Possible ways to reduce this discomfort may be represented by quitting smoking, acquiring new information that reduces the belief that smoking leads to lung cancer, or convince herself that the benefits of smoking outnumber the damages of it.

Closely interconnected with cognitive dissonance, the confirmation bias theory postulates that when people search for or interpret information, they tend to confirm their own beliefs, especially when emotional issues are involved (Nickerson, 1998). Likewise, according to Klayman and Ha (1987), scientific research has the tendency to implement positive test strategies, which means that researchers are more likely to test those cases characterized by the highest probability of verifying existing beliefs, rather than those cases which would probably falsify them. According to this cognitive bias, poor decisions can be undertaken in that it distorts the informational capacity of individuals.

Conversely, the standard paradigm of the economics of information failed to explain many of the logics for which people acquire or avoid information. For this reason in the past decades several economic studies addressed the idea of anticipatory feelings in order to account for the behavioral anomalies towards information registered in real-life scenarios. Akerlof and Dickens (1982) are the first ones to extensively incorporate beliefs in a vNM utility function. They do so by considering economic agents who are not emotionally neutral towards their beliefs, which can be influenced either through self-persuasion or through the selection of informational sources and signals.

A common feature of studies addressing anticipatory feelings is the Bayesian belief updating of subjects, originated by a first imprecise signal of their own condition. Caplin and Leahy (2001) build a psychological expected utility model that explicitly takes into account a wide class of anticipatory emotions, in which the main innovation is that agents' psychological states are directly connected to their utility level. They model how lotteries over future physical outcomes influence these emotions. Building on Caplin and Leahy (2001), Koszegi (2003) introduces a patient-doctor model of free information acquisition by patients, in which the focus is placed on the patient's decision making. The model suggests that, in case of possible very bad news, the patient completely avoids the doctor. Similarly to Koszegi (2003), Caplin and Leahy (2004) present a patient-doctor model. Here the doctor observes the health condition (either good or bad) of a patient and the doctor's aim is to maximize the patient's utility function. Patient's posterior belief about her own health state is incorporated in the model. There exist two types of patients: one who wishes to have a posterior belief about her health condition as close as possible to her prior belief, whereas the other one has opposite preferences. The patient first reveals her type and then the doctor decides to reveal or not her health state.

Few studies have addressed anticipatory utility from a policy perspective. In Caplin and Eliaz (2003), a game theoretical model examines the optimal testing mechanism for AIDS. By taking into account psychological factors such as anxiety in undertaking the test, as well as the externality involved in the spreading of this illness, they show that partially revelatory certificates could enhance people to undertake AIDS tests. A similar policy implication is put forward by Schweizer and Szech (2013). They aim to investigate the optimal mechanism of conveying life-changing information, which results in providing either precise good news or noisy bad news. Only if the diagnosing system is designed in such a manner it would be easier to be undertaken by patients.

Other studies propose different reasons for avoiding or collecting information, respectively. Carrillo and Mariotti (2000) consider strategic ignorance as a self-control device for those individuals who may fear the impact of information on their beliefs. For instance, a non-smoker might want to ignore the real harmful effects of tobacco on health in order to prevent herself from smoking. In fact, she may discover that the objective probability of getting lung cancer from smoking is lower than her subjective probability.

Conversely, a case of anomalous information acquisition is proposed by Eliaz and Schotter (2010), who study the demand for non-instrumental information, that is information which is useless in terms of decision making. Their findings show that people derive an intrinsic utility from posterior beliefs, as they are willing to pay in order to be more confident about the choices they made. In particular, people attach more weight to bad news, as they prefer to know whether the worst case scenario has occurred, rather than the best case scenario.

In this paper we present a laboratory experiment that examines the key dynamics of the ostrich effect, such as the existence of preliminary non-conclusive information, either positive or negative, and a following optional decision of acquiring more conclusive information. The novelty of our study is the possibility of exerting an effort that improves participants' expected outcomes. The effort aims to proxy a prevention component, that introduces policy implications from a new perspective compared to the optimal diagnosing system proposed by Caplin and Eliaz (2003) and Schweizer and Szech (2013).

Few attempts have been made to investigate the extent of information aversion experimentally. Brown and Kagel (2008) mirror the financial setting of Karlsson et al. (2009), with an experiment that simulates a stock market. If the ostrich effect is present, investors should look at their stocks more often when they hold a winning stock than when they hold a losing stock. The authors find evidence of a status quo bias rather than of the ostrich effect.

The topic of our study is also related to the "good news-bad news effect", which mirrors the ostrich effect in representing the aversion to acquire new information when subjects expect negative news. This effect has been tested in two recent papers, in which the relevant information displayed refers to very personal characteristics, such as beauty or IQ scores. Mobius et al. (2011) find an asymmetric updating of beliefs about subjects' performance to an IQ quiz: participants over-weight positive feedback of their IQ scores relative to the negative feedback. Also Eil and Rao (2011) use IQ scores, besides beauty scores based on a ranking of all participants made by participants themselves. Their results reveal the existence of aversion to new information when subjects receive negative feedback about their beauty or IQ.

3.3 Experimental design and procedures

3.3.1 Design

The experiment comprises two sessions in a longitudinal framework: session 1 occurs in week 1, whereas session 2 takes place in week 3. Subjects are paid only at the end of week 3 and are informed of this at the recruitment. At the recruitment subjects are also asked their availability for the date of session 2, in order to minimize attrition in the data. In the experimental instructions of session 1 they are informed of the possibility of completing an online task between week 1 and 3, which might have an impact on their final payoff. A brief description of the task is provided in the instructions of session 1 in order not to push subjects to discover the online task simply out of curiosity.

3.3.2 Baseline treatment

We follow the longitudinal framework of the experiment, by first presenting session 1 and then proceeding with session 2.

Session 1. In week 1 participants come to the laboratory and they are told each of them is assigned a box containing 20 balls. They also know that half of them is assigned box A, containing 14 white balls and 6 black balls, and half of them box B, with 6 white balls and 14 black balls inside. They don't know their own box composition, but they are aware that the different box content in terms of balls colors entails a different expected payoff. In fact, at the end of session 2, the computer will randomly extract a ball from each box, and if the ball extracted is white the corresponding participant wins 25 Euros, otherwise she wins 15 Euros.

After being told this, a random extraction of 3 balls from each participant's box occurs, without reinsertion. The color of the three balls extracted is displayed on each computer screen, as a signal of the box composition.

Finally, participants are asked to fill in a computerized questionnaire including demographic questions, such as gender, age and field of studies.[5]

Between session 1 and 2. Between the first and second experimental sessions, in which participants are requested to come to the laboratory, they are given the opportunity to complete an online task that has an impact on their expected earnings at the end of session 2. They are provided with this opportunity on three nonconsecutive dates, distant 2-4 days one to another. On all these dates, they receive an email reminding them of this opportunity, containing the link to the online task.

The task involves replicating a series of blurred codes correctly, between 00:00 and 23:59 of the specific dates devoted for the task.[6] Subjects should replicate the codes in succession, and after completing them they receive an email indicating the correct replication of the codes. The rationale for this task, which is novel in an experimental setting, is to provide participants with a boring assignment, not related to their intellectual capacities and not capable of entertaining them. Moreover, as the experimenter does not benefit from the task completion, we avoid any reciprocity concerns from the subject towards the experimenter.

For the task to be effective for an increase in expected earnings, it should be completed in all three dates. Participants know that the exertion of the task has two different effects depending on the type of box attached to them. If a participant is given box A and completes the task successfully in all three dates, two black balls from her box become white. If a participant is assigned box B, four black balls become white. If no task is performed or if it is not completed in its entirety in all the due dates, the initial composition of participant's box is maintained.

Session 2. On the same day of the week of session 1, but two weeks later, participants come back to the laboratory and take part in session 2.[7] Payoffs in this session are calculated in points and converted into Euros at the end of the experiment (1 point = 0.20 Euros).

We elicit subjects' risk preferences by adopting Eckel et al. (2012)'s task.[8] This is a very simple

[5] See Appendix 3.A.2.

[6] This type of codes are called CAPTCHA in computer science terminology.

[7] A flexibility of up to two days for the date of session 2 is allowed to participants in case of unforeseen circumstances that otherwise would preclude their participation to session 2.

[8] The task was first introduced by Eckel and Grossman (2008). We use Eckel et al. (2012)'s simpler version, focusing on

risk measurement mechanism, which enables to disentangle subject's preferences about risk from their ability to perform mathematical computations. The task entails a choice among six lotteries, which include one sure gamble with no risk (variance) and five more gambles, with both variance between the payoffs and expected return increasing if we move from lottery 1 to lottery 6. The representation of lotteries is circular, with the high payoff on the right of the circle and the low payoff on the left.

Once all participants complete the risk elicitation task, the computers notify to subjects their box type and a random extraction of one ball from each box takes place. The final box composition may have changed from the initial composition only due to the correct completion of the three online tasks between session 1 and 2. The color of the extracted ball is privately displayed on screens, as well as the result of the risk elicitation task.

After the definitive extraction, participants are asked to fill in a computerized questionnaire, which includes questions regarding their health status and their procrastination propensity.[9] The rationale for this is to investigate a potential correlation between information aversion and health related aspects. Finally, payment procedures follow in a separate room.

3.3.3 Additional information treatment

The main purpose of this treatment is to test the existence of the ostrich effect in a context in which subjects can improve their own condition by exerting effort.

The Additional Information Treatment (AIT) differs from the Baseline Treatment (BT) only in Session 1, as can be noted in Table 3.1. After telling participants that an extraction of 3 balls will occur from their urn, we ask them to indicate for each possible combination of 3 extractable balls whether they are willing to pay 1 Euro for the extraction of two additional balls from their box. The 1 Euro would be subtracted from their final payoff at the end of session 2. We thus apply the strategy method to elicit participants' willingness to pay for additional information, which provides them a better signal of their own box composition.[10]

After this elicitation, the random extraction of three balls occurs and their color is privately displayed on each participant's screen. The total number of balls displayed follows participants' choices from the strategy method. Hence, in case the participant chose to acquire the extraction of two ad-

the gain domain, in order to avoid loss aversion considerations.

[9]The ex-post questionnaire is included in Appendix 3.A.5.

[10]The strategy method was first introduced by Selten (1965). See Charness and Rabin (2005) for a discussion of the strategy method in experiments.

Table 3.1: Procedures across treatments

	BT	AIT
Session 1	control questions extraction of 3 balls demographic questionnaire	control questions additional information acquisition decision (strategy method) extraction of 3/5 balls demographic questionnaire
Between session 1 and session 2	possibility of online task on 3 nonconsecutive days	possibility of online task on 3 nonconsecutive days
Session 2	risk aversion task extraction of the ball from own box ex-post questionnaire payment	risk aversion task extraction of the ball from own box ex-post questionnaire payment

ditional balls for the combination of balls extracted, a total of five balls is displayed on the screen. Otherwise, the participant observes a total of three balls.

3.3.4 Procedures

We conducted the experiment between May and June 2014 at the CEEL laboratory of the University of Trento, Italy. The experiment was computerized and programmed with Borland Delphi. 123 subjects were recruited with the online software used at CEEL, that reaches a pool of around 2000 individuals from different faculties of the University.[11] Since participants in a few circumstances could earn more than 25 Euros, at the recruitment they were asked to be of EU citizenship - Switzerland included - and to hold a bank account, due to Italian jurisdiction laws.

Participants were quizzed on their understanding of the experiment after reading the instructions of session 1 (control questions). Answers were provided in private and their accuracy was checked. Each session consisted of various stages, and a final questionnaire was administered to subjects at the end of session 2.[12] Both sessions lasted on average 25 minutes, whereas each online session lasted approximately 5 minutes. Subjects could not attend more than one session of the experiment; they were paid at the end of the second session in a separate room and received on average 22.3 Euros.[13]

We ran three sessions for the BT and five for the AIT, with 44 and 79 participants, respectively. Three subjects did not show up at session 2, therefore we excluded them from the final data set. Overall the experiment consisted of 16 sessions, of which eight in week 1 and eight in week 3. Upon arrival at the laboratory, each participant drew a tag from a bag, indicating the number of her corresponding computer. Computers were allocated in separated cubicles. Participants also had to randomly pick from an urn a tag which reported their secret code for the experiment, to be entered on computers at session 1 and 2, and also before performing the tasks between the two sessions.[14] With this procedure we were able to track participants' choices and at the same time to assure their

[11] These include Economics and Management, Engineering, Math, Physics, Chemistry, Literature, Computer Science and Sociology.

[12] At the beginning of session 1, participants were informed of the existence of various stages in the experiment, but were not provided with details of each stage in order to avoid strategic behavior. Instructions were neutrally-framed and are available in Appendices 3.A.1 and 3.A.4, whereas a screen-shot of the online task is included in Appendix 3.A.3.

[13] At the end of session 2, subjects were asked to fill in the form for payment and to come to a separate room for payment. Those whose payoff was lower than 25 Euros were paid in cash, whereas for payments above 25 Euros we paid participants by bank transfer, due to Italian laws on cash use. We did not inform subjects in advance regarding this different payment scheme, and they could not be familiar with it since payments above 25 Euros are a real exception at the laboratory in which this experiment was run.

[14] In order not to lose the observations of those subjects who might have lost or forgot the code, we asked them, at the end of session 1, to pick a copy of their secret code and to put it in an envelope, which was sealed in front of them and kept at the laboratory. Each participant's name was written on each respective envelope, as a guarantee in case of code loss. Envelopes were destroyed at the end of the experiment.

anonymity. At the beginning of each stage, subjects received the instructions for the corresponding stage and read them for 5 minutes, after which instructions were read aloud by a member of the laboratory.

3.4 Theoretical background and hypotheses

In this section we highlight the key components of the experimental setting, in connection with the behavioral predictions we aim to test.

First, we represent a generic scenario in which the ostrich effect can take place. Two different types of individuals, characterized by different endowments (box types), face a stochastic phenomenon: the possibility of exerting effort between session 1 and 2 increases participants' expected payoff, but the payoff is eventually determined by the final random extraction of one ball from their box. Hence, effort is not deterministic: even if subjects exert it, they may end up at week 3 with a black ball extracted, which corresponds to the low payoff.

Second, the longitudinal framework of the experiment over two weeks creates a considerable temporal dimension in which subjects face uncertainty about their own condition. They could choose to reduce this uncertainty by buying additional information about their own condition, in week 1 of the AIT.

Third, the informational trade-off which characterizes the ostrich effect is introduced in the experimental design via the asymmetric expected gain from the tasks completion: we consider the tasks completion to be more fruitful in terms of final expected payoff in case a participant is attached to box B, which is the bad box. This adds a more realistic component to the model, as it is likely that, for instance in a health scenario, the exertion of activities such as exercising or preventive effort in general have a higher impact on individuals who are initially characterized by a bad condition. As we incorporate in the model decreasing marginal benefits of effort, the acquisition of information is clearly more useful if participants are attached to the bad box.

Signals about own condition are crucial for modeling the ostrich effect. In week 1 of the BT, participants receive preliminary signals, via the random extraction of three balls from their box, without reinsertion. At this stage participants form their initial beliefs about the type of box attached to them. Beliefs following a Bayesian pattern in the BT are depicted in the left-hand side of Table 3.2, and are computed according to the formula:

$$P(A|X) = \frac{P(X|A) \times P(A)}{P(X)}, \tag{3.1}$$

where A indicates the good/bad box and X the type (white or black) and number of balls extracted. For instance, a participant who has two white balls and one black ball displayed on her screen reports a Bayesian probability of 72.22% of being attached to the good box, versus a Bayesian probability of 27.78% of holding the bad box.[15]

Table 3.2: Bayesian beliefs about box types after an extraction of 3 vs 5 balls

3 balls extracted				5 balls extracted				
# White	# Black	Bayes pr. good	Bayes pr. bad	# White	# Black	Bayes pr. good	Bayes pr. bad	Signal
3	0	94.79%	5.21%	4	1	96.62%	3.38%	C
				5	0	99.70%	0.30%	C
				3	2	75.00%	25.00%	C
2	1	72.22%	27.78%	3	2	75.00%	25.00%	C
				4	1	96.62%	3.38%	C
				2	3	25.00%	75.00%	S
1	2	27.78%	72.22%	2	3	25.00%	75.00%	C
				3	2	75.00%	25.00%	S
				1	4	3.38%	96.62%	C
0	3	5.21%	94.79%	1	4	3.38%	96.62%	C
				2	3	25.00%	75.00%	C
				0	5	0.30%	99.70%	C

Notes: The column Signal refers to a confirmatory (C) or switching (S) signal with respect to the initial 3 balls signal. A confirmatory signal maintains the domain (positive or negative) of the Bayesian probability of being attached to the good box, whereas a switching signal entails the overturning of the domain of holding the good box.

We distinguish among three types of signals: a good signal, if the Bayesian belief of being attached to the good box is greater than 50%, versus a bad signal, which occurs if the Bayesian belief of being attached to the good box is smaller than 50%. A Bayesian probability equal to 50% of having the good box refers to an uninformative signal.

In the AIT signals can become more informative with respect to the BT: participants have the possibility of paying 1 Euro for the extraction of two additional balls from their box, without reinsertion. The aim of information acquisition is to be useful in informational terms with respect to the extrac-

[15]Participants' beliefs about being in the good or in the bad box were not elicited, but will be elicited in future experimental sessions. For further details of future research, see Section 3.6 and Chapter 5.

Table 3.3: Different types of signals in terms of detail

Types of signals		3 balls extracted		5 balls extracted	
		# White	# Black	# White	# Black
Good	Excellent	3	0	5	0
				4	1
	Decent	2	1	3	2
Bad	Poor	1	2	2	3
	Very poor	0	3	0	5
				1	4

tion of three balls: the two additional balls either provide a more informative signal of the initial condition (confirmatory signal), or they change the domain of the belief, from a positive to a negative domain, or viceversa (switching signal). Bayesian beliefs in case of 5 balls extracted are detailed in the right-hand side of Table 3.2.

In order to better investigate the effect of different kinds of signals, we decompose good and bad signals into more detailed signals, which are portrayed in Table 3.3. As a result, a good signal is split into an excellent and a decent signal, whose corresponding Bayesian probabilities of being in the good box are both higher than 70%. Conversely, a bad signal is decomposed into a poor and a very poor signal, for which the Bayesian probabilities of being attached to the good box are both lower than 30%.

The research hypotheses we aim to investigate cover three main issues: the exertion and completion of effort, the acquisition of additional information and the relationship between additional information acquisition and health-related variables.

Effort. The comparison between the BT and the AIT allows us to determine whether the possibility of having a more precise information about own condition has an impact on effort exertion. The underlying rationale is to investigate whether there is complementarity or substitutability between effort exertion and information acquisition. Thus, to explore if subjects, provided with the possibility of acquiring additional information as in the AIT, are driven to complement this information with effort or rather prefer to substitute it with effort. However, it is not clear to determine *a priori* which effect will prevail.

We can further examine the availability of additional information to participants by comparing the effort of those who saw five balls with the effort of those who saw three balls, conditional on having received a good or a bad signal.[16] This comparison enables us to test whether a higher accuracy of

[16] By type of signal we refer to a good versus a bad signal, whereas the accuracy of the signal refers to having observed

the signal leads participants to exert more or less effort, depending on the type of signal received. Since, by the experimental design, effort has a higher impact on expected payoff if a participant holds the bad box, having a more accurate signal of the type of box assigned may influence effort exertion. If significant differences among the types of signals and their accuracy are found, we can presume that the acquisition of information is driven by the willingness to have better information regarding the differential impact of the effort that participants may decide to exert.

Hypothesis 3.1. *Effort is higher the worse is the signal and the higher is the accuracy of the signal. Therefore:* $e_{5,BS} > e_{3,BS} > e_{3,GS} > e_{5,GS}$,

where e stands for effort, the first subscript refers to the number of balls displayed and the second subscript refers to having received a bad signal (BS) or a good signal (GS).

Hypothesis 3.1 hinges on the perceived probability of being attached to the bad box. The higher this probability, the higher is the expected payoff deriving from the completion of effort. Therefore, irrespective of the effort cost function, the higher the perceived probability of the bad box, given a certain accuracy and type of signal, the higher should be the effort exerted for that specific informational condition. Indicating by P the probability of having a bad box, it follows that:

$$P_{5,BS} > P_{3,BS} > P_{3,GS} > P_{5,GS} \Rightarrow e_{5,BS} > e_{3,BS} > e_{3,GS} > e_{5,GS} \quad (3.2)$$

Computations underlying the validity of the first part of Equation 3.2, which leads us to formulate the second part (Hypothesis 3.1), are reported in Appendix 3.A.7.

Moreover, we can distinguish the exertion of effort between those who had the opportunity to buy additional information and bought it ($e_{5,AIT}$) versus those who did have the same opportunity but decided to shy away from it ($e_{3,AIT}$). Comparing these two types of participants we can shed light on the differences in effort exertion between those who decided to be better informed and those who denied a better informational set. This comparison will allow us to dig more deeply in the complementary/substitutability relationship between effort and information. Yet, it is not clear which rationale may eventually prevail.

Additional information acquisition. By focusing on the AIT, it is possible to test the existence of the ostrich effect and correlate it with other issues involved in the process of information acquisition. The ostrich effect arises if the information acquired after a preliminary bad signal is lower than after

either five or three balls.

a preliminary good signal. This leads us to formulate the following testable prediction:

Hypothesis 3.2. *Information acquisition (I) after a preliminary bad signal is lower than after a preliminary good signal:* $I_{BS} < I_{GS}$

Additional information acquisition and health-related variables. Albeit the experimental set up is not framed as a health context, we aim to shed light on possible connections between information aversion and health-related variables. For instance, smoking attitudes may entail present-biased preferences, as highlighted by O'Donoghue and Rabin (1999) and Downs et al. (2009). Smokers may concentrate on present utility and avoid information, even if information may be relevant for the future in terms of expected payoff from effort.

Hypothesis 3.3. *Smokers are more information-averse than non-smokers:* $I_{smokers} < I_{non-smokers}$

Individuals' self-reported health-status may be considered in parallel with a preliminary signal about their own condition. Hence, we posit that those who self-report a low health condition receive a bad signal and hence are less willing to acquire additional information about their condition:

Hypothesis 3.4. $I_{low-health} < I_{high-health}$,

where low-health is defined as a health level between 1 and 8 whereas high-health corresponds to a health level between 9 and 10.[17]

Likewise, we suppose that procrastinators in terms of resolution of a health problem are more information-averse, since they tend to delay the acquisition of more detailed information regarding their health condition:

Hypothesis 3.5. $I_{procrastinators} < I_{non-procrastinators}$,

where we indicate as non-procrastinators those participants who are willing to contact the doctor in case of arising of a problem after one hour to one week, whereas procrastinators delay the resolution of the problem beyond two weeks.

3.5 Results

In this section we provide an overview of the results through descriptive statistics and we follow the analysis with econometric estimates. Tables 3.4 and 3.6-3.7 summarize the effort dynamics and

[17] The value of 8 corresponds to the median health status reported by experimental subjects.

Table 3.4: Effort exerted between week 1 and week 3

	BT	AIT	5 balls	$3balls_{AIT}$	3 balls
Sum of efforts	$2.50^{,**}$	2.32	$2.71^{**,*}$	2.16	2.31
GS	$2.63^{*,*}$	2.24	2.38	2.19	2.40
BS	2.35	2.38	$2.92^{***,**}$	2.14	2.22
Complete effort	$0.77^{**,***}$	0.59	0.81^{**}	0.51	0.62
GS	$0.83^{**,**}$	0.56	0.63	0.54	0.68
BS	0.70	0.62	$0.92^{***,**}$	0.48	0.57

Notes: The Table reports means. We distinguish effort exerted after a good signal (GS) and after a bad signal (BS). 5 balls vs 3 balls refer to the number of balls displayed on the screen in week 1. Wilcoxon rank-sum tests have been estimated to test the equality of means for the sum of efforts, whereas tests of proportions have been estimated in case of complete effort. *** indicate significance at the 0.01 level, ** at the 0.05 and * at the 0.10 level, for comparisons of BT and 5 balls with the following column. Comparisons between BT and $3balls_{AIT}$ are indicated in the BT column, after the comma, if significant at any of the standard levels, whereas comparisons between 5 balls and 3 balls are indicated after the comma in the 5 balls column.

information acquisition, respectively. In Tables 3.5 and 3.8 we provide econometric estimates of the determinants of effort exertion and information acquisition. Descriptive statistics of participants' answers to the demographic questionnaire, the ex-post questionnaire and the risk elicitation task are reported in Appendix 3.A.6, whereas regression results with more detailed signals, following the distinction of Table 3.3 are included in Appendix 3.A.9.

3.5.1 Effort decisions

Table 3.4 depicts both the number of times participants completed the online task (sum of efforts) and whether they completed the task all three times (complete effort), which was a necessary condition in order to improve the composition of their own box.

Result 3.1. *Participants in the BT complete more effort than in the AIT, especially after receiving a good signal:* $e_{BT} > e_{AIT}$

Evidence in favor of Result 3.1 derives from data about effort completion, since the sum of efforts does not present significant differences across treatments: 77% of participants in the BT complete all three tasks, versus 59% in the AIT. The gap increases after a good signal, with 83% and 56% of participants completing all tasks, respectively. As in the AIT participants hold on average more information than in the BT, Result 3.1 indicates that overall a substitution effect is operating between effort completion and information acquisition, which is possible only in the AIT. In fact, by offering subjects the possibility to acquire more information they significantly exert less effort.

Considering Hypothesis 3.1, we find evidence of the first part of the inequality:

Result 3.2. *Effort is higher the higher is the accuracy of the signal, especially after a bad signal:* $e_{5,BS} > e_{3,BS}$

As the other comparisons included in Hypothesis 3.1 do not report significant tests, what seems to matter for the exertion of effort is the higher accuracy of bad signals.

The asymmetric impact of good and bad signals on effort exertion applies also when we focus on those participants of the AIT who observe five balls versus those who observe three balls:

Result 3.3. *A higher effort exertion occurs in the AIT when participants observe five balls instead of three, especially after a bad signal:* $e_5 > e_{3,AIT}$

Table 3.4 also shows that those participants in the AIT who decide to shy away from additional information (the $3balls_{AIT}$ column) significantly exert less effort than those who take part in the BT, especially after receiving a good signal:

Result 3.4. *A higher effort exertion occurs after the appearance of three balls in the BT compared to three balls appeared in the AIT, especially in case of a good signal:* $e_{BT} > e_{3,AIT}$

It is interesting to implement comparisons by rows in Table 3.4, thus verifying whether participants significantly exert more or less effort, conditional on the type of signal. As the marginal utility of effort is higher if one is provided with the bad box, it is rational to complete more effort in case of a bad signal. Therefore, conditional on the column, comparing effort exerted after the two types of signals allows us to shed some light on the condition in which participants behave more rationally. The only significant comparison between effort after a good and a bad signal turns out to be the case of five balls displayed, both for the sum of effort and for complete effort (Wilcoxon rank-sum test, z = 1.695, p = 0.0901; two-sample test of proportions, z = 1.6893, p = 0.0912, respectively). Therefore, after receiving additional information, participants take advantage of it and exert effort in the rational direction.

Result 3.5. *Participants' behavior is more rational, in terms of effort exertion, the more information they receive*

We perform regression analyses on the determinants of effort exertion, employing as dependent variables both the sum of efforts and the dummy variable taking value 1 if participants completed the three tasks. By comparing the two models we check whether there exist some factors influencing the completion of all the tasks versus factors driving participants to exert effort irrespective of the premium in expected payoffs. From Table 3.5 we note that the only significant difference between the two models lies in the coefficient of age: younger participants are more likely to complete all three tasks, whereas this does not occur in case of effort exertion. Estimates displayed in Table 3.5 provide a robustness check for the negative impact of the Additional Information Treatment on the exertion and completion of effort, as well as for the positive impact of the availability of more information, as depicted by the positive and significant coefficient of the Five balls variable. Interestingly, we find that females are less likely to exert and complete effort, whereas the contrary holds for smokers.

Regression results on the determinants of effort exertion employing the more detailed signal variable do not report significant differences in the estimates with respect to Table 3.5 and are reported in Appendix 3.A.9.

Table 3.5: Determinants of effort exertion and completion

Independent variable	Sum efforts	Complete effort
	(1)	(2)
Additional Information Treatment	-0.414**	-1.122***
	(0.196)	(0.338)
Age	-0.069	-0.203**
	(0.052)	(0.085)
Female	-0.605***	-0.994***
	(0.184)	(0.292)
Smoker	0.388*	0.552
	(0.225)	(0.364)
Low health	0.155	0.234
	(0.185)	(0.301)
Procrastinator	-0.198	-0.251
	(0.209)	(0.332)
Not a student as reference category		
Out of due date student	-0.286	-0.455
	(0.461)	(0.804)
Master student	-0.436	-0.854

Continued on next page

Table 3.5 – *Continued from previous page*

Independent variable	Sum efforts	Complete effort
	(0.401)	(0.730)
1st year student	-0.865*	-1.755**
	(0.467)	(0.869)
2nd year student	-0.417	-1.429*
	(0.448)	(0.833)
3rd year student	-0.282	-1.013
	(0.437)	(0.770)
Good signal	0.047	0.148
	(0.178)	(0.281)
Low risk aversion	0.199	0.372
	(0.184)	(0.310)
Five balls	0.481*	0.815**
	(0.246)	(0.409)
Constant	3.979***	5.942**
	(1.430)	(2.457)
Observations	116	116
Log likelihood	–	-58.306
R^2	0.211	0.212 (pseudo)
LR χ^2	–	31.52
Prob > χ^2	–	0.005
F	1.93	–
Prob > F	0.0315	–

Notes: The Table reports OLS and probit coefficients in models (1) and (2), respectively. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

3.5.2 Additional information decisions

In Table 3.6 we display information acquisition decisions following the strategy method procedure. In 31% of cases subjects decide to spend 1 Euro in order to have two more balls extracted from their box in week 1. When we distinguish information acquisition across types of signals we find that participants are more willing to buy additional information after receiving a bad signal, which is the contrary of the ostrich effect.

Result 3.6. *Participants are more willing to acquire information after receiving a bad signal:* $I_{BS} > I_{GS}$

Table 3.6: Information acquisition in the AIT

	Mean	Std. Dev.	Min	Max	Observations
Information acquisition	0.31	0.46	0	1	304
GS	0.22^{***}	0.42	0	1	152
BS	0.40	0.49	0	1	152

Notes: Observations refer to strategy method decisions. A test of proportion has been estimated comparing information acquisition after receiving a good signal (GS) and a bad signal (BS). *** indicate significance at the 0.01 level, ** at the 0.05 and * at the 0.10 level.

Table 3.7: Information acquisition in the AIT after more detailed signals

	Mean	Std. Dev.	Min	Max	Observations
Information acquisition	0.31	0.46	0	1	304
Excellent signal	0.22^{**}	0.42	0	1	76
Decent signal	0.22^{*}	0.42	0	1	76
Poor signal	0.35	0.48	0	1	76
Very poor signal	0.45	0.50	0	1	76

Notes: Observations refer to strategy method decisions. Tests of proportion have been estimated comparing information acquisition after receiving an excellent signal versus a very poor signal, and after receiving a decent signal versus a poor one. A test of proportions between poor and very poor signal did not turn out to be significant. *** indicate significance at the 0.01 level, ** at the 0.05 and * at the 0.10 level.

When good and bad signals are decomposed into more detailed signals, as in Table 3.7, we notice that after excellent and decent signals participants buy, overall, the exact same amount of information, hence suggesting that excellent and decent signals are perceived equally. Even if signals are symmetric in the negative domain with respect to positive signals, poor and very poor signals entail different levels of information acquisition. The means of information acquisition after a poor and a very poor signal are both statistically higher with respect to the information acquisition mean registered after a positive signal, and this strengthens the evidence in favor of higher information acquisition the worse is the signal received. As depicted in Table 3.7, this result follows a gradual pattern when moving towards worse types of signals.

This result is confirmed by the negative and highly significant coefficient of the dummy variable for the good signal in the regression model which investigates the determinants of information acquisition (Table 3.8). The model reports robust standard errors clustered at the individual level, since each participant takes repeated decisions about buying or not buying additional information. Thus, intra-subject correlation is taken into account. Risk aversion, age, gender, health-related variables and levels of study do not impact information acquisition.

Table 3.8: Determinants of information acquisition

Independent variable	Information acquisition
Good signal	-0.516***
	(0.174)
Low risk aversion	-0.151
	(0.232)
Age	-0.032
	(0.084)
Female	0.060
	(0.226)
Smoker	0.309
	(0.259)
Low health	0.092
	(0.255)
Procrastinator	-0.296
	(0.270)
Not a student as reference category	
Out of due date student	0.222
	(0.519)
Master student	-0.133
	(0.487)
1st year student	0.220
	(0.650)
2nd year student	0.170
	(0.520)
3rd year student	-0.223
	(0.550)
Constant	-0.086
	(2.160)
Observations	296
Log pseudolikelihood	-172.808
Pseudo R^2	0.0659
Wald χ^2	17.26
Prob > χ^2	0.1402

Notes: The Table reports probit coefficients with clustering at the individual level and robust standard errors in parenthesis. The dependent variable is Information Acquisition. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

Finally, no evidence of correlation between information aversion and health-related variables emerges from our data, neither from pairwise statistical tests performed on the sub-samples of those who bought additional information and those who did not, nor from the regression analysis of Table 3.8.

When the variable depicting more detailed signals is included in the regression analysis of information acquisition, we find that participants significantly buy more information when receiving a poor signal with respect to an excellent signal, and the coefficient of a very poor signal is even higher and more significant. This finding corroborates the higher information acquisition in case of bad signals, and allows us to shed more light on the gradual effects of more detailed signals on the patterns of information acquisition. Regression results are reported in Table 3.14 of Appendix 3.A.9.

3.6 Discussion and conclusion

Information aversion is frequently observed in real life contexts. These range for instance from financial to health scenarios, in which individuals prefer not to obtain additional information when they fear to receive conclusive bad news, whereas they search for more information when they believe conclusive information will be good. In these contexts individuals do not follow the standard paradigm of choice, since beliefs seem to matter for subjects' utility function and thus anticipatory utility should be taken into account when considering individual preferences. To the extent that individuals may manipulate sources of available information and obtain sub-optimal outcomes, it is important to investigate the determinants of information acquisition.

Moreover, little is known about the interaction between information acquisition/aversion and the possibility of exerting an effort which increases the final expected payoff, where the effort could be considered as a prevention type of activity. Introducing the possibility of exerting effort in an information aversion framework allows us to tackle the important informational trade-off which characterizes the ostrich effect: individuals may fear bad news and suffer from emotional costs, but additional information may be relevant for pushing them to exert effort, which in our design increases the final expected payoff in a higher proportion for those in a bad condition than for those in a good condition. Our experimental design enables us to study the interrelation between information aversion and the possibility to improve own condition, as well as to study how the accuracy of the preliminary signal about own condition impacts the exertion of effort, conditional on the type of signal received.

Our primary finding is that participants significantly buy more information after a preliminary bad signal: in this case, additional information is acquired in 40% of cases, whereas the percentage of information acquisition after a good signal is 22% (Result 3.6). Moreover, by decomposing good and bad signal into finer categories, we find that the effect of higher information acquisition after bad signals is gradual when moving from excellent to very poor signals. Thus, contrary to Karlsson et al. (2009), who do not envisage possibilities of improving own condition, we find an opposite result. One explanation for this may derive from the willingness to have better information about their own box given the higher return in the exertion of effort in case one holds a bad box. If information acquisition were driven by having a better clue about the expected return of effort in case of a bad signal, we would report significantly more participants completing the tasks after a bad signal than after a good signal. However, evidence on effort decisions does not lie in this direction: overall, effort exerted after a bad signal is not statistically different from effort exerted after a good signal. Moreover, there is no general evidence of effort differentials conditional on the quality of the signal, even when we distinguish them on the accuracy of the signal received.

As highlighted in the literature review section, previous studies revealing the existence of an ostrich effect, such as Karlsson et al. (2009), do not envisage the possibility of improving subjects conditions. This, on the contrary, represents the main innovative component of our experiment, through the completion of effort in a real-life situation. It could hence be the case that by providing experimental subjects with the possibility of improving their condition, as represented by expected payoffs in the experimental design, they revert a phenomenon that has been proven to exist in different contexts. In order to disentangle the role played by effort in our experiment, it could be advisable to run a third treatment, which mirrors the already existing Additional Information Treatment, in which we elicit subjects' intentions to buy additional information. Differently from the AIT, the new treatment would not entail the opportunity of exerting effort between week 1 and week 3. By comparing information acquisition between this new treatment and the AIT, it would be possible to shed more light on the main innovative element of this experiment.

Accuracy of signals has an impact on the exertion of effort only when the signal is bad: in this case participants exert significantly more effort than after a good signal, which lends support to the importance of additional information acquisition if someone deems to be in a bad situation (Result 3.2). This is confirmed by the higher effort exertion and completion after a bad signal and a more complete set of information when we focus on the AIT only. In this case, participants faced the same possibility of acquiring additional information, and those who had more information displayed and received a bad signal significantly exerted more effort (Result 3.3). Hence, participants understand the incentives of completing effort when they receive a bad signal about their condition only when the informational level is high, or, in other words, the more informed they decided to be.

Effort exertion and information acquisition seem to be substitutable rather than complementary: providing participants with the possibility of acquiring additional information lowers their effort exertion. A possible explanation for this finding may derive from the decomposition of AIT participants. We can note a duality in the behavior of AIT participants: those who chose to buy additional information and observed five balls from their box are rational types of individuals, as on average they exert more effort and in particular they exert it after a bad signal (Result 3.5). On the contrary, those individuals who decide not to have more information displayed remain irrational in their choices of effort exertion and completion. Thus, they don't significantly exert more effort after a bad signal, even if this would represent the ideal decision to undertake. These irrational types of subjects significantly exert less effort also compared to BT participants, who have the same informational set (Result 3.4).

It is reasonable to expect that these two kinds of subjects coexist in the BT, but since we do not leave them the possibility of acquiring information we cannot distinguish them. In the BT the rational subjects face more uncertainty about their own condition, as they receive less information from their box than in the case of the AIT with five balls. This could lead them to exert even more effort than they exert if they have the possibility of observing five balls. On the other hand, as the irrational subjects in the BT are not provided with the possibility of acquiring more information, they may be driven to exert more effort. Nevertheless, since in the BT we do not find more effort exerted after a preliminary bad signal, as it would be rational, we can posit that irrational subjects outnumber rational ones in the BT.

Our results suggest that the possibility of acquiring additional information about own condition points out a duality of individuals: those who decide to be more informed and thus follow a rational behavior in the exertion of effort, and those who decide to remain with a limited information set and behave irrationally in the exertion of effort. Further analysis could explore whether reducing the cost of information acquisition or making the returns from effort more understandable may foster information diffusion and thus effort exertion by the irrational types of individuals, who may undertake a more rational behavior in terms of effort exertion.

We do not observe any correlation between information aversion and risk aversion. However, what could be more relevant in an ostrich effect framework is ambiguity aversion, which could be measured in further experimental sessions by the Abdellaoui et al. (2011)'s task.

Further research could also be conducted in deepening our understanding of participants' beliefs of being attached to the good or the bad box, and these could be compared with the Bayesian beliefs of Table 3.2. Beliefs could be elicited and incentivized after receiving the preliminary signal and in case participants opt for the extraction of two additional balls from their box, in order to check participants' belief-updating.

3.A Appendix

3.A.1 Instructions - Session 1

Welcome![18]

You are about to participate in an experiment on decision making. Please switch off your mobile phone and remain quiet. It is strictly forbidden to talk to other participants. If you violate this rule, you will be dismissed from the lab and forfeit all earnings. Whenever you have a question, please raise your hand and one of the experimenters will come to your aid and answer your questions privately. If the question is deemed to be relevant for everybody, the experimenter will repeat the answer aloud for all participants.

All instructions are identical for all participants, and we read them aloud so that you can verify them. At the beginning of each session you will receive the corresponding instructions.

The experiment consists of two sessions, run in two different days: the first one is the one at which you are taking part today, while the second session will take place in exactly two weeks from now. Participation to the experiment will be considered completed only after session 2. Therefore, payments will occur only after the second session. In case of unforeseen circumstances, you can postpone your participation to the second session of a maximum of 2 days. After two days of delay we can no longer consider you as participant to this experiment and hence you will not receive any payment. So, it is very important that you commit to come back for the second session. All your decisions and answers will remain anonymous throughout the experiment. During this experiment, your earnings will depend on your decisions. It is therefore important that you read the instructions carefully.

Instructions of session 1

Each of you, at the beginning of the experiment, will receive a secret code. This code is crucial throughout the experiment because it allows us to keep your choices anonymous and it allows you to identify yourself at the beginning of each task you may complete.

We will provide you two copies of the secret code. You will take home one, which you will have to bring back to the laboratory for session 2, in order to continue the experiment. At the end of session 1, we ask you to put the second copy of your secret code in an envelope that we provide you, write

[18] These instructions refer to the Baseline Treatment (BT). In Italics, prior explicit indication, are reported the parts exclusive for the BT or added for the Additional Information Treatment (AIT).

your name on it and close it with adhesive tape. We will keep all envelopes in a safe place at the laboratory, and they will be useful in case someone loses her secret code or forgets to carry it with her at session 2. This procedure will allow you to know that the code is anonymous and that we cannot identify your decisions and answers throughout the experiment.

Each of you is matched to a box containing 20 balls. Half of you is matched to Box A and half of you to Box B. The boxes differ in terms of the composition of white and black balls inside them. Box A contains 14 white balls and 6 black balls, while Box B contains 6 white balls and 14 black balls. The color of balls is associated with a different payoff. In experimental session 2, the computer will randomly extract for each participant a ball from her respective box. If the ball extracted is white, your payoff is equal to 25 Euros. If the ball extracted from the box is black, your payoff is equal to 15 Euros. You do not know to which box you have been matched.

Today, a random extraction of 3 balls, without reinsertion, for each participant's box takes place. Without reinsertion means that each extracted ball is not reinserted inside the box. As an illustrative example, if you have been matched to box A, whose content is 14 white balls and 6 black balls, if the first extraction results in one white ball, this means that for the second extraction there are 13 white balls and 6 black balls available in the box.

There are 4 possible combinations for the 3 balls extracted from your box:

- 3 white
- 2 white and 0 black
- 1 white and 2 black
- 3 black

[BT only] The computer will randomly extract 3 balls from your box and the color of these balls will be displayed on the screen.

[AIT only] For each possible combination of 3 balls extractable from your box, we ask you to express whether you are willing to pay for the extraction of 2 additional balls from your box, without reinsertion. In case you opt in favor of the extraction of two additional balls, this procedure costs you 1 Euro, which will be subtracted from your final earnings at the end of session 2. This means that if you pay 1 Euro, you can have 5 balls extracted from your box, instead of 3. After you have indicated all your hypothetical choices for each different combination of balls extracted, the computer randomly extracts 3 balls from your box and the color(s) of these balls is displayed on your screen.

Then, the color of 2 additional balls randomly extracted from your box will be displayed only if you previously decided to pay 1 Euro for this extraction.

Time between session 1 and session 2

Between session 1 and session 2, you have the possibility of completing a task online that may change your final earnings.

You can complete this task only during some specific days, indicated below:

- this Thursday
- next Tuesday
- next Thursday

You will receive an email on these days, reminding you of this possibility. The email will also contain the link at which you will find the online task.

You can log in on the internet page devoted to the task at any time between 00:00 and 23:59, in these specific days. When you log in, you will have to indicate your secret code. You will receive emails after you will complete the task. If you choose to perform the task, you will have to complete the task straight after you log in on the web site.

The impact of completing the task, **in all these dates**, on your earnings depends on whether you have been matched to box A or box B:

- if you have box A and you complete all the tasks, two black balls will be transformed into two white balls. This means that, since box A contains initially 14 white balls and 6 black balls, you will end up having 16 white balls and 4 black balls

- if you have box B and you complete all the tasks, four black balls will be transformed into four white balls. This means that since box B contains initially 6 white balls and 14 black balls, you will end up having 10 white balls and 10 black balls

If you do not want to complete the task at all or if you do not complete the task in all the specific dates, your box will keep the same composition as initially, hence either with the original Box A or the original Box B. In other words, the composition of your box will change only if you connect on the web site of the experiment at the three specified dates and if you complete the task in its entirety.

The online task consists in replicating a blurred and not well-defined code. You are asked to write a series of such codes. Each online session lasts approximately 5 minutes.

Session 2

After two weeks, at the same day and time of this experimental session, you will have to come back to the laboratory in order to finalize the experiment and proceed for payment. We will send you an email to remind you of this task.

Do not forget to bring your secret code with you.

We will ask you to answer a few questions and to take decisions that reflect your preferences.

After this, you will be notified the type of box you were matched with, and the computer will activate the random extraction of one ball from your box. For the purposes of this extraction, your box contains 20 balls. Indeed, any ball previously extracted will be reinserted in your box. The composition of your box may have changed from the initial composition exclusively if you have completed the three tasks successfully between session 1 and session 2.

Your screen will display the color of the ball extracted at random from your box. You will be paid according to this draw (25 Euros if the ball is white, 15 Euros if the ball is black).

You will receive a feedback on the exact composition of your box only at the end of session 2.

Please answer the following questions. These serve to check your understanding of the decision situation and earnings calculations. When everyone will have answered all the questions correctly, we will proceed with session 1.

Control questions:

[AIT only] 1) In Session 1, if you pay 1 Euro for the extraction of two additional balls from your box, this extraction takes place:

- from 20 balls

- from 17 balls

- from 3 balls

2) Suppose that you have completed the task in all the three dates specified in these instructions. Suppose that you think you have been initially matched to box B. What is the effect of completing the task on your hypothetical situation?

- your box now contains 14 white balls and 6 black balls

- your box now contains 10 white balls and 10 black balls

- your box remains unchanged

3) The experiment will end in two weeks of time and you will be paid even if you don't show up at session 2, in two weeks of time

- true
- false

4) If you have been matched to box A, which contains a higher number of white balls, you will earn 25 Euros for sure

- true
- false: this will occur only if, at the end of session 2, the final extraction will result in one ball extracted and if I haven't bought the extraction of two additional balls

5) At the end of session 2, the random extraction of 1 ball from your box will occur from a total of

- 17 balls
- 20 balls
- 15 balls if you paid 1 Euro for the extraction of 2 balls

3.A.2 Demographic questionnaire

1) Which is your gender?

2) In which year were you born?

3) In which faculty are you enrolled?
(Economics/ Humanities/ Engineering/ Law/ Physics and Mathematics/ Sociology or Psychology/ Other Natural Sciences/ Other Social Sciences/ not a student)

4) In which year of studies are you enrolled?
(first year/ second/ third/ out-of-due date/ master student/ not a student)

3.A.3 Online task

On three dates between session 1 and 2, participants received an email with a link to a web page in which they could find the online task they could perform.
Here we report a screenshot of the page with the task.

Figure 3.1: Online task between week 1 and week 3

3.A.4 Instructions - Session 2

Welcome to the second session of the experiment!

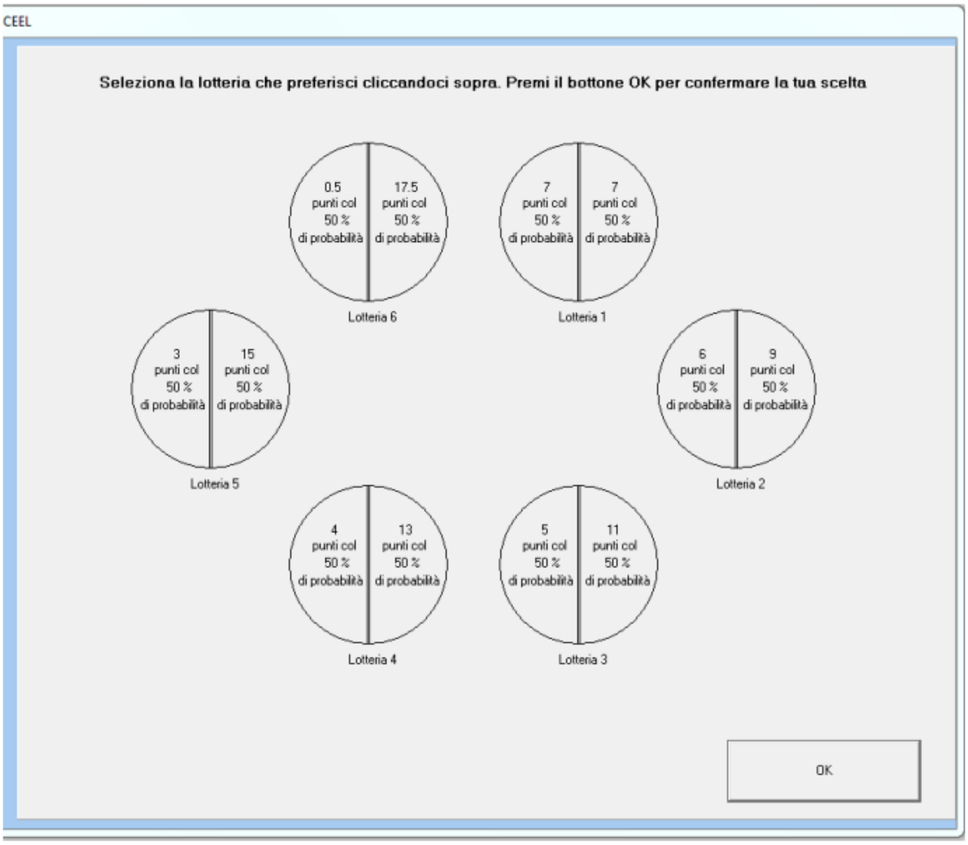
We will ask you to answer a few individual questions at the end of the experiment. As in session 1, all your choices will remain anonymous.

During this session your earnings are calculated in points. Each point corresponds to 0.20 Euros.

You will be paid individually and in a separate room.

We ask you to choose, among the six following lotteries, the lottery you would like to play. The figure below displays the type of screen you will face in order to select your preferred lottery.

Figure 3.2: Risk elicitation task



Each circle represents a different lottery. Each circle is divided in two parts. Each part represents a possible outcome of the lottery.

For each lottery, each outcome is equiprobable, i.e. it has a 50% probability to occur. The amount of points the lottery provides you for each possible outcome is indicated inside the circle.

At the end of this part, the computer will throw a six-sided die to determine which outcome of the selected lottery will take place:

- if the die throws a 1, 2 or 3, you will receive the points to the left of the circle
- if the die throws a 4, 5 or 6, you will receive the points to the right of the circle

Please keep in mind that in any lottery each outcome has 50% of probability to occur.

To select a lottery you should click on it. You can change your choice any time you want. When you are happy with your choice, press the OK button.

Illustrative example:

Suppose that you select lottery 3 and that, afterwards, the die throws a 4, 5 or 6 $\Rightarrow$ in this case you earn 11 points.

If the die throws a 1, 2 or 3 $\Rightarrow$ you earn 5 points.

3.A.5 Ex-post questionnaire

1) Are you currently a smoker?

(Yes / No)

1.1) If yes, are you trying to quit?

(Yes / No)

1.2) If not, have you ever been a smoker?

(Yes / No)

2) How would you consider your health status on a scale from 1 to 10, where 1 represents the minimum and 10 the maximum?

(1/ 2/ 3/ 4/ 5/ 6/ 7/ 8/ 9/ 10)

3) Suppose you have a tooth problem, and you think you should go to the dentist. Since the moment in which you think you should go to the dentist, how much time does it take before you call the dentist to fix an appointment?

(1 hour/ 1 day/ 2 days/ 1 week/ 2 weeks/ 1 month/ 2 months/ more than 2 months)

4) Have you completed the task in all three circumstances between session 1 and 2?

(Yes / No)

4.1) If yes, do you regret having performed it?

(Yes / No)

4.2) If not, do you regret not having performed it?

(Yes / No)

3.A.6 Background of experimental participants

Table 3.9 reports participants' descriptive statistics of the demographic questionnaire performed at the end of session 1, as outlined in Table 3.1.

In Table 3.10 we present participants' descriptive statistics of the ex-post questionnaire performed to subjects at the end of session 2, as also outlined in Table 3.1.

In Table 3.11 we report both frequencies and percentages of experimental participants' choices in the risk elicitation task presented in Figure 3.2.

Table 3.9: Descriptive statistics - Demographic questionnaire

	Frequency	Percentage
Gender		
Female	43	64.17
Male	77	35.83
Faculty		
Economics & Management	67	55.83
Engineering	12	10.00
Law	15	12.50
Literature	9	7.50
Math	1	0.83
Not a student	6	5
Sociology	10	8.33
Degree level		
Out of due date student	11	9.17
Master student	26	21.67
Not a student	7	5.83
1st year student	27	22.50
2nd year student	26	21.67
3rd year student	23	19.17
Age		
Mean	22.88	–
St. Dev.	2.60	–
Min	20	–
Max	32	–

Table 3.10: Descriptive statistics - Ex-post questionnaire

	Frequency	Percentage
Actual smoker		
Yes	23	19.66
No	94	80.34
Smoker trying to quit		
Not applicable	53	45.30
Yes	13	11.11
No	51	43.59
Ever been a smoker		
Not applicable	17	14.53
Yes	27	23.08
No	73	62.39
Health status		
4	1	0.85
5	3	2.56
6	4	3.42
7	24	20.51
8	43	36.75
9	32	27.35
10	10	8.55
Completed three tasks		
Yes	84	71.79
No	33	28.21
Regret having performed them		
Not applicable	18	15.38
Yes	9	7.69
No	90	76.92
Regret not having performed them		
Not applicable	68	58.12
Yes	18	15.38
No	31	26.50
Procrastination		
1 hour	15	12.82
1 day	29	24.79
2 days	24	20.51
1 week	21	17.95
2 weeks	17	14.53
1 month	8	6.84
2 months	2	1.71
More than 2 months	1	0.85

Notes: Health status ranges from a minimum of 1 to a maximum of 10.

Table 3.11: Descriptive statistics - Risk elicitation task

	Frequency	Percentage
Lottery 1	4	0.33
Lottery 2	22	18.33
Lottery 3	23	19.16
Lottery 4	17	14.16
Lottery 5	33	27.5
Lottery 6	18	15

3.A.7 Proof underlying Hypothesis 3.1

According to Hypothesis 3.1, if participants are rational the exertion of effort should comply with the following pattern: $e_{5,BS} > e_{3,BS} > e_{3,GS} > e_{5,GS}$.

Given that effort leads to a higher reward in case of a bad box with respect to a good box, as imposed in the experimental design and known by experimental subjects, the higher the probability of being attached to a bad box, the higher is the expected return from effort.

Therefore, indicating by P the probability of having a bad box, it follows that:

$P_{5,BS} > P_{3,BS} > P_{3,GS} > P_{5,GS} \Rightarrow e_{5,BS} > e_{3,BS} > e_{3,GS} > e_{5,GS}$.

In the following part of the section we aim to compute $P_{5,BS}$, $P_{3,BS}$, $P_{3,GS}$ and $P_{5,GS}$ in order to prove the first part of Equation 3.2.

Case of $P_{3,GS}$

<u>Bad box</u>

The bad box contains 6 white balls and 14 black balls. A good signal with an extraction of 3 balls can derive either from 3 white balls or from 2 white balls.

Probability that 3 white balls and 0 black balls are extracted from a bad box:

$p_{3-0,BB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{4}/_{18} = 0.018$,

where the first subscript refers to the number of white and black balls extracted, respectively, and the second subscript to a bad box.

Probability that 2 white balls and 1 black ball are extracted from a bad box:

$p_{2-1,BB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{14}/_{18} \times 3 = 0.184$

The probability of receiving a good signal, given a bad box, is hence equal to: 0.018 + 0.184 = 0.202

Probability that the good signal derives from 3 white balls extracted: ${}^{0.018}/_{0.202} = 0.087$

Probability that the good signal derives from 2 white balls and 1 black ball extracted: $^{0.184}/_{0.202} = 0.913$

Good box

The good box contains 14 white balls and 6 black balls. A good signal with an extraction of 3 balls can derive either from 3 white balls or from 2 white balls.

Probability that 3 white balls are extracted from a good box:

$p_{3-0,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{12}/_{18} = 0.319$,

where the first subscript refers to the number of white and black balls extracted, respectively, and the second subscript to a good box.

Probability that 2 white balls and 1 black ball are extracted from a good box:

$p_{2-1,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{6}/_{18} \times 3 = 0.479$

The probability of receiving a good signal, given a good box, is hence equal to: 0.319 + 0.479 = 0.798

Probability that the good signal derives from 3 white balls: $^{0.319}/_{0.798} = 0.4$

Probability that the good signal derives from 2 white balls and 1 black ball: $^{0.479}/_{0.798} = 0.6$

Since being attached to the good or the bad box is equally probable, the overall probability that the good signal derives from the extraction of 3 white balls, irrespective of being attached to the good or the bad box, is equal to: $^{1}/_{2} \times 0.087 + {}^{1}/_{2} \times 0.4 = 0.24$

With the same argument, the overall probability that the good signal derives from the extraction of 2 white balls and 1 black ball is equal to: $^{1}/_{2} \times 0.913 + {}^{1}/_{2} \times 0.6 = 0.76$

These are the weights of the Bayesian probabilities of being in the bad box, considering an extraction of 3 white balls and an extraction of 2 white and 1 black balls, respectively, which are indicated in Table 3.2:

$0.24 \times 5.21 + 0.76 \times 27.78 = 22.28$,

where 22.28 is the probability of being attached to the bad box, given 3 balls extracted and a good signal ($P_{3,GS}$).

Case of $P_{5,GS}$

Bad box

The bad box contains 6 white balls and 14 black balls. A good signal with an extraction of 5 balls can occur in three cases: from 5 white balls, from 4 white balls or from 3 white balls.

Probability that 5 white balls are extracted from a bad box:

$p_{5-0,BB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{4}/_{18} \times {}^{3}/_{17} \times {}^{2}/_{16} = 0.0004$

Probability that 4 white balls and 1 black ball are extracted from a bad box:

$p_{4-1,BB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{4}/_{18} \times {}^{3}/_{17} \times {}^{14}/_{16} \times 5 = 0.0135$

Probability that 3 white balls and 2 black balls are extracted from a bad box:

$p_{3-2,BB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{4}/_{18} \times {}^{14}/_{17} \times {}^{13}/_{16} \times 10 = 0.1174$

The probability of receiving a good signal, given a bad box, is hence equal to: 0.0004 + 0.0135 + 0.1174 = 0.1313

Probability that the good signal derives from 5 white balls: ${}^{0.0004}/_{0.1313} = 0.00295$

Probability that the good signal derives from 4 white balls and 1 black ball: ${}^{0.0135}/_{0.1313} = 0.1031$

Probability that the good signal derives from 3 white balls and 2 black balls: ${}^{0.1174}/_{0.1313} = 0.8939$

<u>Good box</u>

The good box contains 14 white balls and 6 black balls. A good signal with an extraction of 5 balls can occur in three cases: from 5 white balls, from 4 white balls or from 3 white balls.

Probability that 5 white balls are extracted from a good box:

$p_{5-0,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{12}/_{18} \times {}^{11}/_{17} \times {}^{10}/_{16} = 0.1291$

Probability that 4 white balls and 1 black ball are extracted from a good box:

$p_{4-1,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{12}/_{18} \times {}^{11}/_{17} \times {}^{6}/_{16} \times 5 = 0.3874$

Probability that 3 white balls and 2 black balls are extracted from a good box:

$p_{3-2,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{12}/_{18} \times {}^{6}/_{17} \times {}^{5}/_{16} \times 10 = 0.3522$

The probability of receiving a good signal, given a good box, is hence equal to: 0.1291 + 0.3874 + 0.3522 = 0.8687

Probability that the good signal derives from 5 white balls: ${}^{0.1291}/_{0.8687} = 0.1486$

Probability that the good signal derives from 4 white balls and 1 black ball: ${}^{0.3874}/_{0.8687} = 0.4459$

Probability that the good signal derives from 3 white balls and 2 black balls: ${}^{0.3522}/_{0.8687} = 0.4054$

Since being attached to the good or the bad box is equally probable, the overall probability that the good signal derives from the extraction of 5 white balls is equal to: ${}^{1}/_{2} \times 0.00295 + {}^{1}/_{2} \times 0.1486 =$

0.08

With the same argument, the overall probability that the good signal derives from the extraction of 4 white balls and 1 black ball is equal to: $1/2 \times 0.1031 + 1/2 \times 0.4459 = 0.27$

And the overall probability that the good signal derives from the extraction of 3 white balls and 2 black balls is equal to: $1/2 \times 0.8939 + 1/2 \times 0.4054 = 0.65$

These are the weights of the Bayesian probabilities of being in the bad box, considering an extraction of 5 white balls, an extraction of 4 white and 1 black balls, and 3 white and 2 black balls, respectively, which are indicated in Table 3.2:

$0.30 \times 0.08 + 3.38 \times 0.27 + 25 \times 0.65 = 17.19,$

where 17.19 is the probability of being attached to the bad box, given 5 balls extracted and a good signal ($P_{5,GS}$).

Case of $P_{3,BS}$

Bad box

The bad box contains 6 white balls and 14 black balls. A bad signal with an extraction of 3 balls can derive either from 1 white ball or from 0 white balls.

Probability that 1 white ball and 2 black balls are extracted from a bad box:

$p_{1-2,BB} = 6/20 \times 14/19 \times 13/18 \times 3 = 0.4789$

Probability that 0 white balls and 3 black balls are extracted from a bad box:

$p_{0-3,BB} = 14/20 \times 13/19 \times 12/18 = 0.3193$

The probability of receiving a bad signal, given a bad box, is hence equal to: 0.4789 + 0.3193 = 0.7982

Probability that the bad signal derives from 1 white ball and 2 black balls: $0.4789/0.7982 = 0.60003$

Probability that the bad signal derives from 2 white balls and 1 black ball: $0.3193/0.7982 = 0.40002$

Good box

The good box contains 14 white balls and 6 black balls. A bad signal with an extraction of 3 balls can derive either from 1 white ball or from 0 white balls.

Probability that 1 white ball and 2 black balls are extracted from a good box:

$p_{1-2,GB} = 14/20 \times 6/19 \times 5/18 \times 3 = 0.1842$

Probability that 0 white balls and 3 black balls are extracted from a good box:

$p_{0-3,GB} = {}^{6}\!/_{20} \times {}^{5}\!/_{19} \times {}^{4}\!/_{18} = 0.01754$

The probability of receiving a bad signal, given a good box, is hence equal to: 0.1842 + 0.01754 = 0.2017

Probability that the bad signal derives from 1 white ball and 2 black balls: ${}^{0.1842}\!/_{0.2017} = 0.9132$

Probability that the bad signal derives from 2 white balls and 1 black ball: ${}^{0.01754}\!/_{0.2017} = 0.08696$

Since being attached to the good or the bad box is equally probable, the overall probability that the bad signal derives from the extraction of 1 white ball and 2 black balls is equal to: ${}^{1}\!/_{2} \times 0.6 + {}^{1}\!/_{2} \times 0.91 = 0.755$

With the same argument, the overall probability that the bad signal derives from the extraction of 0 white balls and 3 black balls is equal to: ${}^{1}\!/_{2} \times 0.4 + {}^{1}\!/_{2} \times 0.08696 = 0.2434$

These are the weights of the Bayesian probabilities of being in the bad box, considering an extraction of 1 white ball and 2 black balls and an extraction of 3 black balls, respectively, which are indicated in Table 3.2:

$0.755 \times 72.22 + 0.2434 \times 94.79 = 77.60$,

where 77.60 is the probability of being attached to the bad box, given 3 balls extracted and a bad signal ($P_{3,BS}$).

Case of $P_{5,BS}$

Bad box

The bad box contains 6 white balls and 14 black balls. A bad signal with an extraction of 5 balls can occur in three cases: from 2 white balls, from 1 white ball and from 5 black balls.

Probability that 2 white balls and 3 black balls are extracted from a bad box:

$p_{2-3,BB} = {}^{6}\!/_{20} \times {}^{5}\!/_{19} \times {}^{4}\!/_{18} \times {}^{13}\!/_{17} \times {}^{12}\!/_{16} \times 10 = 0.3522$

Probability that 1 white ball and 4 black balls are extracted from a bad box:

$p_{1-4,BB} = {}^{6}\!/_{20} \times {}^{14}\!/_{19} \times {}^{13}\!/_{18} \times {}^{12}\!/_{17} \times {}^{11}\!/_{16} \times 5 = 0.3874$

Probability that 0 white balls and 5 black balls are extracted from a bad box:

$p_{0-5,BB} = {}^{14}\!/_{20} \times {}^{13}\!/_{19} \times {}^{12}\!/_{18} \times {}^{11}\!/_{17} \times {}^{10}\!/_{16} = 0.1291$

The probability of receiving a bad signal, given a bad box, is hence equal to: 0.3522 + 0.3874 +

0.1291 = 0.8687

Probability that the bad signal derives from 2 white balls and 3 black balls: ${}^{0.3522}/_{0.8687} = 0.4054$

Probability that the bad signal derives from 4 white balls and 1 black ball: ${}^{0.3874}/_{0.8687} = 0.4459$

Probability that the bad signal derives from 3 white balls and 2 black balls: ${}^{0.1291}/_{0.8687} = 0.1486$

Good box

The good box contains 14 white balls and 6 black balls. A bad signal with an extraction of 5 balls can occur in three cases: from 2 white balls, from 1 white ball and from 5 black balls.

Probability that 2 white balls and 3 black balls are extracted from a good box:

$p_{2-3,GB} = {}^{14}/_{20} \times {}^{13}/_{19} \times {}^{6}/_{18} \times {}^{5}/_{17} \times {}^{4}/_{16} \times 10 = 0.1174$

Probability that 1 white ball and 4 black balls are extracted from a good box:

$p_{1-4,GB} = {}^{14}/_{20} \times {}^{6}/_{19} \times {}^{5}/_{18} \times {}^{4}/_{17} \times {}^{3}/_{16} \times 5 = 0.01354$

Probability that 0 white balls and 5 black balls are extracted from a good box:

$p_{0-5,GB} = {}^{6}/_{20} \times {}^{5}/_{19} \times {}^{4}/_{18} \times {}^{3}/_{17} \times {}^{2}/_{16} = 0.0004$

The probability of receiving a bad signal, given a good box, is hence equal to: 0.1174 + 0.01354 + 0.0004 = 0.13134

Probability that the bad signal derives from 2 white balls and 3 black balls: ${}^{0.1174}/_{0.13134} = 0.8939$

Probability that the bad signal derives from 1 white ball and 4 black balls: ${}^{0.01354}/_{0.13134} = 0.1031$

Probability that the bad signal derives from 0 white balls and 5 black balls: ${}^{0.0004}/_{0.13134} = 0.00304$

Since being attached to the good or the bad box is equally probable, the overall probability that the bad signal derives from the extraction of 2 white balls and 3 black balls is equal to: ${}^{1}/_{2} \times 0.4054 + {}^{1}/_{2} \times 0.8939 = 0.6497$

With the same argument, the overall probability that the bad signal derives from the extraction of 1 white ball and 4 black balls is equal to: ${}^{1}/_{2} \times 0.4459 + {}^{1}/_{2} \times 0.1031 = 0.2745$

And the overall probability that the bad signal derives from the extraction of 0 white balls and 5 black balls is equal to: ${}^{1}/_{2} \times 0.1486 + {}^{1}/_{2} \times 0.00304 = 0.07582$

These are the weights of the Bayesian probabilities of being in the bad box, considering an extraction of 2 white balls and 3 black balls, an extraction of 1 white ball and 4 black balls, and 0 white balls and 5 black balls, respectively, which are indicated in Table 3.2:

$75 \times 0.6497 + 96.62 \times 0.2745 + 99.7 \times 0.07582 = 82.81,$

where 82.81 is the probability of being attached to the bad box, given 5 balls extracted and a bad signal ($P_{5,BS}$).

Since $82.81 > 77.60 > 22.28 > 17.19$, it is rational to exert effort following the pattern indicated in Hypothesis 3.1, for any effort cost function.

3.A.8 List of variables

In Table 3.12 we report a list of variables used in the non parametric tests of Tables 3.4, 3.6 and 3.7, as well as in the regression analyses reported in Tables 3.5, 3.8, 3.13 and 3.14.

Table 3.12: List of variables

Variable	Description
Sum of efforts	Times the participant completed the online task between week 1 and week 3 (it ranges from 0 to 3)
Complete effort	Dummy variable assuming value 1 if the participant completed the online task in all three circumstances and value 0 otherwise
Information acquisition	Dummy variable assuming value 1 if the participant spent 1 € in order to have two more balls extracted from her box in week 1 and value 0 otherwise
Additional Information Treatment	Dummy variable assuming value 1 if the participant belongs to the Additional Information Treatment and value 0 otherwise
Age	Age in years
Smoker	Dummy variable assuming value 1 if the participant answered "Yes" to Question 1) of the ex-post questionnaire and value 0 otherwise
Female	Dummy variable assuming value 1 if the participant is female and value 0 otherwise
Low health	Dummy variable assuming value 1 if the participant answered a value between 1 and 8 to Question 1) of the ex-post questionnaire and value 0 otherwise[19]
Procrastinator	Dummy variable assuming value 1 if the participant answered a value between "2 weeks" and "more than 2 weeks" to Question 3) of the ex-post questionnaire and value 0 otherwise
Good signal	Dummy variable assuming value 1 if the participant visualized minimum 2 white balls out of 3 balls or minimum 3 white balls out of 5 balls

Continued on next page

[19]The value 8 corresponds to the sample average of the variable.

Table 3.12 – *Continued from previous page*

Variable	Description
Dummies for more detailed signals	
Excellent signal	3 white balls out of 3 balls extracted OR 5 white balls out of 5 balls extracted OR 4 white balls out of 5 balls extracted
Decent signal	2 white balls out of 3 balls extracted OR 3 white balls out of 5 balls extracted
Poor signal	1 white ball out of 3 balls extracted OR 2 white balls out of 5 balls extracted
Very poor signal	0 white balls out of 3 balls extracted OR 0 white balls out of 5 balls extracted OR 1 white ball out of 5 balls extracted
Low risk aversion	Dummy variable assuming value 1 if the participant selected one lottery from 4 to 6 in the risk elicitation task represented in Figure 3.2 and value 0 if she selected one lottery from 1 to 3
Five balls	Dummy variable assuming value 1 if the participant visualized 5 balls and value 0 if she visualized 3 balls
Dummies for study levels	
Out of due date student	Student employing more years than those expected for the specific degree she is enrolled for
Master student	Student enrolled in a master program
1st year student	Student enrolled in the first year of the Bachelor
2nd year student	Student enrolled in the second year of the Bachelor
3rd year student	Student enrolled in the third year of the Bachelor
Not a student	Participant not enrolled in any University course

3.A.9 Regression analyses with more detailed signals

Tables 3.13 and 3.14 report the regression analyses of Tables 3.5 and 3.8, in which the dummy variable "Good signal" has been replaced with the variable indicating different types of signals, as detailed in Table 3.3.

Table 3.13: Determinants of effort exertion and completion with more detailed signals

Independent variable	Sum efforts	Complete effort
	(1)	(2)
Additional Information Treatment	-0.402**	-1.155***
	(0.199)	(0.344)
Age	-0.067	-0.207**
	(0.052)	(0.086)
Female	-0.643***	-1.006***
	(0.190)	(0.305)
Smoker	0.373	0.588
	(0.230)	(0.370)
Low health	0.172	0.195
	(0.190)	(0.310)
Procrastinator	-0.203	-0.289
	(0.214)	(0.339)
Not a student as reference category		
Out of due date student	-0.252	-0.592
	(0.488)	(0.852)
Master student	-0.407	-1.003
	(0.423)	(0.779)
1st year student	-0.835*	-1.926**
	(0.491)	(0.923)
2nd year student	-0.362	-1.583*
	(0.472)	(0.885)
3rd year student	-0.236	-1.092
	(0.447)	(0.791)
Excellent signal as reference category		
Decent signal	-0.200	-0.167
	(0.267)	(0.431)
Poor signal	-0.213	-0.150
	(0.262)	(0.412)
Very poor signal	-0.087	-0.372
	(0.286)	(0.456)
Low risk aversion	0.209	0.431
	(0.192)	(0.325)
Five balls	0.415	0.885**

Continued on next page

Table 3.13 – *Continued from previous page*

Independent variable	Sum efforts	Complete effort
	(0.265)	(0.442)
Constant	4.073***	6.377**
	(1.461)	(2.544)
Observations	116	116
Log likelihood	–	-58.107
R^2	0.217	0.216(pseudo)
LR χ^2	–	31.92
Prob > χ^2	–	0.010
F	1.72	–
Prob > F	0.0557	–

Notes: The Table reports OLS and probit coefficients in models (1) and (2), respectively. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

Table 3.14: Determinants of information acquisition with more detailed signals

Independent variable	Information acquisition
Excellent signal as reference category	
Decent signal	0.004
	(0.168)
Poor signal	0.371*
	(0.201)
Very poor signal	0.661***
	(0.222)
Low risk aversion	-0.151
	(0.233)
Age	-0.032
	(0.084)
Female	0.060
	(0.227)
Smoker	0.312
	(0.260)
Low health	0.091
	(0.256)
Procrastinator	-0.297
	(0.270)

Continued on next page

Table 3.14 – *Continued from previous page*

Independent variable	Information acquisition
Not a student as reference category	
Out of due date student	0.223
	(0.521)
Master student	-0.135
	(0.489)
1st year student	0.221
	(0.652)
2nd year student	0.169
	(0.522)
3rd year student	-0.223
	(0.551)
Constant	-0.611
	(2.176)
Observations	296
Log pseudolikelihood	-171.880
Pseudo R^2	0.0710
Wald χ^2	19.52
Prob > χ^2	0.1462

Notes: The Table reports probit coefficients with clustering at the individual level and robust standard errors in parenthesis. The dependent variable is Information Acquisition. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

4 Selection into Research: do gender and connections matter?[1]

4.1 Introduction

Over the past decades, researchers have sought to explain the great gender imbalance in research: women outnumber men by 18% when they finish university, but are under-represented compared to their male colleagues when we look at top academic positions. More specifically, their share represents 37% of Associate Professors and just 20% of Full Professors.[2] This phenomenon is commonly known as the "leaky pipeline" or "glass ceiling" effect, with both metaphors evoking the difficulty for women to reach equal access at the top of several professions (Clark Blickenstaff, 2005; Cotter et al., 2001).

What makes the university environment peculiar is that research contexts, differently from most jobs in the private sector, do not exhibit large disparities in work load across junior, intermediate and senior positions (Bosquet et al., 2014). Accordingly, child rearing - which, on average, affects women more than men - should not be a major reason for preventing women from reaching higher academic and research ranks.[3]

A recent report on the conditions in academia in Italy (Frattini and Rossi, 2013) shows a rather discouraging scenario for female academics. It is still significantly more difficult for women than for men to pursue a career in academia. The aim of this article is to contribute to the debate on gender discrimination in the Italian research community. This paper complements and extends the already existing investigation on the Italian research environment: we study the determinants of entrance to a non-academic research environment.

In 2005 the former president of Harvard University Larry Summers suggested that differences in

[1]This chapter is based on a joint work with Daniele Checchi and Nevena Kulic.
[2]European Commission data, see Meulders and O'Dorchai (2013).
[3]See Ceci and Williams (2011) for a meta-analysis of the causes of women's under-representation in science.

innate aptitude are responsible for the failure of women to progress to top scientific positions (Barres, 2006). This statement provoked a heated debate in the academic community. Although it is hardly defendable, it stimulated the debate on this important topic. For a country to be competitive in the globalized world, it is crucial to maximize its human intellectual capital, and data from around the world suggest that a great contribution could be provided by women (Sugimoto, 2013).

For more than a decade, in many disciplines, affirmative action policies that "require pro-active steps"[4] have been designed in order to level the playing field for men and women and help to increase the number of female academics (Sugimoto, 2013). A direct affirmative action in this direction, currently hotly debated, consists of imposing gender quotas in selection boards when filling research positions. This has already been introduced in a few European countries such as Sweden, Norway, Finland and Spain, where the compulsory share of commission members of each gender is 40%. Also the European Commission has recently brought into the discussion for suggested legislation such type of policy as a desirable tool to countervail the gender inequality which is still prevailing in the academic sector (Meulders and O'Dorchai, 2013, p. 7).

Nevertheless, imposing gender quotas in scientific evaluating commissions does not represent a zero-cost policy since it requires a disproportionate share of senior academic women's time for attending evaluating commissions, a "non-productive" activity from a research perspective. This policy could be detrimental to female academics research activities and could thereby foster the vicious cycle of women being trapped in lower academic positions. Therefore, effectiveness of such a policy should be carefully examined.

This paper aims to investigate the potential effectiveness of gender quotas in evaluating commissions. The focus of this study is a research environment in Italy, a country which is characterized by substantial gender inequality in many spheres of public and private life (De Paola and Scoppa, 2011; Frattini and Rossi, 2013). More specifically, we analyze the selection procedures within the Foundation Bruno Kessler (FBK), a top Italian research institution in hard science, located in the north-eastern part of the country. We aim at identifying the main drivers for winning a research competition, and in particular we are interested in knowing whether a higher proportion of women in evaluating commissions increases the proportion of women selected for research jobs.[5]

The contribution of this paper is threefold. First, in contrast to previous studies, we investigate

[4]For an overview of affirmative action policies, see Holzer and Neumark (2000, p.484).

[5]We thank the Human Resource division of FBK for providing the data within the project "FESTA", supported by the EC under the 7th FP.

gender discrimination in a non-academic research environment. The research institution we study is not related to any specific university, which allows us to shed light on research careers of women outside university. Second, we examine recruitment processes and not promotions. The recruitment strategy to the institution studied here mainly operates at the lowest research levels, equivalent to post-doc or research fellowship positions at universities. To the best of our knowledge, no other study has so far investigated gender discrimination at the start of a research career. This is an extremely interesting level of the research ladder, because it may be the starting point of the scissors effect in women and men research careers. Third, in almost all cases only one candidate is selected. Therefore, the competition among candidates is more evident and explicit than in previous qualification data sets, and it allows us to use different econometric techniques that take into account the interdependent success probability of candidates applying for the same call.

Building on these innovative components, our study suggests that the introduction of gender quotas may be useful in promoting female researchers in their early careers, although we do not find evidence of discrimination at higher hierarchical levels. Surprisingly, our analyses indicate that the most important factor in explaining the success of candidates are pre-existing ties with the institution or commission members, which however seem to be gender neutral. We also find interesting distinctions according to the quality of the commission in terms of research output, as well as a different productivity across candidates selected through prior ties or not.

The paper is organized as follows. In Section 4.2 we discuss the related literature and Section 4.3 presents the institutional environment in which FBK operates, and outlines the selection process followed. In Section 4.4, we describe the sources of our data set and we report candidate and call level characteristics. We proceed in Section 4.5 with the estimation strategy and results, where we focus on gender dynamics and pre-existing ties with the institution or one commissioner. Section 4.6 discusses and concludes.

4.2 Literature review

The economic literature has provided several explanations for the gender imbalance in academia and research. A number of studies attribute the poorer academic performance of women to their lower productivity. The lower productivity might be due to a lower number of international connections and coauthorships (Sugimoto, 2013), to the lack of role models in the upper echelons of the academic world (Blau et al., 2010), and to a higher burden of family responsibilities on women than men (Shen, 2013). A different reasoning is put forward by the supporters of the pipeline theory. According to this theory gender inequalities registered in academia are due to a shorter permanence of women in the academic pipeline. This theory suggests that women would reach gender parity in universities if they acquire the same tenure as their male colleagues (Gregory, 2003). However, evidence throughout the years does not seem to support the pipeline theory and hence other interpretations are needed.

A large recent body of literature based on behavioral and experimental studies posits the existence of a different preference system across genders with regard to risk aversion and competition: women are found to be more risk averse, less competitive, less self-confident and to have lower performance in competitions than men.[6] Since competition for top positions in academia is more intense, women could shy away from these jobs and prefer to remain at lower levels of the career ladder. If this is what women want and what they aim for, then, in economic terms no inefficiency would exist and no policy intervention would be required. Moreover, we cannot exclude the possibility that women are prevented from reaching the upper levels of the academic career because of gender discrimination. A possible channel through which discrimination might operate could be the larger presence of males in evaluating commissions. Male evaluators could be more likely to hire or promote a male researcher instead of a female if they are subject to gender stereotypes, if they have gendered connections to male candidates, or if they share the same research interests as male candidates. Therefore, examiners' gender-biased preferences could foster further the gender segregation of women in research.

A substantial number of studies have recently investigated the impact of the gender composition of evaluating commissions on the likelihood of female candidates to be selected for a research position. However, none of these has examined entry-level research positions, such as post-docs or research fellowships, which may be the starting levels of the gender imbalance in this type of profession.[7] These recent studies analyze centralized selection for Associate and Full Professor positions in Italy and Spain, where the promotion mechanisms are relatively similar and consist of a first phase of centralized selection, after which candidates apply for specific positions at the university level. The identification strategy hinges on the random assignment of evaluators to commissions (see De Paola

[6]For a detailed review of experimental studies see Gneezy and Rustichini (2004), Croson and Gneezy (2009), De Paola et al. (2013), Flory et al. (2012), Gneezy et al. (2009), Charness and Gneezy (2012), Niederle et al. (2013), Frick (2011), Kleinjans (2009), Datta Gupta et al. (2005), Datta Gupta et al. (2013), Villeval (2012) and Marianne (2011).

[7]Another strand of the gender discrimination literature has investigated non-academic competitions. Based on 150,000 applications to enter the Spanish Judiciary system, Bagues and Esteve-Volart (2010) find opposite sex preferences: commissions with relatively more females are more likely to hire a male candidate. Other studies examine the impact of evaluators' gender on decisions such as accepting articles in a leading journal in economics (Abrevaya and Hamermesh, 2012) or approving grant proposals for the economics Program of the National Science Foundation (Broder, 1993). In the former study no discrimination is documented against female authors, whereas in the latter there is a clear preference for opposite sex grant applicants.

and Scoppa (2011); Zinovyeva and Bagues (2011); Bagues et al. (2014)). This avoids endogeneity due to the possible existence of unobservables that may be correlated with commissions' and candidates' characteristics. The exogenous variation in the gender composition of the commission enables the estimation of the impact of one more female commissioner on the likelihood of a female candidate to be selected. Although the papers share the same methodology, the evidence is mixed. De Paola and Scoppa (2011) examine 1,000 candidates in Chemistry and Economics for the Italian qualifications to Associate and Full professorship held in 2008 and document same-sex preferences, while Bagues et al. (2014), analyzing 66,000 applications to Associate and Full Professorships in all academic fields in Italy in 2013, report the opposite result: namely, each additional female commissioner decreases the success rate of female candidates by 2%. A mixed and different result is provided by Zinovyeva and Bagues (2011) on Spanish data from all academic fields: opposite-sex preferences are found in competitions for Associate Professor, whereas female evaluators tend to prefer female candidates in competitions to become Full Professor. The authors explain their results with the internalization of the glass ceiling effect in academia by female evaluators, who may tend to discriminate against potential future female competitors. However, little is known on the gender dynamics operating at the start of a research career.

4.3 Institutional background

4.3.1 The research institute

This paper is based on a novel data set about selection procedures for research positions in the Bruno Kessler Foundation (FBK). FBK is a private non-profit research organization based in Trento (Italy), which focuses on the physical sciences but also carries out some research in the humanities. It was established in March 2007 as a transformation of the pre-existing Trentino Institute of Culture (ITC), a major provincial research institute founded in 1962 and fully funded by the local government (Provincia Autonoma di Trento).[8] Nowadays, it hosts around 350 researchers who work in five

[8] According to the provincial law 14/2005, art.28, regulating the transformation process from ITC to FBK, "Since the date indicated at subsection 2, the personnel working at ITC with open-ended contracts that does not establish a working contract with FBK, is transferred in the unique role of the provincial personnel, and is made available to FBK. The personnel transferred to the Province can ask to be hired at FBK within 120 days from the adoption of the governing body decision that identifies the collective contract indicated at subsection 13". Therefore, the data set investigated in this paper does not include calls dedicated to ITC personnel, since specific provincial laws regulated the transfer of personnel from ITC to FBK. More details of the laws regulating the transformation process are included in Appendix 4.A.3.

research centers.[9] Each center is composed of several independent units that manage various research projects and compete for national and international funds. The institute has an excellent reputation and attracts many young researchers not only from the local University of Trento, but also from other parts of Italy and from abroad.

4.3.2 The selection process

The selection procedure for hiring new researchers is competitive, public and follows the rules and stages outlined in the institutional guidelines. The guidelines cover two different selection processes: selection procedures for level I and II researchers, irrespective of whether contracts are fixed-term or permanent, and selection procedures for level III and IV researchers and employees with an equivalent salary. First and second level researchers correspond to the academic positions of Full and Associate Professor, whereas third and fourth level researchers are equivalent to the positions of Assistant Professor and Research Assistant, respectively.

For level I and II researchers, the analysis of the CVs by the evaluating commission results in a short-list of candidates. This is followed by seminar presentations by short-listed candidates. A ranking of eligible applicants is produced after these two stages.

The selection of researchers for level III and IV researchers and employees with an equivalent salary consists of four stages. In the first stage, the evaluating commission (or its president) carries out a first screening and provides a short-list with at most 20 candidates. These candidates are then interviewed by both personnel officers and researchers in charge of center(s) and/or research unit(s) that have required the hiring. This second phase determines the suitability of candidates. The list of suitable candidates is then ranked according to the quality of the candidates and adequacy for the position. Finally, the post is offered to highest ranked candidate(s).

Each evaluating commission in both cases consists of up to five members, with typically one from the Human Resource department. The remaining members are researchers from FBK (from the relevant unit or center that posts the vacancy) or from other research institutions (in Italy and abroad).

Within the sluggish Italian academic labor market, a salient feature of FBK is the high turnover of researchers, leading to a large number of selection procedures each year (between 30 and 50 per year). Given both the significantly smaller size of the two research centers working in the field of

[9]The five centers are: Center for Information Technology-CIT, which focuses on computer science; Center for Materials and Microsystems - CMM, concerned with microsystems and microelectronics, as well as computational physics and materials; European Center for Theoretical Physics - ECT; Center for Italian-German Historical Studies - ISIG, which deals with the historical connections between Italy and Germanic countries, and Center for Religious Sciences - ISR.

humanities (history and religious studies) and the unavailability of clear bibliometric criteria in this field, we focus on science and engineering competitions in the other three centers. These are typically male-dominated disciplines.

4.4 Data

The data used in this study come from three main sources: candidates' CVs, administrative archives of the FBK and the bibliometric database Scopus, which was used to retrieve information about candidates' research output. Candidates applied online or via email, and their applications are kept within the institution for up to 5 years. Administrative archives hold official final reports, as well as the job advertisement of the public call. The three sources lead to a multilevel design of the analysis, since data relate to both the applicant and the call level.

The data analyzed include information on 672 candidates from 112 calls posted between 2009 and 2011.[10][11] The number of applications and calls, each year, varied: 37 calls were posted in 2009, 47 in 2010 and 28 in 2011. Overall we examined 191 candidates in 2009, 290 candidates in 2010 and 191 candidates in 2011. Some of these competitions failed to fill the position, while a few resulted in recruitment of more than one candidate. For 27 competitions only one application was received. Our final sample excludes both competitions with only one applicant and without any successful candidate. As a result, the dataset contains 616 applicants and 79 calls: 26 calls and 179 candidates were analyzed for 2009, 31 calls and 266 for 2010, as well as 22 calls and 171 candidates were considered for 2011. Out of the 79 calls, 73 resulted in the recruitment of one candidate and they amount to 510 candidates, whereas five calls led to two recruitments and one to three, with 73 and 33 applicants, respectively.

The calls examined are all public calls, hence open to both external and internal candidates of FBK. The personnel already working at FBK, as employees and not as temporary collaborators, can proceed in the internal career ladder through the advancement procedures outlined in the institutional guidelines.[12] However, this does not hamper applications from the internal staff.

[10]The calls are observations at the call level and the candidates are observations at the candidate level.

[11]2008 was not included in the analysis because the transformation process from the Trentino Institute of Culture to FBK was still in progress in 2008.

[12]Among the positions included in the analysis, internal advancements in the career ladder do not apply only for co.co.pro. workers, who are considered as fixed-term collaborators.

4.4.1 Candidate level variables

We analyze socio-demographic characteristics of the candidates such as gender, age, the country of birth and, for Italians only, the region of birth, the highest educational level and its mark, the year of graduation, the years of work experience and the field of study.[13] Candidates' age was missing in 47% of CVs, but the proportion was higher for candidates from abroad at 66%. When no age was given, the age was estimated based on the year of graduation.

Several categorical variables were constructed. Origin was coded distinguishing between Trentino-Alto Adige (local), the rest of Italy, the European Union (including Switzerland) and the rest of the world. The field of study was recoded in three main areas: 1) social sciences and humanities (within engineering fields); 2) engineering, computer sciences, architecture, environmental sciences; 3) physical sciences without engineering (mathematics, physics, chemistry, geography). Educational attainment consists of four categories: no degree, bachelor degree, masters and Ph.D. level. We also generated a variable to capture the level of excellence of the candidate in her educational career, by distinguishing four ranges of the final mark: 110 and 110 *cum laude*[14]; between 100 and 109; between 90 and 99 and less than 90.[15]

The data set also includes information on candidates' scientific output. Candidates self-reported their publications on their CVs. Nevertheless, in order to make the comparison more objective, we decided to retrieve data regarding the number of publications and citations directly from Scopus, the Elsevier's bibliographic database founded in 2004.[16] In order to obtain applicants' publication records, the Scopus Author search page was queried with researchers' first and last names. If the author's name was not unique, the results were cross-checked with data appearing on candidate's CV, such as age, origin and field of study, in order to refine results and ensure the correct attribution of publications to candidates.[17] The database was accessed in August 2014. Both publications and h-index data were retrieved for the specific year of the FBK call for which the candidate applied. We

[13] Ph.D. activity alone is not considered as work experience, whereas lecturing and/or other types of employment, even if conducted during the Ph.D., are taken into account.

[14] This mark corresponds to distinction in the Italian university system.

[15] Foreign marks were converted to the Italian standard classification.

[16] Although other bibliographic sources such as Google Scholar and Web of Science are available, many studies suggest that Scopus is superior both in terms of coverage and accuracy. According to Falagas et al. (2008, p. 338), "Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy.". Moreover, "Scopus helps distinguish between the researchers in a more nuanced fashion than Web of Science."(Meho and Rogers, 2008, p. 1711).

[17] Out of 616 candidates, 50 candidates, corresponding to 40 unique individuals were not uniquely identified. In order not to lose these observations, the average value for the h-index and the number of publications of namesakes in the same field as the individual applying to FBK were entered in the data set.

opted for collecting candidates' h-index besides the number of publications because we believe that this combined measure is a good proxy for the importance and significance of candidates' contributions (Hirsch, 2005).

Finally, candidates' pre-existing ties with FBK are captured by a dummy variable taking value 1 if the candidate had prior ties with FBK or with one member of the evaluating commission, either in the past or at the time of the application. Defining a pre-existing tie is to some degree arbitrary. Candidates were classified as having a pre-existing tie/connection with FBK if they complied with at least one of the following criteria:
1) a co-author with a member of the evaluating commission;
2) supervision at the master or Ph.D. level by one of the commission members;
3) prior work experience (including internship) at FBK;
4) current work experience (including internship) at FBK.

We refer to 1) and 2) as commission ties and to 3) and 4) as institution ties, with the underlying hypothesis that ties with the institution may exert a stronger effect in the hiring procedures, since a candidate with prior or current work experience at FBK may have had more opportunities to show her skills and at the same time to create networks with the hiring institution.

4.4.2 Call level variables

For each call, the characteristics of the members of the evaluating commission are included in the data set, jointly with information about the position advertised. For commission members, gender, age and country of origin were obtained from administrative archives of FBK, and their bibliometric measures retrieved from Scopus for the year of the call. The latters provide a measure of the intrinsic quality of the commission, which according to the literature is positively correlated with the meritocracy of the hiring system.[18] To distinguish between researchers and HR staff, a dummy variable HR (Human Resources department) is created for each commission member.

In order to investigate whether the commissions are gendered, two variables were created: a dummy variable taking value 1 if the commission presented at least one female member, and another variable measuring the fraction of female members in the commission. At the call level, we also obtained information on the centers to which the call belongs, the duration of the contract, the

[18] De Paola et al. (2014), in a recent piece of work on the 2013 Italian qualifications to Associate and Full Professor, find that a commission whose members had an h-index above the median tends to weight a candidates' publications in the selection process more than other commissions; moreover, in the lower tail of the distribution of commissions' quality, candidates with a weaker publication record tend to be qualified, at the expense of stronger candidates.

salary and the type of position advertised.[19] The advertised positions are classified into the following categories: R1 are level I researchers, R3 level III researchers and R4 level IV researchers. Finally, our data set includes T4 positions, which are for level IV technicians, post-doc positions and fixed-term contracts (called "co.co.pro." in the Italian legislation).

4.4.3 Descriptive statistics

Almost 8 candidates applied per call.[20] The variation of the number of candidates across calls however is considerable: it ranged from two applicants to more than 30 applicants. Table 4.1 depicts the number of candidates and calls across the different positions included in the data set. We distinguish the different research positions into upper and lower positions. To the former type belong R1, R3 and R4 calls, which represent higher positions in the career ladder with respect to T4, post-doc and co.co.pro., which we refer to as lower positions. As can be noted in Table 4.1, the vast majority of the observations included in our data set apply for lower positions. Descriptive statistics regarding monthly wage, applicant's age and contract length across the different research positions are reported in Table 4.21 in Appendix 4.A.3.

Table 4.1: Number of candidates and calls across different positions

Position	N. applicants	N. calls	Average N. of candidates per call
R1	8	1	8
R3	65	13	5
R4	124	13	9.54
T4	22	2	11
post-doc	15	2	7.5
co.co.pro.	382	48	7.96
Total	616	79	7.8

Table 4.2 reports descriptive statistics for the candidate level variables. As can be seen from Table 4.2, the average candidate was around 30 years old, with 4.6 years of work experience; he had around 6 publications and a corresponding h-index of 1.38. The proportion of women was about 20%. Almost 14% of the candidates were recruited (15.7% of the female candidates and 13.5% of the male candidates) and almost 11% had pre-existing ties with FBK members or with commissioners (similar for men and women, with 12.3% of the women and 10.5% of the men).

Table 4.2 also allows us to shed a preliminary light on some marked differences between male

[19]The duration of the posted job is given in months of activity.

[20]Applicants per call range from 6.88 in 2009, to 8.58 in 2010 and 7.77 in 2011, with an overall mean of 7.80.

Table 4.2: Descriptive statistics - Candidates

Candidates					
	Mean	Std. Dev.	Min	Max	Observations
Female	0.196	0.397	0	1	616
Age	30.646	5.774	19	60	616
Females	29.942	4.736	20	53	121
Males	30.818	5.992	19	60	495
Success	0.139	0.346	0	1	616
Females	0.157	0.365	0	1	121
Males	0.135	0.342	0	1	495
Publications	5.982	21.659	0	424	616
Females	3.190	6.284	0	29	121
Males	6.664	23.917	0	424	495
H-index	1.388	3.188	0	39	616
Females	0.851	1.552	0	7	121
Males	1.519	3.461	0	39	495
Work experience	4.628	4.913	0	34	616
Females	3.609	3.723	0	25	121
Males	4.877	5.135	0	34	495
Ties with FBK	0.108	0.311	0	1	616
Females	0.123	0.330	0	1	121
Males	0.105	0.306	0	1	495

and female candidates: male applicants had on average one more year of work experience than female ones; they had on average three more publications, and their h-index was higher, although the disparity across genders was proportionally lower for the h-index.[21]

By decomposing the variable representing prior ties with FBK into the four categories indicated at the end of Section 4.4.1, we note that the majority of ties is represented by ties with the institution (Table 4.3). Forty candidates have been working in the past for FBK, whereas twenty-six are currently working for the institute. Only nine and eight observations have a co-author or a Ph.D./Master thesis supervisor among the commissioners, respectively.

It can be the case that one candidate has more than one type of tie with FBK. In order to investigate the role of a different amount of ties, we constructed a “tie indicator” variable, which ranges between 0 and 4, where 0 indicates that the candidate does not have any tie with FBK, and 4 stands for having all the four types of ties with FBK. Of the 616 observations of the data set, 549 have no ties, 55 have one tie, 9 have two types of ties, 2 have three types of ties whereas one individual boasts of all types

[21]The differences in the standardized number of publications between male and female candidates are statistically significant - Wilcoxon rank-sum test, z= 4.093, p= 0.000 -, whereas they are not significant for the standardized h-index - Wilcoxon rank-sum test, z= 1.531, p = 0.126 -.

of ties.

Table 4.3: Decomposition of types of ties with FBK

	Co-author	Supervisor	Prior work	Current work	At least one tie	Institution	Commission
Frequency	9	8	40	26	67	56	16
Percentage	1.46	1.30	6.49	4.22	10.88	9.09	2.60

Notes: The percentage is computed with respect to the total amount of candidates (616).

Looking at the geographical origin of candidates, almost half come from Italy, with one out of six candidates from the local Trentino-Alto Adige region (see Table 4.4). It is interesting that a large number of applications came from international researchers, and that the share of their applications increased over the years, especially for those candidates applying from outside Europe. This confirms the attractiveness of FBK for foreign researchers.

When we consider more in depth candidates' bibliometric measures, we notice large differences across fields of research, due to different publication propensities in different sectors of research. In order to compare candidates from different fields, we standardize the number of publications, as well as the value of candidates' h-index, according to 3 macro fields of research, more uniform in terms of bibliometric patterns: social sciences; math, physics and chemistry; environmental sciences, engineering and computer sciences.[22] Table 4.5 reports candidates' h-indexes across these macro fields of work, as well as the number of observations in each case.

Table 4.6 shows information about commission members. Of all commission members 26% were female, but if female HR members are excluded, the percentage of female commissioners decreases to 8.6%. Women accounted for the 74% of the total HR personnel at FBK.

Commission members were mainly Italian: only 30% were not Italian, and among female commissioners only 5% were non-Italian. The average age of commissioners was 45 years, with an h-index of 15. Commissioners' h-index was standardized following the same procedure as for candidates' h-index.[23]

If we look more closely at the gender composition of commissions, only 16% of the commissions included a female researcher. Thus, all-male researchers commissions were dominant, reached 92% in 2009 but the percentage was lower in the subsequent years, about 78-82%. Furthermore, the percentage of female researchers in commissions varies little, as generally only one member was female. Table 4.7 shows all combinations of female percentage of commissioners and commission's

[22]The field of research was indicated by each candidate on the CV sent to FBK at the time of the application.

[23]Unlike for candidates, who had to indicate their field of research in the CV, commissioners' field was retrieved from Scopus, as the principal sector of their publications appearing on the Elsevier database.

Table 4.4: Geographical origin of candidates

Origin	2009	2010	2011	2009-2011
Trentino- Alto Adige	14.53	21.05	10.53	16.23
Rest of Italy	45.81	25.56	26.90	31.82
EU and Switzerland	13.97	17.29	14.62	15.58
Rest of the world	25.70	36.09	47.95	36.36
Total	100	100	100	100

Table 4.5: Candidates' h-indexes across macro fields of research

Field	Mean	Std. Dev.	Min.	Max.	Obs.
Field 1	.116	.448	0	2	43
Field 2	2.507	4.057	0	39	227
Field 3	.812	2.440	0	32	346

Notes: Field 1 includes social sciences within engineering fields; Field 2 math, physics and chemistry; Field 3 environmental sciences, engineering and computer sciences.

size. The very high percentage of all-male commissions makes discrimination against female applicants a possibility if commissioners tend to have same-sex preferences as regards candidates. This will be further examined in the econometric analysis.

Table 4.8 depicts a series of descriptive statistics of successful and unsuccessful candidates, over the years under analysis. The percentage of successful candidates among male and female applicants is gender neutral in some years but varies in other years. The fraction of successful female candidates over the whole set of candidates is quite constant over the years at about 20%, but while the year 2009 presents the same proportion of female successful candidates, in 2010 women are favored and the opposite occurs in 2011.

Finally, Table 4.9 indicates that the disparity between men and women is greatest at high level research positions. No women applied for any level I position, while their share in applications and recruitment increases as the level of the position decreases. For instance, men and women are almost equally represented at the T4 level (which corresponds to the position of research assistants and technologists).

4.5 Empirical analysis

In this section we analyze whether women applying at FBK are discriminated during the selection process. We analyze the factors influencing the selection process, and among others the gender of

Table 4.6: Descriptive statistics - Commission members

Evaluators					
	Mean	Std. Dev.	Min	Max	Observations
Female commissioner	0.263	0.440	0	1	2058
Female commissioner (no HR)	0.086	0.280	0	1	1626
Origin	0.296	0.456	0	1	2058
Females	0.042	0.201	0	1	542
Males	0.387	0.487	0	1	1516
Age commissioner	45.267	11.385	25	71	1940
Females	40.094	9.606	27	58	542
Males	47.273	11.391	25	71	1398
Human Resources	0.210	0.407	0	1	2051
Females	0.741	0.438	0	1	542
Males	0.019	0.139	0	1	1509
H-index commissioner	15.071	11.698	0	47	1626
Females	4.821	4.445	0	19	140
Males	16.037	11.707	0	47	1486

the applicant and the gender composition of the commission. The descriptive overview showed that women are in the minority both at the application and the recruitment stages. Is it possible that one of the reasons for this is the under-representation of women in the evaluating commissions?
We first conduct an analysis on the whole sample of candidates at FBK, and we proceed to examine hiring patterns on sub-samples of applicants, obtained by stratifying on relevant characteristics of the hiring process. We then investigate the relationship between gender, origin and pre-existing ties.

4.5.1 Analysis of the full sample

To examine the effects of the gender composition of the commission on the probability of being hired we estimate Equation 4.1 by a conditional logit model:

$$Success_{ij} = \beta_0 + \beta_1 Female_{ij} + \beta_2 Female_{ij} * Female\ in\ commission_j + \beta_3 X_{ij} + \mu_j + \varepsilon_{ij} \qquad (4.1)$$

where $Success_{ij}$ is a dummy variable taking value 1 if candidate i was recruited as a result of the call j and value 0 otherwise. We control for the gender of the candidate, $Female_{ij}$, and for the interaction between the gender of the candidate and a dummy variable indicating the existence of at least

Table 4.7: Percentage of women in evaluating commissions and number of commissioners

	N. of commissioners					
% of women	1	2	3	4	5	Tot.
0	2	27	23	8	6	66
0.2	0	0	0	0	1	1
0.25	0	0	0	4	0	4
0.33	0	0	4	0	0	4
0.5	0	3	0	0	0	3
0.8	0	0	0	0	1	1
Tot.	2	30	27	12	8	79

Notes: Women in this table are considered as researchers and not belonging to the HR department of FBK.

Table 4.8: Applicants by final outcome and year

Year		Unsuccessful	Successful	Total
2009	Female share	0.22	0.19	0.22
	Age	30.86	29.32	30.60
	Publications	9.84	6.55	9.27
	H-index	1.70	1.61	1.69
	Work experience	4.49	3.81	4.37
2010	Female share	0.18	0.33	0.20
	Age	29.85	30.24	29.89
	Publications	4.50	9.12	5.08
	H-index	1.28	1.91	1.36
	Work experience	4.14	5.12	4.26
2011	Female share	0.19	0.09	0.18
	Age	32.18	29.73	31.87
	Publications	3.78	5.09	3.95
	H-index	1.05	1.59	1.12
	Work experience	5.71	3.80	5.46

one woman in the commission, $Female_{ij} *$ Female in commission$_j$.[24] Thus, β_1 indicates the effect of being female on the probability of being selected by a all-male evaluating commission, whereas β_1 + β_2 that of being female on the probability of being selected by a mixed-gender commission. X_{ij} is a vector of candidates' attributes, such as the Scopus h-index at the year of the competition, age and origin, a dummy for holding a Ph.D., years of work experience and the presence of pre-existing ties with FBK. Call fixed effects are captured by μ_j, which includes both the type of position and the area of specialization, as well as other factors that may influence candidates' probability of success.

Besides estimating Equation 4.1 by a conditional logit model, we also estimate a linear probability

[24]Given the low variability of the share of women among commissioners, we decided to opt for this specification in order to study the effect of the gender composition of the commission.

Table 4.9: Gender distribution of unsuccessful and successful applicants at different research levels

	Unsuccessful		**Successful**	
Position	Males	Females	Males	Females
R1	100.00	0.00	100.00	0.00
R3	80.39	19.61	64.29	35.71
R4	82.57	17.43	80.00	20.00
T4	52.63	47.37	66.67	33.33
post-doc	92.31	7.69	50.00	50.00
co.co.pro.	80.97	19.03	82.35	17.65
Total	80.75	19.25	77.91	22.09

model as a robustness check. The conditional logit allows us to take into account candidates' interdependent probabilities of being recruited. The likelihood of the data in a conditional logit depends on the conditional probabilities, conditional on the number of positive outcomes (in our case the success) within group. For almost all calls in our data set only one candidate was selected per call and hence the group at the basis of the conditional logit is composed of those candidates applying for the same call. In other words, the conditional logit fits a logistic model that explains why one candidate had a positive outcome in a certain group, conditional on one of the candidates in the group having a positive outcome. Hence, the differences across candidates are considered at the call level. Estimates for the conditional logit are displayed in Table 4.10 and refer to coefficients.

In the first specification reported in Table 4.10, column 1, we are interested in knowing whether there is overall a lower probability of female candidates to be recruited. Our results do not support this hypothesis. Examining the effect of the gender composition of the commission in column 2, it turns out that commissions with a female evaluator tend to favor female candidates. This holds true also if we include a measure of candidates' scientific productivity, the standardized h-index, as in column 3. Being more productive in bibliometric terms has a positive impact on the probability of being appointed for the job. Other individual variables statistically significantly influence the chances of success, as reported in column 4 of Table 4.10: age has a negative impact, whereas Italian origin and Ph.D. positively influence the probability of being recruited, as so does an additional year of work experience. Yet, by incorporating these variables the positive effect of the h-index is not statistically significant anymore.

If the estimation encompasses the dummy variable representing pre-existing ties, as in column 5, the interaction term between candidates' gender and the gender composition of the commission loses significance, while pre-existing ties is positive and highly significant. Therefore, it might be that

Table 4.10: Probability of success - Conditional logit

	1	2	3	4	5
Female	-0.033	-0.415	-0.371	-0.369	-0.404
	(0.309)	(0.379)	(0.383)	(0.406)	(0.432)
Female*Dummy fem. in Comm.		1.587**	1.545**	1.422*	1.101
		(0.760)	(0.763)	(0.798)	(0.817)
H-index standardized			0.221**	0.181	0.129
			(0.103)	(0.114)	(0.124)
Age squared				-0.002***	-0.002***
				(0.001)	(0.001)
Italian origins				1.052***	1.074***
				(0.339)	(0.352)
Ph.D.				0.790**	0.683*
				(0.384)	(0.413)
Work experience				0.090**	0.099**
				(0.045)	(0.047)
Ties with FBK					1.457***
					(0.351)
Pseudo R^2	0.000	0.016	0.030	0.104	0.164
Observations	614	614	614	614	614

Notes: The Table reports conditional logit coefficients, computed at the competition level. The dependent variable is a dummy variable for being successful. Standard errors are reported in parenthesis. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

the interaction term in the previous specifications of the model hides connections between female commissioners and female candidates. Moreover, all individual variables are significant, except for the productivity measure of candidates as captured by the h-index. This shows the importance of prior ties, possibly at the expense of purely meritocratic patterns of selection, based on research output. Results of the final conditional logit model are confirmed both in significance and in signs when we look at the estimates of Table 4.11. Linear probability models do not change substantially the results, thus providing a robustness check. It is important to note that in both final specifications the probability of selection is not statistically significantly different for female and male candidates and the presence of a female member in the evaluating commission is also not statistically significant. This means that a female quota in evaluating commissions seems unlikely to have an impact, if we consider the whole data set.

4.5.2 Analysis on sub-samples of candidates

In this section we analyze if a gender gap in hiring patterns arises in sub-samples obtained by splitting the original data set according to relevant variables. All results displayed in Table 4.12 are from a

Table 4.11: Probability of success - Linear Probability Model

	1	2	3	4	5
Female	-0.004	-0.054	-0.046	-0.045	-0.036
	(0.043)	(0.048)	(0.048)	(0.047)	(0.046)
Female*(Dummy fem. in Com.)		0.205**	0.199**	0.173*	0.139
		(0.090)	(0.089)	(0.090)	(0.094)
H-index standardized			0.032	0.021	0.014
			(0.019)	(0.018)	(0.018)
Age squared				-0.002***	-0.001***
				(0.000)	(0.000)
Italian origins				0.115***	0.112**
				(0.042)	(0.043)
Ph.D.				0.114**	0.091**
				(0.044)	(0.044)
Work experience				0.009**	0.011**
				(0.005)	(0.005)
Ties with FBK					0.283***
					(0.082)
Call FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.005	0.013	0.019	0.051	0.109
Observations	616	616	616	616	616

Notes: The Table reports OLS estimates. The dependent variable is a dummy variable for winning the competition. We control for call fixed effects. Standard errors, clustered at the competition level, are reported in parenthesis. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels respectively.

conditional logit model. The model is the same as the full model specification of Table 4.10, column 5. First, we are interested in distinguishing commissions in terms of a bibliometric measure, as a proxy for the quality of the commission. To this end, we compute the average h-index of all evaluating commissions and divide the full sample of candidates in two sub-samples: those assessed by commissions belonging to the fourth quartile of the h-index distribution and those in the three lower quartiles. Estimations performed on these two sub-samples are shown in columns 1 and 2 of Table 4.12.

The fourth quartile of commissioners in terms of research quality seems to be more responsive to candidates' research production, as highlighted by the positive and significant (at the 10% significance level) coefficient for the standardized h-index. Also Italian origin increases the probability of recruitment and so do more years of work experience, *ceteris paribus*. For candidates evaluated by the three lower quartiles in terms of h-index, the criteria that matter most for being recruited are being younger, having Italian origin and pre-existing ties with FBK. The same estimation has been performed on two sub-samples divided in above or below the median of commission's h-index. No

Table 4.12: Probability of success sub-categories - Conditional logit

	1	2	3	4	5	6
	H-index commission:		**Position:**		**Discipline:**	
	4^{th} quartile	1^{st}- 3^{rd} quartiles	lower	upper	lower h	higher h
Female	-0.059	-0.359	-1.425**	0.637	-0.207	0.468
	(1.221)	(0.465)	(0.699)	(0.610)	(0.548)	(1.006)
Female*	15.444	0.814	2.000*	0.720	-0.504	15.592
*(fem. in Com.)	(1508.159)	(0.863)	(1.062)	(1.570)	(1.255)	(2216.181)
H-index stand.	0.321*	0.020	0.499***	-0.098	0.029	0.885***
	(0.183)	(0.257)	(0.186)	(0.239)	(0.153)	(0.332)
Age squared	-0.002	-0.002**	-0.002**	-0.001	-0.001	-0.009***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Italian origins	1.947**	0.890**	0.931**	1.578**	1.166***	1.555**
	(0.850)	(0.392)	(0.430)	(0.680)	(0.436)	(0.786)
Ph.D.	0.318	0.737	0.679	0.647	0.469	0.889
	(1.284)	(0.452)	(0.586)	(0.618)	(0.533)	(1.023)
Work experience	0.308**	0.081	0.141***	0.014	0.077	0.273*
	(0.153)	(0.052)	(0.055)	(0.102)	(0.056)	(0.148)
Ties with FBK	-0.447	1.590***	1.645***	1.790***	1.710***	2.806**
	(1.706)	(0.370)	(0.499)	(0.552)	(0.430)	(1.160)
Pseudo R^2	0.238	0.176	0.200	0.239	0.166	0.410
Observations	168	446	417	197	340	190

Notes: The Table reports conditional logit coefficients, computed at the competition level. The dependent variable is a dummy variable for for being successful. Standard errors are reported in parenthesis. In models 1 and 2 observations are divided according to the average h-index of commissioners, whether belonging to the first three quartiles of the h-index of the commissions or to the fourth quartile; in models 3 and 4 the distinction is made according to the level of the position of the competition: lower refers to co.co.pro., post-doc and T4 levels, whereas upper stands for R1, R3 and R4 types of positions; in models 5 and 6 observations are divided in two sub-samples according to the average h-index of the discipline involved by the competition: model 5 includes fields 1 and 3, whereas model 6 refers to field 2. Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

relevant differences between the two sub-samples are found in this case, meaning that a different hiring pattern emerges only for top quality evaluating commissions.

Next, we examine whether different characteristics are statistically significant if we consider separately lower and upper research positions at FBK.[25] As we pointed out in the introduction, what distinguishes our study from previous studies is the availability of data for entry research jobs: indeed, 417 out of 616 applications are for lower level research jobs, whereas only 197 are for upper research positions (columns 3 and 4 of Table 4.12). Females applying for lower level jobs have, *ceteris paribus*, a lower probability of being hired, although this is reverted if the commission includes a

[25] Due to the scarce numerosity of candidates applying for level I and level II positions, we split the sample in two groups with respect to the hierarchy of research positions present in the sample. To this end, T4 candidates, although belonging to the same level as R4 positions, were included in the lower group, given the less research-oriented nature of their contracts.

female researcher. Being younger, having more years of work experience and being born in Italy positively influence the probability of selection, and so does having a higher h-index. Still, pre-existing ties with FBK play an important role in being recruited. If we look at the recruitment for upper levels of research positions, important determinants for being recruited are being born in Italy and having prior ties. However, not only do we lose statistical significance of the gender dummy, but we also lose significance of the h-index, age, and work experience, but not of pre-existing ties with FBK. The recent Italian reform on Italian universities, which strengthens the role of bibliometric measures in selection processes for academic positions, seems to impact mainly entry level recruitment .[26]

Finally, we divide candidates according to the h-index of the field of research. For this purpose we take the fields of research of Table 4.5. We group fields 1 and 3, which represent social sciences, engineering, computer sciences and environmental sciences, which are characterized by a lower h-index, separately from field 2, which corresponds to math, physics and chemistry. Looking at columns 5 and 6 of Table 4.12 we note that in both sub-samples being selected depends on being born in Italy and having pre-existing ties with the institution. For field 2, age and the h-index are statistically significant, which is not the case for the other fields. In both sub-samples, no indication of discrimination against female candidates and no effect of a female researcher in the commission can be found.

4.5.3 Gender, origin and prior ties

In the previous sections it emerged that in almost any model specification and sub-sample, the most important factors affecting the probability of recruitment are Italian origin and pre-existing ties with FBK. In half of the sub-samples, the probability of being recruited decreases with age and increases with work experience. Interestingly, in half of sub-samples of Table 4.12, scientific productivity is not statistically significant, while holding a Ph.D. does not statistically increase the chances of selection.

We find clear indications that pre-existing ties with FBK increase the probability of recruitment. This represents an alternative channel through which a gendered selection may take place. Therefore, we explore in more detail whether pre-existing ties are gender specific. Prior ties precede the actual selection process. Could they be a mechanism to attract more women into research? The decomposition of prior ties with FBK across candidates' gender, however, shows that connections are gender

[26]The reform is based on the "Legge Gelmini 30 Dicembre 2010, n. 240".

Table 4.13: Gender composition of successful and unsuccessful candidates, with and without connections with FBK

	Females	%	Males	%
Overall				
Ties	15	(12.4%)	52	(10.51%)
No ties	106	(87.6%)	443	(89.49%)
Successful				
Ties	6	(31.58%)	22	(32.84%)
No ties	13	(68.42%)	45	(67.16%)
Unsuccessful				
Ties	9	(8.82%)	30	(7.01%)
No ties	93	(91.18%)	398	(92.99%)

neutral. Table 4.13 reports the percentage of candidates with and without prior ties: overall, 12.4% of female applicants have previous ties with FBK, and the percentage for males stands at 10.51%.[27] If we look at successful and unsuccessful applicants separately, again, we do not find any statistically significant differences across genders.

We also examine the correlation between pre-existing ties with FBK and the region of origin of candidates. It may not be surprising that the proportion of candidates born in Trentino- Alto Adige or in the rest of Italy have more prior ties with FBK, as portrayed in Table 4.14.

Table 4.14: Pre-existing ties with FBK by country of origin

Origin	No ties	Ties	Total
Trentino- Alto Adige	76	24	100
Rest of Italy	90.8	9.2	100
EU and Switzerland	97.9	2.1	100
Rest of world	89.7	10.3	100
Total	89.1	10.1	100

We now examine the impact of different types of ties with FBK on the probability of being recruited. We do this in two ways: first, we estimate the conditional logit model (5) of Table 4.10 by replacing the variable "Ties with FBK" with the four variables representing the different types of ties with FBK. Second, we estimate the same model by replacing two dummy variables indicating ties with commissioners or with the institution, in order to examine which dimension prevails. Results are reported in Tables 4.15 and 4.16, respectively. The analyses are performed on the full sample of candidates, since estimating the model on the sub-samples of candidates as in Table 4.12 entails huge

[27] Differences in proportions are not statistically significant; two-group test of proportions: z = -0. 599, p = 0.549.

standard errors. This is due to the low frequencies of the different types of ties in each sub-sample considered. Linear probability model estimations on the full data set in both cases did not provide significant differences.

The estimation including the four different types of ties sheds light on the significant positive effect of two types of ties on the chances of being recruited. More specifically, we find that having one's own supervisor among the commissioners and currently working at FBK are important drivers in the recruiting process, with the former type of tie having a stronger effect. However, when we consider ties distinguishing them into ties with the commission or with the institution, we can see that both dimensions are significant, with the coefficient of the dummy variable commission being slightly higher.

Table 4.15: Probability of success - Conditional logit with four types of ties

Female	-0.284
	(0.425)
Female*(Dummy fem. in Com.)	0.989
	(0.830)
H-index standardized	0.143
	(0.129)
Age squared	-0.002***
	(0.001)
Italian origins	0.913**
	(0.356)
Ph.D.	0.636
	(0.417)
Work experience	0.096**
	(0.047)
Co-author	0.672
	(1.024)
Supervisor	2.562**
	(1.002)
Prior work	0.738
	(0.480)
Current work	1.292**
	(0.566)
Pseudo R^2	0.180
Observations	614

Notes: Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

Finally, in the model of Table 4.17 we aimed at testing whether having incremental types of ties with FBK increases the chances of being recruited, treating equally the different ties. The coefficient

Table 4.16: Probability of success - Conditional logit with Commission and Institution ties

Female	-0.335
	(0.432)
Female*(Dummy fem. in Com.)	0.945
	(0.832)
H-index standardized	0.109
	(0.131)
Age squared	-0.002**
	(0.001)
Italian origins	1.015***
	(0.356)
Ph.D.	0.695*
	(0.414)
Work experience	0.096**
	(0.047)
Commission	1.815***
	(0.701)
Institution	1.225***
	(0.381)
Pseudo R^2	0.170
Observations	614

Notes: Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

of the tie indicator variable shows that having an additional tie with FBK significantly increases the chances of being hired, irrespective of the type of tie.

Table 4.17: Probability of success - Conditional logit with tie indicator

Female	-0.341
	(0.424)
Female*(Dummy fem. in Com.)	0.990
	(0.824)
H-index standardized	0.128
	(0.124)
Age squared	-0.002***
	(0.001)
Italian origins	0.972***
	(0.353)
Ph.D.	0.669
	(0.416)
Work experience	0.100**
	(0.047)
Tie indicator	1.120***
	(0.287)
Pseudo R^2	0.170
Observations	614

Notes: Symbols *, ** and *** indicate that coefficients are statistically significant at 10%, 5% and 1% levels, respectively.

Having identified in pre-existing ties an important determinant of candidates' success at FBK, we conduct a post-competition analysis on those candidates who won a competition, splitting them into two groups: applicants with pre-existing ties and without pre-existing ties with the institution. The rationale for this is to investigate the role of ties in the ex-post candidates' research productivity, since hiring through networks can carry two opposite effects: the negative effect of an evaluation bias driven by acquaintances, and the positive effect of reducing information asymmetries regarding candidates' quality and research potential. It could hence be the case that the connections mechanism present at FBK competitions might contribute to hiring the most productive individuals and hence to making the selection more efficient.

We limit the analysis to the sub-sample of successful candidates, consisting of 86 observations, and we compare the number of publications in the two years following the competition between those individuals with prior ties and those without prior ties.[28] Table 4.18 shows that the h-index

[28] In this case we opted for a comparison of the number of publications rather than of h-indexes, since for a period of only two years gathering citations for the articles published in this very short time lapse is extremely rare and variability in the data would not be sufficient.

and number of publications of successful candidates with prior ties is on average higher than for successful candidates without prior ties. If we perform a Wilcoxon rank-sum test on the number of publications the difference is statistically significant at the 10% level (z= -1.667, p= 0.0954), although it is not for the standardized number of publications (z= -1.574, p= 0.1156).

Table 4.18: Post-competition productivity of successful candidates

	Ties	No Ties
Average N. of publications	4.78	3.29
Average N. of publications (std)	0.38	0.13

Notes: Data refer to the two years following the year of the competition and are retrieved from Scopus. The standardization occurs at the field level, as of Table 4.5.

There is hence some evidence of a positive informational effect of ties, suggesting that networks may help to select more productive candidates in terms of scientific publications.

When we decompose the post-competition productivity of successful candidates across the different types of ties, as in Table 4.19, we see that candidates with a co-author in the commission are on average more productive than candidates with other ties. Performing Wilcoxon rank-sum tests on all the average number of publications of the different tie variables, we find that only the co-author dummy is significant, both in levels and in standardized values (z = -2.131, p = 0.0331; z = -1.998, p = 0.0457). This means that the post-competition productivity of successful candidates who had a co-author in the commission is higher than the post-competition productivity of successful candidates who did not have a co-author among the commissioners.
This may suggest that commissioners who probably work at FBK and who are co-authors with successful candidates exert a positive effect on the propensity of publication of candidates. By replicating this analysis on the Commission and Institution dummy variables for ties, we find that only the post-competition productivity of candidates with a commission tie is significantly higher than the post-competition productivity of candidates without a tie in the commission, both in levels and in standardized values (Wilcoxon rank-sum test, z = -2.215, p = 0.0267; z = -2.002, p = 0.0453). Therefore, it emerges that having networks with commissioners is positively correlated with the post-competition propensity of publication, especially when the tie is represented by a co-authorship.

Table 4.19: Post-competition productivity of successful candidates across ties

	Average N. of publications	Average N. of publications (std)
Co-author	9	0.997
Supervisor	7.3	0.765
Prior work	5.63	0.539
Current work	6.08	0.680
Commission		
Yes	6.8	0.629
No	3.38	0.156
Institution		
Yes	5	0.442
No	3.36	0.132

Notes: Data refer to the two years following the year of the competition and are retrieved from Scopus. The standardization occurs at the field level, as of Table 4.5.

4.6 Discussion and conclusion

Gender disparities in some professional fields are a highly-debated issue. Several policy solutions have been proposed to reduce gender disparities, such as gender quotas for employees and, lately, gender quotas in evaluating commissions for research positions. The latter is motivated by possibly gender-biased preferences towards same-sex candidates in evaluating commissions, which are generally male-dominated.

In this paper we analyze whether female commissioners foster the hiring of female candidates. The analysis uses data from recruitments at FBK, an Italian research center mainly operating in hard science, a typically male-dominated field. Based on data on recruitment processes occurred between 2009 and 2011, we investigate the main determinants in the selection process. We study if the gender composition of the evaluating commissions influences the selection of candidates in terms of gender. Our data set allows us to analyze factors influencing recruitment at the start of a research career. Unlike other studies, our analysis includes mainly post-docs, temporary jobs and research fellowships.

Furthermore, our study is the first one in the strand of gender discrimination research to use bibliometric data retrieved from Scopus, which is widely recognized as being more reliable that other databases previously used for such types of analysis.

The most important determinants for successful applications are Italian origin of the candidate and prior ties with FBK. Ties play a major role if a commissioner member is the Ph.D./Master thesis supervisor of the candidate or if the candidate is currently working at FBK. The absence of evidence of discrimination on the whole data set is in contrast to previous research. Previous studies found either same-sex preferences (De Paola and Scoppa, 2011), opposite-sex preferences (Bagues et al., 2014) or mixed preferences depending on the level of the position (Zinovyeva and Bagues, 2011).

Candidates' h-index increases the probability of being recruited more when commission members themselves have a higher h-index. The effect of prior ties vanishes when candidates are evaluated by commissions belonging to the fourth quartile of the distribution of commissions' h-index. We find that for commissions with lower h-index, prior ties with FBK are the most important factor for recruitment.

Interestingly, for entry level research positions we find that female candidates are discriminated against by all-male commissions, even after controlling for their research productivity and education, as well as for socio-demographic variables and prior ties with the institution. The presence of a female researcher in the evaluating commission increases the probability of hiring female candidates, thus lending support to a potential benefit of gender quotas at lower research levels. The same does not hold for top research positions, for which ties with FBK and Italian origin are the main drivers for selection. We do not find evidence of gender discrimination in hiring, except for lower research positions. However, looking at the internal organizational chart of FBK, women share declines with the hierarchy rank, as in most organizations. The real challenge is thus to understand how to increase the percentage of female researchers at top positions.

From our analysis, it is evident that prior ties with the research center play a major role in the recruitment decisions. For this reason we investigated more thoroughly the effect of pre-existing ties and we found that having won a competition having prior ties with FBK entails significantly more new publications in the two years following the competition. This implies that there might be a possible positive role for prior ties in reducing the information gap between candidates and the evaluating commissions about candidates' research quality, confirming a result found by Bagues and Zinovyeva (2012). They posit the existence of an optimal distance in terms of pre-existing ties between evaluators and candidates in their analysis on academic promotions to Associate and Full Professorships in Spain. Being selected through weak ties with commissioners enhances ex-post research productivity, whereas being promoted with strong ties or without ties hinders candidates' research outcome.[29] They explain this result suggesting that with strong ties the selection bias driven by the acquaintance would be too prevalent, while with no pre-existing ties there would be no informational advantage

[29]Strong connections are those with a Ph.D. thesis advisor or with a coauthor, while weak ties are identified in same university colleagues and to a lesser extent in Ph.D. thesis defense members.

about applicants' potential. Unlike Bagues and Zinovyeva (2012), in our analysis we are not able to distinguish between strong and weak ties. The aspect of different kinds of ties is worth additional research. Further research should also be carried out on promotion mechanisms operating at FBK: it would probably be possible to isolate in a more explicit way contingent discrimination dynamics and the role of the gender composition of promoting commissions because prior ties should lose their importance. Promotions represent indeed another significant portion of the story of why women are under-represented in top rank positions.

4.A Appendix

4.A.1 List of variables

Table 4.20: List of variables

Variable	Description
Female	Dummy variable assuming value 1 if the candidate is female and value 0 otherwise
Success	Dummy variable assuming value 1 if the candidate is recruited and value 0 otherwise
Dummy fem. in Comm.	Dummy female in commission: dummy variable assuming value 1 if the candidate is evaluated by a commission in which there is at least one female researcher (no HR) and value 0 otherwise
H-index standardized	Candidate's h-index at the year of the call, standardized by her macro field of research as indicated in Table 4.5
Age	Candidate's age in years
Age squared	Candidate's age squared
Italian origins	Dummy variable assuming value 1 if the candidate has an Italian nationality and value 0 otherwise
Ph.D.	Dummy variable assuming value 1 if the candidate has earned a Ph.D. and value 0 otherwise
Work experience	Candidate's work experience in years, excluding years exclusively devoted to Ph.D.

Continued on next page

Table 4.20 – *Continued from previous page*

Variable	Description
Ties with FBK	Dummy variable assuming value 1 if the candidate complies with at least one of the following criteria, and value 0 otherwise: 1) a co-author with a member of the evaluating commission; 2) supervision at the master or Ph.D. level by one of the commission members; 3) prior work experience (including internship) at FBK; 4) current work experience (including internship) at FBK
Commission	Dummy variable assuming value 1 if the candidate has ties with FBK of type 1) or 2) and value 0 otherwise
Institution	Dummy variable assuming value 1 if the candidate has ties with FBK of type 3) or 4) and value 0 otherwise
Tie indicator	Variable assuming value 0 if the candidate has no ties with FBK, value 1 if she has one type of tie with FBK, value 2 if she has two types of ties with FBK, value 3 in case of three types of ties and value 4 in case of four different types of ties with FBK
Upper positions	R1, R3, R4 positions
Lower positions	T4, post-doc, co.co.pro. positions
Lower h	Competitions in social sciences within engineering fields, environmental sciences, engineering and computer sciences (fields characterized by lower h-indexes)
Higher h	Competitions in math, physics and chemistry (fields characterized by higher h-indexes)
Female commissioner	Dummy variable assuming value 1 if the commissioner is female and value 0 otherwise
Female commissioner (no HR)	Dummy variable assuming value 1 if the commissioner is a female researcher, not from the HR division, and value 0 otherwise
Origin commissioner	Dummy variable assuming value 1 if the commissioner comes from outside Italy and value 0 otherwise
Age commissioner	Commissioner's age in years
Human resources	Dummy variable assuming value 1 if the commissioner is from the HR division and value 0 otherwise
H-index commissioner	Commissioner's h-index at the year of the call
Monthly wage	Monthly wage of the posted positions expresses in Euros
Contract length	Contract length of the posted positions expressed in months

4.A.2 Descriptive statistics of different research levels

Table 4.21 aims to shed light on some descriptive statistics of the various research levels at FBK.

Table 4.21: Descriptive statistics of different research levels

	Mean	Std. Dev.	Min	Max
Monthly wage	2687	905.228	750	7234
R1	7234	0	7234	7234
R3	3113	124.320	2667	3156
R4	2842	124.221	2750	3150
T4	2750	0	2750	2750
post-doc	3156	0	3156	3156
co.co.pro.	2394	815.445	750	4388
Applicant's age	30.65	5.774	19	60
R1	42.88	9.978	25	52
R3	32.94	5.431	26	55
R4	30.65	5.295	24	56
T4	26.59	5.378	20	37
post-doc	32.33	3.244	26	37
co.co.pro.	30.17	5.539	19	60
Contract length	21.46	11.010	3	44
R1	36	0	36	36
R3	29.63	9.117	12	42
R4	25.58	9.411	8	36
T4	8.36	1.177	8	12
post-doc	30.4	6.197	24	36
co.co.pro.	18.83	10.503	3	44

Notes: The monthly wage is specified in €, applicant's age in years and contract length in months of the posted job.

4.A.3 Transformation process from ITC to FBK

In this appendix we report the original documents which regulated the transformation process from ITC to FBK. These documents highlight that at the date of the expiration of ITC (1st March 2007), all the personnel working for ITC was offered to be transferred to FBK with an equivalent contract.

Reg.delib.n. 338, Prot. n. 890-GEN/07/S007, VERBALE DI DELIBERAZIONE DELLA GIUNTA PROVINCIALE

O G G E T T O: Approvazione schema di convenzione disciplinante i rapporti finanziari e organizzativi per la gestione del personale dell'Istituto Trentino di Cultura, che verr inquadrato nel ruolo unico della Provincia Autonoma di Trento e messo a disposizione della Fondazione Kessler, ai sensi dell'articolo 28, comma 8 della Legge provinciale 2 agosto 2005, n. 14.

Il relatore comunica:

con Legge provinciale 2 agosto 2005, n. 14

"Riordino del sistema provinciale della ricerca e dell'innovazione. Modificazioni delle L.P. 13 dicembre 1999, n. 6, in materia di sostegno dell'economia, L.P. 5 novembre 1990, n. 28, sull'Istituto agrario di San Michele all'Adige, e di altre disposizioni connesse", la Provincia promuove la costituzione di una fondazione denominata "Fondazione Bruno Kessler", quale Ente di interesse pubblico senza fini di lucro. Con questa legge riconosciuta alla fondazione la personalit giuridica di diritto privato.

Ai sensi di tale legge l'Istituto trentino di cultura, istituito dalla legge provinciale 29 agosto 1962, n. 11 soppresso e il personale, gi dipendente dello stesso, con rapporto di lavoro a tempo indeterminato, che non istauri un rapporto di lavoro con la Fondazione, trasferito nel ruolo unico del personale della Provincia ed contemporaneamente messo a disposizione della Fondazione Bruno Kessler.

Premesso che il rapporto di lavoro del personale trasferito regolato dai contratti collettivi applicati alla data del trasferimento, con i successivi adeguamenti, a norma del citato articolo 28, occorre approvare un apposito schema di convenzione, al fine di regolare i reciproci rapporti tra Provincia autonoma di Trento e la Fondazione Kessler, per la gestione del personale messo a disposizione.

Come previsto dall'articolo 7 della legge provinciale 29 dicembre 2005, n. 20, la spesa per la messa a disposizione di personale a favore di altri soggetti privati, deve sempre far carico al soggetto utilizzatore.

L.P. 2-8-2005 n. 14 Riordino del sistema provinciale della ricerca e dell'innovazione. Modificazioni delle L.P. 13 dicembre 1999, n. 6, in materia di sostegno dell'economia, L.P. 5 novembre 1990, n. 28, sull'Istituto agrario di San Michele all'Adige, e di altre disposizioni connesse. Pubblicata nel B.U. Trentino-Alto Adige 9 agosto 2005, n. 32.

Capo VI

Disposizioni transitorie e finali.

Art. 28

Disposizioni per l'avvio della fondazione Bruno Kessler e soppressione dell'Istituto trentino di cultura.

1. Il Presidente della Provincia assume gli atti necessari affinch la fondazione Bruno Kessler sia costituita entro diciotto mesi dalla data di entrata in vigore di questa legge.

2. Dalla data fissata dalla Giunta provinciale, comunque entro il 1 marzo 2007, l'Istituto trentino di cultura, istituito dalla legge provinciale 29 agosto 1962, n. 11, soppresso e i suoi organi decadono, ad eccezione del direttore e del collegio dei revisori dei conti, che rimangono in carica per la redazione del rendiconto generale finale e, rispettivamente, per l'esame e l'attestazione della correttezza dei valori riportati, da ultimare entro tre mesi dalla soppressione dell'ente. (10) (11)

3. Dalla data di cui al comma 2, la fondazione Bruno Kessler subentra nei rapporti giuridici attivi e passivi gi facenti capo all'Istituto trentino di cultura secondo le modalit e i criteri stabiliti dalla Giunta provinciale, fatto salvo quanto previsto da quest'articolo in materia di personale. A seguito della soppressione dell'Istituto trentino di cultura i riferimenti ad esso contenuti nella vigente legislazione provinciale s'intendono sostituiti con il riferimento alla fondazione Bruno Kessler. (12)

4. Dalla data prevista dal comma 2, il personale con rapporto di lavoro a tempo indeterminato dipendente dell'Istituto trentino di cultura, che non instauri il rapporto di lavoro con la fondazione, trasferito nel ruolo unico del personale della Provincia ed messo a disposizione della fondazione Bruno Kessler. Il personale trasferito alla Provincia pu comunque chiedere di essere assunto presso la fondazione entro centoventi giorni dall'adozione da parte del consiglio di amministrazione della deliberazione che individua il contratto collettivo di cui all'articolo 13. (13)

5. Al personale con rapporto di lavoro a tempo indeterminato, che instaura il rapporto di lavoro con la fondazione o trasferito nel ruolo unico del personale della Provincia, riconosciuta tutta l'anzianit maturata nell'ente di provenienza.

6. Dalla data indicata dal comma 2, la fondazione Bruno Kessler subentra all'Istituto trentino di cultura nei rapporti di lavoro a tempo determinato per la durata residua dei contratti.

7. Il rapporto di lavoro del personale trasferito nei ruoli provinciali regolato dai contratti collettivi applicati nei suoi confronti alla data del trasferimento nel ruolo della Provincia, con i successivi adeguamenti.

8. La gestione giuridica ed economica del personale di cui al comma 4, cui possono provvedere la Provincia o la fondazione, le sue modalit di utilizzo e quant'altro necessario sono regolati mediante intese fra la Provincia e la fondazione Bruno Kessler.

5 Summary and concluding remarks

This thesis contains two experimental essays that explore the role of cognitive costs in imitative dynamics and the role of effort decisions on information aversion, respectively. The other essay provides an empirical application which examines whether gender quotas in evaluating commissions foster the hiring of female researchers.

The aim of this general conclusion to the thesis is to discuss the limitations of the three essays, that constitute a baseline for future research directions.

In the first essay we aimed at studying the extent of imitation as a possible cognitive short-cut when individuals face decisions cognitively demanding. To this end, the experimental design develops across two dimensions of treatments, that allow us to disentangle the role played by the availability of social influence in a high-demanding scenario in cognitive terms versus a low-demanding one. The first dimension involves cognitive costs: via the novel experimental task we devised we can set a low-cost scenario and a high-cost one. The second dimension of treatments concerns the social influence that subjects may receive. We model this by the amount of information provided to subjects about a default choice. In one type of treatments we tell participants that this choice belongs to the upper half of choices in terms of quality, whereas in the other set of treatments, on top of this information, we notify subjects that the choice derives from a majority of participants in a previous experimental session. Given the low frequencies of the majority option, our results suggest that imitation, conceived as a cognitive shortcut, does not seem to be present in participants' choices.

A possible explanation of this finding could be identified in the way in which the imitative component has been introduced in the experiment setting. This indeed might have been too weak in order to generate some imitative behavior. For this reason, further research could be devoted towards an endogenous imitation pattern within the experiment: the majority could be formed during the experiment and feedback about majority choices could be provided during the game.

A possible way of strengthening the social component of the experiment could be achieved by incorporating an ingroup/ outgroup framework. The idea would be to induce group identity and thus to provide information about ingroup members' choices, versus those of outgroup members. Chen and Li (2009), with their recent contribution in the social identity literature, propose an effective tool for inducing group identity. The task is characterized by social-welfare-maximizing actions and by positive reciprocity. It entails subjects to chat online prior to the experiment, regarding the solution of a task, whose correct solution provides positive payoffs to the members belonging to the same group.

Another line of argument refers to the general overconfidence of subjects. With this in mind, following what a majority previously did, without knowing the details of what kind of individuals the majority consisted of, might have been too risky for such kind of overconfident agents. It is thus conceivable to design future follow-up studies of the current one with different information, across the treatments, about the composition of who has previously chosen the majority card. Telling this kind of information to participants could help increasing the attractiveness of this card and shed more light on the distinction between social and personal heuristics.

In the second essay we were interested in extending the study of information aversion and we coupled it with the possibility of exerting an effort that may improve final payoffs. The idea was to investigate the informational trade-off of the ostrich effect, by allowing subjects to acquire information that can be useful for their final payoff. Individuals receive preliminary signals and form their beliefs about the probabilities of being attached to the good and the bad box. In the experiment we relied on Bayesian beliefs rather than eliciting subjects' beliefs.

It would be interesting to extend our design by eliciting subjects' beliefs of being in the good or in the bad condition. Such further experimental sessions would help to understand how individuals update their beliefs after receiving additional information. Such evidence would provide us a glimpse in whether there exists an asymmetric belief updating after different types of signals. It could also occur that the mere belief elicitation has an effect on information acquisition or effort exertion. Participants would be asked to reflect about their condition, and thus may change their choices. If this were the case, the further experimental sessions with beliefs elicitation could represent an "attention treatment".

Our results show that risk aversion is not correlated with information aversion, although one would expect that risk averse subjects are willing to buy additional information in order to reduce the risk of their effort choices. Yet, what could be more related to information aversion is ambiguity aversion rather than risk aversion, that is the tendency of individuals to prefer outcomes with known probabilities instead of outcomes with unknown probabilities. Further experimental sessions could implement Abdellaoui et al. (2011)'s methodology in order to elicit ambiguity aversion and correlate it with information acquisition and effort completion. This would help us to shed more light on the behavioral determinants of information acquisition and effort exertion.

Further analysis could explore whether reducing the cost of information acquisition or making the returns from effort more understandable may foster information diffusion and thus effort exertion by the irrational types of individuals, who may undertake a more rational behavior in terms of effort exertion.

Finally, in the third essay we aimed to study the main determinants of victory of research competitions in an Italian context, and in particular whether the presence of one female commissioner impacts the likelihood of female participants to be selected for the research job. One main drawback of the analysis is the absence of randomization in the composition of evaluating commissions. This lack in the identification strategy does not allow us to consider our estimates in terms of causality. In fact, endogeneity concerns may arise if unobservable factors are correlated with committees' and candidates' characteristics. However, one way to deal with this problem would be to apply the sensitivity analysis proposed by Altonji et al. (2005), who propose an impact evaluation in which the identification strategy hinges on subjects' observable characteristics and in which there is no information on the selection process. The rationale is to use the degree of selection on observables as a guide to the degree of selection on the unobservables, in an attempt to measure the degree of omitted variables bias.

Moreover, a natural extension of the study could be carried out on data about promotion mechanisms operating at the institution analyzed, as promotions represent another significant portion of the story of why women are under-represented in top rank positions. By gathering data on promotions it would probably be possible to isolate in a more explicit way contingent discrimination dynamics and the role of the gender composition of promoting commissions, as prior ties would apply to all subjects already working at the research center.

Bibliography

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review*, 101(2):695–723.

Abrevaya, J. and Hamermesh, D. S. (2012). Charity and favoritism in the field: Are female economists nicer (to each other)? *Review of Economics and Statistics*, 94(1):202–207.

Akerlof, G. A. and Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *American Economic Review*, pages 307–319.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151.

Apesteguia, J., Huck, S., and Oechssler, J. (2007). Imitation: theory and experimental evidence. *Journal of Economic Theory*, 136(1):217–235.

Bagues, M., Sylos-Labini, M., and Zinovyeva, N. (2014). Do gender quotas pass the test? Evidence from academic evaluations in Italy. *Scuola Superiore Sant'Anna, LEM Working Paper Series*, 14.

Bagues, M. and Zinovyeva, N. (2012). The role of connections in academic promotions. IZA Discussion Paper 6821, Bonn.

Bagues, M. F. and Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, 77(4):1301–1328.

Bala, V. and Goyal, S. (1998). Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, pages 261–292.

Barres, B. A. (2006). Does gender matter? *Nature*, 442(7099):133–136.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, pages 841–877.

Beugnot, J., Fortin, B., Lacroix, G., and Villeval, M. C. (2013). Social networks and peer effects at work. IZA Discussion Paper 7521, Institute for the Study of Labor (IZA).

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Blau, F. D., Currie, J. M., Croson, R. T. A., and Ginther, D. K. (2010). Can mentoring help female assistant professors? Interim results from a randomized trial. *American Economic Review*, 100(2):348–352.

Bosquet, C., Combes, P.-P., and Garcia-Peñalosa, C. (2014). Gender and promotions: Evidence from academic economists in France. Technical report, Sciences Po.

Broder, I. E. (1993). Review of NSF economics proposals: Gender and institutional patterns. *American Economic Review*, pages 964–970.

Brown, A. L. and Kagel, J. H. (2008). Behavior in a simplified stock market: the status quo bias, the disposition effect and the ostrich effect. *Annals of Finance*, 5(1):1–14.

Calvo-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267.

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

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, pages 306–318.

Camerer, C. F. (2003b). Strategizing in the brain. *Science*, 300(5626):1673–1675.

Caplan, L. (1994). Patient delay in seeking help for potential breast cancer. *Public Health Reviews*, 23(3):263–274.

Caplin, A. and Eliaz, K. (2003). Aids policy and psychology: A mechanism-design approach. *RAND Journal of Economics*, pages 631–646.

Caplin, A. and Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics*, pages 55–79.

Caplin, A. and Leahy, J. (2004). The supply of information by a concerned expert. *The Economic Journal*, 114(497):487–505.

Carrillo, J. D. and Mariotti, T. (2000). Strategic ignorance as a self-disciplining device. *The Review of Economic Studies*, 67(3):529–544.

Ceci, S. J. and Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8):3157–3162.

Çelen, B. and Kariv, S. (2004). Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, 94(3):484–498.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Charness, G. and Rabin, M. (2005). Expressed preferences and behavior in experimental games. *Games and Economic Behavior*, 53(2):151–169.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, pages 431–457.

Cicognani, S. and Mittone, L. (2014). Over-confidence and low-cost heuristics: An experimental investigation of choice behavior. *Economics: The Open-Access, Open-Assessment E-Journal*, 8(2014-36).

Clark, J. and Friesen, L. (2009). Overconfidence in forecasts of own performance: An experimental study. *The Economic Journal*, 119(534):229–251.

Clark Blickenstaff, J. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education*, 17(4):369–386.

Conte, A., Di Cagno, D. T., and Sciubba, E. (2009). Strategies in social network formation. Technical report, Jena economic research papers.

Cotter, D. A., Hermsen, J. M., Ovadia, S., and Vanneman, R. (2001). The glass ceiling effect. *Social Forces*, 80(2):655–681.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, pages 448–474.

Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education*, 1977(17):1–15.

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2005). Male and female competitive behavior: experimental evidence. IZA Discussion Paper 1833, Bonn.

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1):816–835.

De Paola, M., Gioia, F., and Scoppa, V. (2013). Are females scared of competing with males? Results from a field experiment. IZA Discussion Paper 7799, Bonn.

De Paola, M., Ponzo, M., and Scoppa, V. (2014). Chi sale in cattedra. Lavoce.info.

De Paola, M. and Scoppa, V. (2011). Gender discrimination and evaluators gender: Evidence from the Italian academy. Technical report.

Downs, J. S., Loewenstein, G., and Wisdom, J. (2009). Strategies for promoting healthier food choices. *American Economic Review*, pages 159–164.

Dubra, J. (2004). Optimism and overconfidence in search. *Review of Economic Dynamics*, 7(1):198–218.

Duffy, J. and Feltovich, N. (1999). Does observation of others affect learning in strategic environments? an experimental study. *International Journal of Game Theory*, 28(1):131–152.

Earl, P. E., Peng, T.-C., and Potts, J. (2007). Decision-rule cascades and the dynamics of speculative bubbles. *Journal of Economic Psychology*, 28(3):351–364.

Earl, P. E. and Potts, J. (2004). The market for preferences. *Cambridge Journal of Economics*, 28(4):619–633.

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.

Eckel, C. C., Grossman, P. J., Johnson, C. A., de Oliveira, A. C., Rojas, C., and Wilson, R. K. (2012). School environment and risk preferences: Experimental evidence. *Journal of Risk and Uncertainty*, 45(3):265–292.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, pages 114–138.

Eliaz, K. and Schotter, A. (2010). Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior*, 70(2):304–324.

Eliaz, K. and Spiegler, R. (2006). Can anticipatory feelings explain anomalous choices of information sources? *Games and Economic Behavior*, 56(1):87–104.

Facione, N. C. (1993). Delay versus help seeking for breast cancer symptoms: a critical review of the literature on patient and provider delay. *Social Science & Medicine*, 36(12):1521–1534.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., and Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google scholar: strengths and weaknesses. *The FASEB Journal*, 22(2):338–342.

Festinger, L. (1962). *A theory of cognitive dissonance*, volume 2. Stanford University Press.

Flory, J. A., Gneezy, U., Leonard, K., and List, J. A. (2012). Sex, competitiveness, and investment in offspring: On the origin of preferences. *Artefactual Field Experiments*, 72.

Fortin, B., Lacroix, G., and Villeval, M.-C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11 - 12):2089–2112.

Frattini, R. and Rossi, P. (2013). Report sulle donne nell'università italiana. Technical report, Mendizero, Rivista dell'Università in movimento.

Frechette, G. R. and Schotter, A. (2015). *Handbook of Experimental Economic Methodology*. Oxford University Press.

Frick, B. (2011). Gender differences in competitiveness: Empirical evidence from professional distance running. *Labour Economics*, 18(3):389–398.

Galai, D. and Sade, O. (2006). The Ostrich effect and the relationship between the liquidity and the yields of financial assets. *The Journal of Business*, 79(5):2741–2759.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62(1):451–482.

Gill, D. and Prowse, V. (2013). A novel computerized real effort task based on sliders. MPRA Paper 48081, University Library of Munich, Germany.

Glaeser, E. and Scheinkman, J. (2001). Measuring social interactions. In Durlauf, S. and Young, P., editors, *Social Dynamics*, pages 83–132. Boston, MA: MIT Press.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, pages 377–381.

Golman, R. and Loewenstein, G. (2013). Curiosity, information gaps, and the utility of knowledge. SSRN Scholarly Paper 2149362, Social Science Research Network.

Gregory, R. F. (2003). *Women and Workplace Discrimination: Overcoming Barriers to Gender Equality*. Rutgers University Press.

Hastie, R. and Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage.

Hayakawa, H. (2000). Bounded rationality, social and cultural norms, and interdependence via reference groups. *Journal of Economic Behavior & Organization*, 43(1):1–34.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 102(46):16569–16572.

Holzer, H. and Neumark, D. (2000). Assessing affirmative action. *Journal of Economic Literature*, 38(3):483–568.

Jones, S. R. (1984). *The economics of conformism*. Blackwell.

Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2):95–115.

Klayman, J. and Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2):211.

Kleinjans, K. J. (2009). Do gender differences in preferences for competition matter for occupational expectations? *Journal of Economic Psychology*, 30(5):701–710.

Koszegi, B. (2003). Health anxiety and patient behavior. *Journal of Health Economics*, 22(6):1073–1084.

Lerman, C., Hughes, C., Lemon, S. J., Main, D., Snyder, C., Durham, C., Narod, S., and Lynch, H. T. (1998). What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. *Journal of Clinical Oncology*, 16(5):1650–1654. PMID: 9586874.

Lerman, C., Hughes, C., Trock, B. J., Myers, R. E., Main, D., Bonney, A., Abbaszadegan, M. R., Harty, A. E., Franklin, B. A., Lynch, J. F., et al. (1999). Genetic testing in families with hereditary nonpolyposis colon cancer. *Jama*, 281(17):1618–1622.

Lyter, D. W., Valdiserri, R. O., Kingsley, L. A., Amoroso, W. P., and Rinaldo Jr, C. R. (1987). The hiv antibody test: why gay and bisexual men want or do not want to know their results. *Public Health Reports*, 102(5):468.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Marianne, B. (2011). New perspectives on gender. *Handbook of Labor Economics*, 4:1543–1590.

Meechan, G., Collins, J., and Petrie, K. (2002). Delay in seeking medical care for self-detected breast symptoms in new zealand women. *The New Zealand Medical Journal*, 115(1166):U257–U257.

Meho, L. I. and Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11):1711–1726.

Meulders, D. and O'Dorchai, S. P. (2013). She Figures 2012: Women and science: Statistics and Indicators. ULB Institutional Repository 2013/135739, ULB – Universite Libre de Bruxelles.

Miron-Shatz, T., Stone, A., and Kahneman, D. (2009). Memories of yesterdays emotions: Does the valence of experience affect the memory-experience gap? *Emotion*, 9(6):885.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. NBER Working Paper 17014, National Bureau of Economic Research.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, pages 1067–1101.

Nosarti, C., Crayford, T., Roberts, J., Elias, E., McKenzie, K., and David, A. (2000). Delay in presentation of symptomatic referrals to a breast clinic: patient and system factors. *British Journal of Cancer*, 82(3):742.

O'Donoghue, T. and Rabin, M. (1999). Incentives for procrastinators. *Quarterly Journal of Economics*, pages 769–816.

Offerman, T. and Sonnemans, J. (1998). Learning by experience and learning by imitating successful others. *Journal of Economic Behavior & Organization*, 34(4):559–575.

Payne, J. W. (1976). Heuristic search processes in decision making. *Advances in Consumer Research*, 3(1):321–327.

Pokorny, K. (2008). Pay - but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behavior & Organization*, 66(2):251–264.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2):243–281.

Price, P. C. (2006). Are you as good a teacher as you think? *Thought & Action*, pages 7–14.

Richard, M. A., Grob, J. J., Avril, M. F., Delaunay, M., Gouvernet, J., Wolkenstein, P., Souteyrand, P., Dreno, B., Bonerandi, J. J., Dalac, S., et al. (2000). Delays in diagnosis and melanoma prognosis (i): the role of patients. *International Journal of Cancer*, 89(3):271–279.

Schlag, K. H., Tremewan, J., and Van der Weele, J. J. (2014). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics*, pages 1–34.

Schweizer, N. and Szech, N. (2013). Optimal revelation of life-changing information.

Selten, R. (1965). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperimentes. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie.

Shen, H. (2013). Mind the gender gap. *Nature*, 495(7439):22–24.

Sorensen, A. T. (2007). Bestseller lists and product variety. *The Journal of Industrial Economics*, 55(4):715–738.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, pages 355–374.

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy*, pages 213–225.

Sugimoto, C. R. (2013). Global gender disparities in science. *Science*, 504:211–213.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2):143–148.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Villeval, M. C. (2012). Ready, steady, compete. *Science*, 335(3):544–545.

Weber, E. U. and Johnson, E. J. (2006). Constructing preferences from memory. *The Construction of Preference, Lichtenstein, S. & Slovic, P.,(eds.)*, pages 397–410.

Weinstein, N. D. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioral Medicine*, 5(4):441–460.

Zanella, G. and Banerjee, R. (2014). Experiencing breast cancer at the workplace. Technical report, Quaderni-Working Paper DSE.

Zinovyeva, N. and Bagues, M. F. (2011). Does gender matter for academic promotion? Evidence from a randomized natural experiment. IZA Discussion Paper 5537, Bonn.