Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN ECONOMIA

Ciclo XXIII

Settore scienntifico disciplinare di afferenza: SECS-P06

Essays in behavioral economics

Presentata da: Mattia Nardotto

Coordinatore Dottorato
Andrea Ichino

Relatore
Giacomo Calzolari

Esame finale anno 2011

A Sofia

Contents

Introduction

In the last decade much effort has been devoted to enrich the standard approach to modeling economic behavior. Field evidence, economic experiments and established results in other disciplines of social sciences, spurred a large debate among economists on topics which are commonly labeled as behavioral economics. The contributions to the discussion have been both empirical and theoretical. In the latter case, the need to explain the observed behavior of economic agents, triggered the proposal of behavioral models aimed to capture what is left unexplained by the standard model of rational choice, adapting the analysis of trade-offs to contexts where expectations and preferences may not be shaped consistently with the conventional assumptions.

From the point of view of the scientific method, the observation of empirical regularities which are difficult to reconcile with the standard theory have been the main motivation for the proposal of extensions. In this respect however, three observations can be made. First, from a theoretical point of view it is still missing a unified framework to analyze deviations from the so called perfectly rational behaviour and the vast majority of models which have been proposed are designed *ad-hoc* to explain the observed phenomenons in particular contexts. Second, given the previous point, also the empirical analysis, especially when it deals with field data as opposed to laboratory experiments, finds some difficulties in showing uncontroversially that behind these deviations there is indeed the proposed explanation and not some other competing ones. Third, even when behavior is shaped accordingly to the proposed theory, it is not always clear which are the implication of these findings in terms of the general theory of economic behavior. This last point is related crucially to the fact that very often deviations from rational behavior are considered as biases, meaning that they are departures from the the prescriptions of the rational model. This poses an straightforward question: once established the existence and the relevance of a bias, how stable, or persistent, is it? Economists recognize that most of the strength of the standard paradigm of rational economic choice is based on two considerations: the normative aspect of the theory and the idea that rationality arises with experience and as an equilibrium outcome. Hence, it is fundamental, in proposing a behavioral assumption to explain a particular phenomenon, to evaluate its empirical relevance and to understand its theoretical implications. These fundamental issues, which are at the basis of the process of construction, enlargement and revision of any paradigm in science, motivate our analysis.

In the first two chapters we conduct an inquiry aimed to test a behavioral model of choice. To address the first issue we raised, namely the relevance of the bias, we test some implications of both the rational and the behavioral model with a rich dataset built on this purpose. The environment we study is a subscription market for sport activity (i.e. a gymnasium) and we are interested in analyzing the choice behavior of users both from a static and a dynamic perspective, in order to address also the third issue we have highlighted: the persistence of the bias.

One of the main difficulties in the analysis is given by what we have listed has the second issue: the need to distinguishing among alternative explanations for the observed behaviour. In our, as in many other contexts, this is due to fact that behind choices and outcomes there can be many different decision processes (and each one of them can be explained by a different behavioral theory) which are observationally equivalent. To overcome this problem, we perform an extensive empirical investigation which relies not only on the data on choice and consumption but also contemplates to recover crucial information through surveys, laboratory experiments (performed online) and a field experiment.

The first chapter starts contrasting the rational expectation model of choice with a model based on two assumptions: firstly, that agents are subjected to self control costs in being stick to their investment plan (exercising in the gymnasium is considered an investment activity in future health since benefits are delayed in time while costs are anticipated) and secondly that they may be partially unaware of this costs. Based on these two assumptions we propose a behavioral model of overconfidence and we derive predictions on choice behavior especially in relation with foreseeable costs that we are able to quantify in the empirical analysis. This allows us to contrast the predictions of the rational model of choice with those of the behavioral one. Then we quantify, both from a static and a dynamic perspective, the role of expectations in determining the contractual choice of agents. In this respect we go one step further the current research which has focused mainly on the presence of the bias without answering to the question of whether it is transitory or not. We find that a relevant share of the population of users tend to behave accordingly to the behavioral model and, more importantly, they tend to adjust their expectations, and their choices, very slowly if not at all. These findings have relevant implications since we show that overconfidence is a persistent bias in the perception of own performance, leading to inefficient choices and possibly underinvestment.

Another contribution we make to the literature is to show the relationship between cognitive skills and the observed behavior. As highlighted by the recent literature which investigates the role of cognition in shaping preferences and influencing the decision processes, we find significant relationships between these cognitive characteristics and both the economic choices and beliefs we observe in our economic context.

The second chapter is motivated by the first one since we test the possibility to help agents in investing in their activity. We achieve this goal providing two types of information: reminding the availability of the activity and providing information about their actual performance. Moreover, we extend the model based on overconfidence incorporating two elements which are likely to affect behavior on our context: limited attention and mental accounting. To test our predictions we conduct a field experiment where randomized groups of users are assigned respectively to an informational treatment (i.e. they receive reminders and feedbacks) and to the control treatment (i.e. they do not receive any information as regular users in the gymnasium).

The first contribution of this work is to show the positive and significant effect of reminders on the number of visits. In particular, we show two effects: first, an aggregate effect on monthly visits which increase by 11% on average and 25% if we consider the group of users which was displaying the lower monthly attendance before the start of the treatment; second, we quantify the immediate effect of reminders, showing that the probability of observing a visit in the 24 hours after receiving the reminder increases by 8 percentage points. The result on the aggregate attendance is important since it is comparable in size with the ones obtained in other studies where monetary incentives to exercise have been provided. Moreover, we show that when users also look at the content of feedbacks, they tend to perform even better than with reminders only.

The second main contribution of this work is to highlight the relevance of inattention and mental accounting in influencing economic behavior. We show that reminders are able to induce higher attendance because they make consumers more attentive. Moreover, we highlight that their effect is non constant over the month and in particular it becomes very small when the account (defined as the difference between the price paid by the user and one associate the best alternative in the menu) switches from red to black. These two elements are not novel in the literature (especially in psychology) but our study can be considered one of the first to provide a robust evidence in support of their importance in economics.

In the third chapter we investigate the relationship between both cognitive skills and personality traits and two fundamental elements of the preferences of individuals: the risk attitude and the degree of impatience. The methodology we follow consists in setting a controlled environment, an online experiment, where participants have to make economic choices. We employ a standard laboratory design for the elicitation of these two economic parameters so choices are incentive compatible and rewarded through real payments. Participants also go through standard tests borrowed from the psychological literature in order to recover information on cognitive skills and personality.

Our findings confirm what have been highlighted both by the psychological and economic literature when more complex outcomes have been considered, such as labor status, life expectancy, the development of unhealthy habits etc. With respect to cognitive skills, we confirm the positive association of general IQ measures with the propensity to take higher level of risk. Moreover, we highlight an interesting relationship between numeracy skills and risk neutrality. The latter finding, which is strongly significant from a statistical point of view, demonstrates what seminal studies in economics have shown in recent years. Cognitive skills are also associated with a higher preference for delayed, but larger, gratifications, meaning greater patience. However, this characteristic of individual preferences is shown too be associated to personality traits and in particular to neurocitism (or emotional stability), which is the trait defined as the degree to which a person tend to experience angriness, anxiety and in general perceives the world as threatening. Consistently with the psychological literature, we find that economic choices of participants are affected by this characteristic of their personality in a sizable way.

This study makes also two contributions to the experimental literature. First, we show that online experiments can be a reliable alternative to laboratory ones, suggesting them as a valuable option especially when the experimenter needs to overcome the usual limitations of laboratory experiments, in particular the restricted number of participants and the problem of their homogeneity in terms of socio-demographic characteristics. Second, we highlight the importance of controlling for certain individual characteristics which are likely to affect behavior in experiments. In particular, we focus on the impulsivity of participants in performing their task, since it can be the cause of frequently observed paths in behavior which are otherwise difficult to explain and produce noise in the data.

The fourth chapter investigates the effect of different incentive schemes on schooling achievement hence making a bridge between the literature in personnel economics on incentives and the growing economic literature on education. We run a field experiment in an undergraduate class at the University of Bologna assigning randomly students to incentive schemes which differ in terms of how they reward the academic performance. Three schemes are considered: a tournament one that fosters competition between coupled students, a baseline scheme such that each student is rewarded on the basis of his own performance and a cooperative scheme which incentivizes, through joint rewarding, students to cooperate. These schemes are also analyzed in a formal model and we compare the theoretical predictions with the actual data coming from the experiment.

Consistently with the model, we find that the effort exerted by students in the competitive scheme proves to be higher than in the other two treatments. However, also a strong gender effect emerges and we show that the effect of competition is entirely driven by male students while female do not display any significant difference across schemes. Finally, the size of the effect we induced in the competitive treatment is comparable with the recent studies which incentivized schooling performance through monetary incentives.

Chapter 1

Stable naive and math sophisticated, a detailed field study on overconfidence

abstract

Evidence suggests that people experience self control problems and a large body of theoretical literature has modeled overconfidence on future behaviour. To assess empirically these models we analyze data on contractual choices and attendance to a gymnasium and perform experiments on the same population. The rational choice model is rejected as a comprehensive explanation of observed behaviour while a clear field evidence of users with optimistic beliefs on future behaviour emerges in the data. Following the categorization proposed in the literature on self control, we find that the population is mostly divided between *naive* and *rational* users, while *sophisticated* users are a small minority. We analyze the role of experience in changing behaviour and beliefs of users and we find that it plays a little role in improving expectations on future behaviour and optimality of choices. Finally, we run some standard tests to elicit cognitive and non cognitive traits on the same users and we find that *naive* users perform poorly in the cognitive test designed to measure numeracy skills. Moreover, the data show a significant positive relation between an important personality trait, conscientiousness, and the ability to stick to plans.

JEL codes: D12, D81, D91

Keywords: overconfidence, empirical analysis, experience, cognitive skills

1.1 Introduction

In many environments agents experience difficulties to stick to planned decisions. Moreover, when they have to choose between different plans of consumption, evidence shows that very often the expense is not minimized when compared to some available alternatives. Examples of this finding are reported in Heidhues and Köszegi (2010), Grubb (2009) and Della Vigna and Malmendier (2006) who bring evidence respectively on the credit card market, the telephone market and the sport industry. Together with the accumulation of evidence a number of models on overconfidence and self control have been proposed such as O'Donoghue and Rabin (2001) and Gul and Pesendorfer (2001), and to derive predictions for the market outcome, as in Della Vigna and Malmendier (2004). So far the literature has been very active in proposing models to rationalize this behaviour but clear field evidence on the presence of self control problems is still limited. The aim of our work is to fill this gap.

As in the case of Della Vigna and Malmendier (2006) (hereafter DVM) we focus on the sport industry and we analyze data provided by a gymnasium of the University of Bologna. This market is generally characterized by a simple contract schedule: users can exercise buying a ticket for each visit (i.e. *pay as you go*) or they can pay in advance a fee that allows them to attend at zero cost until the expiration date (i.e. *flat* contracts). Despite the simple contract options available, this environment is rich enough to make the economic problem of choice interesting since users must plan in advance their future activity and coherently select their contract. The data we analyze encompass the whole list of contracts bought by 1500 users from January 2008 to August 2010 and the visits to the gym for the entire period. We complement the data on contractual choice and attendance with information on preferences, beliefs, cognitive abilities, personality traits and demographics for a relevant subsample of the population under analysis. To gather these information users participated in payed experiments and filled a detailed survey. Since users are students, we have also access to information on academic background and performance.

Our environment, as those cited above, is characterized by a relevant share of users incurring in high costs for exercising. This means that, given their *ex post* attendance, the price payed is greater than under some available alternatives. In particular, as in DVM, users buy too frequently *flat* contracts and then attend too few with respect to the number of visits that makes this type of contract less costly than attending the gym on a *pay as you go* basis.

This finding is compatible with many models of choice. The first candidate we must consider is the model of rational choice among contracts with correct beliefs on future attendance where a high cost for exercising is determined by unexpected shocks which affect users at random. In other words, agents form correct expectations on future attendance taking into account the relevant factors that will affect it and coherently select the contract that minimizes the total expense. Whenever the *ex post* attendance is not coherent with the choice, this is due to unforeseeable contingencies. We label agents who behave accordingly to this model as *rational* users. The leading alternative explanation has been proposed by the literature on self control and assumes that users are characterized by self control costs that arise at the moment of exercising. So far however, we still miss a field study that clarifies whether users are aware of their self control problems, i.e. they are *sophisticated*, and choose *flat* contracts even if they are able to forecast their future number of visits (to buy a contract that fosters attendance and so allows to reach a desired level of monthly visits) or if they are *naive* in the sense that contractual choices are simply based on a wrong belief.

We begin our empirical analysis verifying whether the population of users we observe is composed entirely by *rational* users. To address this point we exploit information on elements that affect the every day decision to exercise such as the distance from home to the gym and predictable future weather conditions like the temperature. *Rational* agents that form a correct expectation on future attendance should anticipate these elements at the moment of signing the contract and sort consistently in the menu of contracts. From an *ex-ante* perspective, users should be less likely to sign *flat* contracts when predictable costs are high, due to lower expected attendance, and no systematic relation should emerge *ex-post* between these costs and the probability that the contract chosen by the users results to be more expensive than some alternative. Contrary to the prediction of the rational sorting model, we find both a positive relation between distance and the probability of signing *flat* contracts and a strongly significant positive relation between predictable costs and the probability of incurring in high costs for exercising.

Once discarded the rational choice model as a comprehensive explanation for the observed behaviour, we identify the different types of users in the population using data on expectations about future attendance and the contractual choices they make. A crucial point of the self control model in case of naivete is the presence of wrong beliefs. On one side, *naive* users are too optimistic about future attendance and select contracts that imply a high number of visits. On the other side, *rational* and *sophisticated* users are able to forecast future attendance and their contractual choices respond to different aims. *Rational* users do not suffer any self control cost. *Sophisticated* users buy contracts that foster attendance even if they are incurring in high costs for exercising (to overcome self control issues). The data show that a large share of users are too optimistic about their future behaviour. Comparing expectations with actual attendance, we see that the size of the forecast error is positive and amount roughly to 30% of the expected number of visits. This means that on average a user who expects to visit the gym ten times in a month actually attends seven times. Combining the information on expectations with the contractual choice we identify to which group a user belongs to. The population is mainly composed by *naive* and *rational* users, while *sophisticated* users are a small minority. This result improves the analysis of DMV who illustrate that their evidence is compatible with an heterogeneous population of users but do not characterize the groups of users and their relative size.

We also test whether other explanations are compatible with observed behaviour. In particular we consider the pressure of the staff to select *flat* contracts and the role of risk aversion. We find that only the former has some importance. Users who declare to ask for suggestions from the staff then select *flat* contracts. However, this holds only for users at the very first experience and they account for a very small fraction of the whole population of users. With respect to risk aversion, the reason for selecting *flat* contracts would rely on the fact that they imply no variance in the cost of exercising while the *pay as you go* system can be *ex post* much more expensive. However, we do not find any empirical support for this claim.

Having determined the composition of the population, we address the role of experience. Do expectations improve with the accumulation of experience? Do *naive* users become *sophisticated*? We find that experience plays very little role. Beliefs on future attendance do not improve significantly with experience and users who are paying a high price for exercising at a certain point in time are likely to do so later. This completes our picture. The vast majority of *naive* users we find in the data are not inexperienced users who after the first spell either quit or become *sophisticated*, but reflects instead the presence of stable types in the population. This finding is in line with Agravwal, Driscoll, Gabaix, and Laisbon (2008) who analyze the learning process in the credit card market. Moreover, our claim is strengthened by a result on the perception of past attendance. We find that users who are too optimistic about the future display a significant positive bias in the evaluation of past performance.

We conclude analyzing the relationship between types of users previously determined and both cognitive abilities and personality traits. Here we try to answer the following questions. Who are the *naive* users? What are their characteristics, cognitive and personality traits? We rely on standard tests and questions taken from the psychological literature and employed in economic studies such as in Dohmen, Falk, Huffman, and Sunde (2010). We find that *rational* and *sophisticated* users are characterized by higher cognitive skills, in particular those related to numeracy. This is in line with Heckman, Stixrud, and Urzua (2006) and the literature on the role of cognitive abilities in determining economic outcomes. With respect to personality traits, we find that users who are overconfident on future behaviour have also a lower score in the personality trait which is related with the degree to which a person is able to stick to plans, namely conscientiousness. This finding contributes to the literature on the role of cognitive abilities and personality traits in determining economic outcomes.

The paper is organized as follows: in section 1.2 we propose a simple model of self control and we derive implications on rational sorting in our context. In section 1.3 the data are present and we describe the activities we have designed to elicit preferences and expectations. In section 1.4 is it performed the first part of the empirical analysis, testing the rational sorting model and identifying types of users. In section 1.5 we discuss the stability of types over time and the effect of experience on behaviour and beliefs. In section 1.6 we relate types with both cognitive and personality traits. We conclude in section 1.7 where we highlight the main results, the implications and possible improvements.

1.2 A simple Model

An user is willing to enter a subscription market where she can repeatedly buy an investment good which, at any consumption, involves an immediate non-monetary cost c and a delayed benefit b. For example, exercising in a gym, the user pays upfront the cost of going to the gym and the physical cost for exercising, whilst the benefit of better health and strength are postponed. Similarly, the decision to quit smoking, change alimentary habits toward a more healthy diet or save according to a specific plan can be considered examples of investment goods. For the sake of concreteness, in the sequel we will refer to consumption as exercising in a gym.

As it is often the case, the user can choose from a menu of alternative payment schemes or contracts with different properties. A *flat* contract of length x contemplates a fixed fee L_x to be paid upfront and then the user can exercise at any time for the next x periods with no additional costs. A *pay-per-visit* (PPV) contract involves no fix fee but requires the payment of a price p at any consumption date. Finally, a carnet for x consumptions contemplates an upfront payment of C_x and allows to exercise for x times with no additional price, at any future date. The last contract option is not present in DVM who study an environment where only the first two contracts are available. In our case instead this contract must be considered since the gym provides it in the menu.

To simplify exposition we make the following assumptions. At date $t = 0$ the user decides whether to buy a contract that requires an upfront payment. At any future date $t \in \{1, 2\}$ she decides whether to consume or not and at the beginning of period t she learns the realization of the stochastic cost $c \in \{\underline{c}, \overline{c}\}$, with γ and $1 - \gamma$ being respectively the probability of $\underline{c}$ and $\overline{c}$. At the end of period t she obtains the (delayed) benefit b. The user knows about the presence of future costs of exercising and realizations of cost are IID across periods. Being δ the discount factor, the decision of whether consuming or not at any t is taken considering the net present value of consumption $-c + \delta b - p$ where c is the observed realization of the cost of date t and p is the monetary cost of the visit, which is zero with a flat contract or a carnet and it is instead p with a PPV. To make the analysis interesting we assume that that:

$$-\overline{c} + \delta b > 0 > -\overline{c} + \delta b - p \quad \text{and} \quad -\underline{c} + \delta b - p > 0 \tag{1.1}$$

so that, when the realized cost is low, the user consumes independently of the contract chosen at $t = 0$ and, when instead the cost realization is high, she only consumes when she has signed in $t = 0$ either a flat contract or a carnet.

The contract choice in $t = 0$ is a bit more complicate since the user has to anticipate her future attendance in the gym. Since there are only two periods of possible consumption here we consider only a flat contract that lasts 2 periods (i.e. $x = 2$) and a carnet with 1 period of consumption available. In this case where only two consumption periods are available the difference between a *PPV* and a *carnet* is given by the fact that the latter is paid upfront while the former only at the moment of exercising.

Equation (1.2) expresses the condition under which a user prefers a *flat* contract with respect to the *PPV*. In this case L_2 must be lower than the expected payment under the *pay as you go* plus the value of the extra-visit that the *flat* contract induces in case of high cost:

$$\frac{L_2}{\delta(1+\delta)} < p\gamma + (1-\gamma)(-\overline{c} + \delta b) \tag{1.2}$$

Using the fact that $(\delta b - \overline{c}) < p$ the following result holds:

Result 1 *Given assumption (1.1) the choice of the user is such that if the flat contract is preferred to using a PPV then:*

$$\frac{L_2}{\delta(1+\delta)} < p \tag{1.3}$$

An user who chooses the *flat* contract expects to attend a number of times such that the cost of a single visit (the l.h.s. in (1.3)) is lower than the cost under the *pay as you go*.

Another useful comparison is between the *carnet* and the *PPV*. In this case we have that the former is preferred to the latter if the following condition holds:

$$\frac{C_1}{\delta} < p\gamma + (1-\gamma)(-\overline{c} + \delta b) \tag{1.4}$$

Exploiting again the condition $(\delta b - \overline{c}) < p$ then the following result holds:

Result 2 *Given assumption (1.1) the choice of the user is such that if the carnet is preferred to using a PPV then:*

$$\frac{C_1}{\delta} < p \tag{1.5}$$

Also in this case the actual cost of the visits with the carnet (the l.h.s. in (1.5)) must be at a lower price with respect to the single entrance with the *pay as you go*[1].

The last comparison is between a *flat* contract spanning two periods and a carnet followed by a *PPV*. Following the same reasoning as above:

Result 3 *Given assumption (1.1) if the flat contract is preferred to using a carnet, plus possibly a PPV then:*

$$\frac{L_2 - C_1}{\delta^2} < p \tag{1.6}$$

Equation (1.6) shows that the *flat* contract is preferred to the *carnet* when the price difference is lower with respect to the price to attend implied by the *carnet* in the second period[2].

Self control costs. So far we have assumed that the user consistently evaluates at any date t the net benefit of consumption of any future date. The economic and psychology literature has illustrated both empirically and theoretically that this may not be the case. In particular, here we are interested in introducing the possibility that although at date $t = 0$ the benefit of consumption at any future date $t > 0$ is $[-E(c) + \delta b]\delta^t$ (net of any price she may have to pay at date t depending on the chosen contract), when the consumer reaches date t she realizes that the actual benefit is lower and equal to $-E(c) + \delta b - k$ where k is a non negative scalar. A first possible interpretation of k is that the discount rate of the close future is higher than the discount rate of the distant future, as it has been emphasized by the literature on hyperbolic discounting. In particular, we may have that at t the actual value of the benefit that accrues at the end of the period is not δb but $\beta\delta b$ so that $k = (1 - \beta)\delta b$. Alternatively, the literature on temptation has pointed out that in many real situations users assign different utility to the same consumption bundle depending on the state in which the evaluation takes place. In our simple environment, when reaching date t the consumer may realize that she can opt for a previously unforeseen and tempting alternative to consumption (i.e. watching the TV instead of exercising at the gym) with value k (so that exercising will be undertaken only if better than watching the TV, i.e. $-E(c) + \delta b > k$). Users who completely ignore the fact that they will face this additional cost k for consumption at the

[1] It is worthy to notice that the carnet is also characterized by flexibility. Consider the decision to attend in $t = 1$ when the cost is known. Here the user can exploit the possibility to postpone the entrance in the following period when the realization of cost c is high. The payoffs are the following: $-\overline{c} + \delta b + \delta\gamma(-\underline{c} - p + \delta b)$ if she uses the carnet in $t = 1$ and buys the *PPV* in the second period while $\delta[\gamma(-\underline{c} + \delta b) + (1 - \gamma)(-\overline{c} + \delta b)]$ is the payoff in case of not going and preserve the carnet for the next period. Comparing the two payoffs, it is easy to see that the user is better off not going to the gym in $t = 1$ when she observes high costs to exercise. The reason for this derives from the fact that in $t = 2$ the user will face low costs of attendance with probability γ and this makes the expected utility of exercising in the following period higher with respect to do it in $t = 1$.

[2] Conversely, if the discount in the first period entrance provided by the carnet is high enough, then the user is better off choosing this option and then paying the single entrance ticket in $t = 2$.

date of consumption have been indicated in the literature as *naive* (e.g. see O'Donoghue and Rabin, 2001, or Eliaz and Spiegler, 2006). Those instead who are able to recognize at date $t = 0$ that they will face an additional cost $\hat{k} < k$ are partially *naive* and finally those with a $\hat{k} = k$ fully anticipate this future issue and have been indicated as *sophisticated* users.

To make the analysis interesting, we will assume in the following that:

$$-\overline{c} + \delta b - k < 0 < -\underline{c} + \delta b - k \tag{1.7}$$

In this case a *naive* user may choose the flat contract anticipating an optimistic high attendance. This is because she expects to exercise both with low and high cost c but, due to k, only low costs will induce a visit to the gym. This user would be better off by consuming with a PPV only when the cost realization is low if the following condition on costs and benefit is satisfied:

$$-\underline{c} + \delta b - p - k > 0 \tag{1.8}$$

When instead (1.8) is reversed, then a user is not willing to attend when she has to pay the price p at the moment of exercising. In this case a *sophisticated user* in $t = 0$ foresees her problem of self control and consciously buys a *flat* contract or a *carnet* to commit to attend in $t = 1$ or $t = 2$, at least when the cost realization is $\underline{c}$. This may determine, depending on the cost realizations in the two periods, a higher per visit cost with respect to the *PPV* which is, however, expected by sophisticated users.

Foreseeable costs and contract choice. We proceed examining in more details the role of costs in determining the choice in the menu and the subsequent observed price per visit. Suppose users are heterogeneous in some additional dimension of costs. One example can be their location in the neighborhood of the gym, determining different costs to attend due to different distances to reach the facility (we refer to distance as d).

In a population of *rational* users the effect of distance on choice is to decrease the share of users willing to buy *flat* contracts with respect to *PPV*. This is due to the reduction in the number of expected visits which makes the *PPV* more convenient.

However, this relation may be reversed when *naive* and *sophisticated* users are present in the population. Consider first the case of *naive* users. When distance is such that:

$$-\overline{c} - d + \delta b - k < 0 < -\underline{c} - d + \delta b - k \quad \text{and} \quad -\underline{c} - d + \delta b - p - k > 0 \tag{1.9}$$

then users who choose the *PPV* attend *ex-post*. However, when the distance increases enough to determine the following condition:

$$-\overline{c} - d + \delta b - k < 0 < -\underline{c} - d + \delta b - k \quad \text{and} \quad -\underline{c} - d + \delta b - p - k < 0 \tag{1.10}$$

then *naive* users who think *ex-ante* to attend with *PPV* do not show up in the gym *ex-post*. This means, in terms of observed choice and per visit price, observing a negative relation between distance and the share of *PPV* sold by the gym and at the same time a positive relation between price per visit and distance due to the low *ex-post* attendance of *naive* users who purchased *flat* contracts.

In the same way, under (1.10), *sophisticated* users who realize that the *flat* contract is the only way to commit to exercise, will not buy *PPV* when the distance is too large and attend *ex-post* at a high per visit price. These observations lead to the following testable implication:

Result 4 *In a population of rational users, a negative relation between distance and the proportion of flat contract must emerge. On the contrary, when users with self control costs are present in the population, the share of PPV may decrease. Ex-post, since naive and sophisticated users are observed only when they purchase flat contract, a positive relation between distance and high per-visit cost should emerge.*

Summarizing, the starting point of our empirical investigation is twofold. First, we discuss how choice is affected by foreseeable costs. Second, we test how these costs affect the *ex-post* price to attend. The prediction of the choice model we derived says that choices made by a population of *rational* users are characterized by a negative relation between elements affecting the cost of attending the gym and the share of flat contract signed with respect to *PPV*. On the contrary, in a heterogeneous population of users, this relation may be reversed. At the same time, in the latter case, we expect to find a positive relation between these underestimated cost factors and a high per visit price for attending evaluated considering the *ex-post* attendance. On the contrary, under the rational model, a systematic relation of this kind should not emerge. A high per visit cost to attend, labeled for simplicity "loss", derives from Results 1, Result 2 and Result 3 stated above. Hence, to compute this loss associated to a contract, we compare the observed price to attend with the price implied by the alternative as the predictions of the choice models suggest.

1.3 Data

Data come from the records of a gymnasium owned by a sport association for university students in Bologna (Italy). This gym is not strictly reserved to students but they are the vast majority of costumers since the association has the objective to promote sport activities among university students. Our analysis focuses on this subsample. In general, fees are 30% lower with respect to non student gyms but are still relevant for a student. Indeed, the typical monthly disposable income of a student is around 200 euro net of the cost of board and lodging[3] so the cost of the monthly contract amounts to 22% of this budget.

Our dataset ranges from January 2008 to July 2010. Overall, we have information on all the contracts signed and on the sport activity of 2330 users attending the gym in this time span. However, the empirical analysis is performed on a subsample of the whole population exercising in the gym. In particular, we exclude: users for whom it was not possible to recover the distance between the gym and their home[4]; users who bought contracts for specific courses since they are not suitable for computing a loss (see note 5); those who lost the electronic key and have some periods of unrecorded attendance. We end up with a sample of contractual choices made by 1535 users. Moreover, since the data on attendance are available from January 2008 but contractual choices have been recorded since September 2007, the analysis of contractual choice is performed on a slightly larger sample which includes also these contracts. For a summary of the variables employed in the whole section see appendix 1.8.

The menu of contracts. The gym offers a menu of contracts with single tickets, a set of flat contracts with different durations and two carnet of single tickets[5]. Flat contracts last for: one, two three, six months (available only in 2008) and there is a contract that lasts for the entire academic year (from 1st of September to 31th of July). A last flat contract available in the menu is the "1 euro per day promotion" which gives the right to access the gym from the date of sign to 31st of July at a price of 1 euro per day[6]. In 2009 this option was available only in summer months (from June to July 2009) while in the 2010 from November onward. Table 1.1 reports the prices of the different contracts for year 2010, together with the volume sold and the relative frequency (in the whole sample) and the average number of monthly visits attended by users.

After the expiration of any contract there is no automatic renewal. This makes our environment different from the one in DVM where monthly contracts are renewed automatically. The most important difference is that upon expiration users can adjust their contract on future expected consumption since they must explicitly choose a new contract.

Attendance. Each user has an electronic key to access the gym. To get in, the user must

[3]This figure is obtained through a survey filled by the students in the gym. More information in the appendix.

[4]The number of cases is limited and it amounts to less than 4% of the population.

[5]The actual menu is richer, with contracts for specific courses (like yoga, spinning, gymnastic for injured persons, karate etc.) and some promotions. In the former case only one option is available so we cannot use these contracts to evaluate the optimality of observed choices. Promotions will be instead used.

[6]As an example, if a user signs the contract the 3rd of June, then the cost is 59 euro (28 days in June and 31 in July).

Table 1.1: Attendance by contract category and prices (year 2010)

Type of contract	**Price (€)**	No.	Rel. freq.	Mean visits (monthly)	Conf. int.
Single entrance	6,00	548	8.8%	–	–
Carnet (10 visits)	50,00	184	3.0%	1.67	[1.49,1.84]
Carnet (20 visits)	90,00				
One month	45,00	3918	63.1%	7.88	[7.73,8.02]
Two months	85,00	308	5.0%	7.30	[6.79,7.82]
Three months	118,00	518	8.3%	6.61	[6.26,6.97]
Academic flat	270,00	58	0.9%	6.74	[5.75,7.74]
Flat 1 euro promotion	–	281	4.5%	6.65	[6.19,7.10]
Specific contracts	–	384	6.2%	5.47	[5.07,5.87]

a. Pooled average attendance of carnet is reported.

b. Price of promotions depend on the moment of signing.

c. Specific contracts have different prices.

plug the key in an electronic device and a member of the staff at the reception regularly monitors the output on the screen of a computer[7]. Those who decide to buy a single visit ticket have to provide to the receptionist the university badge and pay cash.

Average attendance for each contract category is reported in table 1.1 which shows that the attendance is higher with flat contracts than with carnet and among flat contracts it decreases with contract length.

Other variables. One key variable that we will use is the distance between the gym and the user's home. To recover this measure we employed an application provided by Google Maps which computes the linear distance between two addresses. As table 1.13 shows, 90% of the users is less than 3km far from the gym

Information on weather conditions are provided by the regional authority for environmental monitoring, which collects hourly data on precipitation and temperature.

Information on academic history of users is provided by the University of Bologna. Anonymous data include the school, the list of passed exams with their standardized grades[8]

[7]The monitoring of the entrance is in the interest of the owners of the gym. In our case the facility is not directly managed by the university association because a private shareholder of the facility is in charge of managing it.

[8]More precisely, we have information on the mean grade and the standard deviation of grades (up to four years) for each exam undertaken by the users in our sample. With such information we can compute the performance as a relative measure taking the standard ratio $\frac{mark_{i,j}-\mu_j}{\sigma_j}$ where $mark_{i,j}$ is the mark of agent i

and the date at which the exam has been undertaken. From this data we construct two variables: the average standardized mark of the user in her academic carrier and an index that accounts for the timing of the exams. The latter measure is particularly interesting because it tells how an user is able to stick to his academic commitment. In the Italian university system a student can postpone or retake an exam and only the passed exams are recorded. Finally, for each academic year, whether an user receives financial aid from the university. The provision of financial aid is granted on a wealth basis conditional on a given average mark and the amount of the grant varies from a small reduction of the tuition fees to the complete exemption or to a subsidy for studying. The information we have is related to the aid provided for the economic status of the user.

Experimental data. To collect information on preferences, expectations and traits, we implemented some web based activities that have been administered among the users of the gym in 2009 and 2010. These activities include: a general survey on sport activity, a test focused on cognitive and personality traits, an online experiments on risk and time preferences. Users have been paid for participation from 5 up to 85 euro (see the appendix for more details). The number of users who participated is 405 (65% of the typical active cohort in the gym)[9].

Through the survey on sport activity we gather information on elements that influence contractual choice and attendance. Among other questions we ask about i) the motivations for exercising; ii) the perception of the past performance; iii) the expectation on future attendance; iv) the perceived characteristics of the contract options available in the menu. The survey also includes some questions on socio demographic characteristics. We asked for example about the monthly budget, if the user is working and in case of positive answer, the type of job.

Measures on cognitive and non-cognitive traits are obtained through standardized test and questions. We employ a symbolic test as the one performed by Lang, Weiss, Stocker, and Rosenbladt (2007) to have a measure of computational speed (which is known to be strongly correlated with IQ). To evaluate the skills of users in doing calculations we also ask them to answer a set of questions designed for this purpose (see the appendix for details).

Measures of non-cognitive traits are obtained through a set of standard questions. We

---

in the course j and μ_j, σ_j are the mean and standard deviation of marks in course j. In the Italian system a student is supposed to collect 60 academic credits per year and each exam has an associated number of credits. If a student is not aligned with this schedule then every year the number of credits earned is lower than 60.

[9]The technology we adopted for the online implementation of these activities is php/MySQL. In the interactive sections (symbolic digit test and risk and time experiments) we have designed a dynamic interface in Javascript.

follow the literature in psychology and rely on self-reported traits (for a discussion of the pros and cons of this approach see Borghans, Duckworth, Heckman, and Bas 2008). We employ the same battery of questions of Dohmen, Falk, Huffman, and Sunde (2010) which are taken from the German panel survey SOEP (for more on this see Heineck and Anger 2008). The traits we are interested in belong to the "big five" personality traits list defined in the psychological literature: openness, conscientiousness, extraversion, agreeableness and neuroticism (or emotional stability). In general, each trait is elicited through different questions (each one is associated with a facet) which are combined in order to get an aggregate measure. Since exercising, and in particular being consistent with plans, is mostly related with conscientiousness, we focus more on that trait. Conscientiousness is defined as the degree to which a person is willing to comply with conventional rules, norms and standards. The main facets of this trait are: competence, order, dutifulness, achievement striving and self discipline. Emotional stability is another trait we are interested in and it is defined as the degree to which a person experience world as threatening and beyond her control. To have a measure of it, other than standard question, we also run a test developed by Frederick (2005) to measure impulsivity. We look also at two other traits: the internal and the external locus of control. With the former we identify people who believe that that the outcome they experience is determined by their own skills and behaviour. With the latter, we identify people who believe that outcomes are mostly affected by external elements beyond their control.

We employ a web version of two standard laboratory experiments on risk and time preference. The experimental design we follow is similar to Holt and Laury (2002) where subjects must fill the rows of some tables where different payments are proposed (see the appendix for more details). In the experiment on the risk preference, subjects face two tables with 10 rows each. Each row is made by a couple of alternative payments: option A and option B. The former is an amount of money that is granted for sure (e.g. 20 euro); whereas the latter is a lottery that have a small prize (e.g. 5 euro) and big prize (e.g. 35). The two payments of option B are granted with a probability 1/2. All the rows are displayed from the beginning of the stage and the sequence is such that option B remains constant while option A becomes at each row (moving downward) less attractive since it is reduced at each step by a fixed amount. In the case of time preferences, the rows present to the user two payments: a smaller amount of money in 2 days versus a larger one in a later date (e.g. 10 euro in 2 days versus 20 euro in 15 days). In order to achieve saliency, we pay subjects accordingly to their choices. To do this, at the end of the experiment one of the rows filled by the subject is chosen at random and the choice made in this row becomes the actual payment[10] Finally,

[10]Online experiments have proved to be comparable with laboratory ones. However, in online experiments

we address the issue of transaction costs which can be relevant in the time experiment. To overcome this problem, participants are paid electronically with telephone credit in order to eliminate transaction costs[11].

In order to be sure that subjects have understood the instructions before starting, the experiment requires first to complete a brief tutorial, with no associated payment. In the tutorial subjects fill a demonstrative table and a simulation of the final payment draw is performed. Moreover, we administer some questions are administered at the end of the tutorial and users cannot start the payed stages until the answers are all correct. One user out of five is actually paid at random (for more information on this experiment see Nardotto 2011a).

1.4 Empirical analysis

Contract choices and observable costs. Our analysis on contract choice and costs is focused on the role of distance in determining contractual choices made by users. We divide the contracts into three categories: *PPV*, *carnet* and *flat* contracts, where the last category contemplates all flat contracts of different durations. Computing at increasing distances the frequencies of contracts purchased, we find a diminishing path for the share of *PPV*. In particular, within the 2km circle, they amount to 12% of the contracts while *carnet* and *flat* contracts are respectively 3% and 85% of the sales. Outside the 2km circle, the fraction of *PPV* falls to 7%, mainly in favour of *flat* contracts.

To test this negative relation, we estimate a multinomial logit model where the dependent variable is the category of contract. Results of the estimations are reported in table 1.2. The table is divided in two panels. The left one reports the estimation of the multinomial logit model with the three categories defined above while in the right part we report the estimated coefficient of a logit model where *carnet* and *flat* contracts are in the same category (opposed to *PPV*). All regressions have the observed contractual choice as the dependent variables so the estimation is performed clustering at the user level, since a user signs in general more than one contract. Contractual choice is explained by the following variables: the distance from home to the gym and the distance squared, demographic characteristics

we observe a shorter answering time and more noise. A comparative study on this issue is Anderhub, Müller, and Schmidt (2001).

[11] Subject have been informed about the nature of the payment in the instructions. When the experiment took place, the vast majority of consumers in the mobile telephone market in Italy has a *pay as you go* contract so our payment is very likely to be valuable for our subject pool. Moreover, those who could not receive the payment had the possibility to contact the experimenter and ask for a different type of payment. Nobody however did use this option.

such as gender, age, a proxy for budget constraint given by the dummy for receiving financial aid from the university, a dummy that takes value 1 if the contract is the first contract bought by the agent, the experience maturated in the gym measured by the number of days spanned by previous contracts and finally the mean temperature experienced during the contract (and its square).

Table 1.2: Multinomial logit on contractual choice

Baseline outcome: *flat* contract **Dep. var: probability to purchase:**	*PPV*		*Carnet*		*PPV*	
	Whole sample	Excluding 1st contract	Whole sample	Excluding 1st contract	Whole sample	Excluding 1st contract
Gender	-.459**	-.475**	-1.02***	-1.46***	-.43**	-.43**
	(.195)	(.216)	(.291)	(.401)	(.193)	(.214)
Age	-.063	-.065	-.011	.004	-.063	-.065
	(.040)	(.044)	(.053)	(.072)	(.04)	(.044)
Distance	.782***	.614**	-.085	.122	.785***	.611**
	(.291)	(.313)	(.192)	(.26)	(.29)	(.312)
Distance^2	-.247***	-.221***	.015	-.037	-.248***	-.22***
	(.076)	(.081)	(.034)	(.05)	(.076)	(.081)
Mean temp	-.336***	-.322***	-.014	-.061	-.333***	-.339***
	(.035)	(.039)	(.06)	(.061)	(.035)	(.039)
Mean temp^2	.012***	.011***	-.0013	.0002	.012***	.011***
	(.0011)	(.0013)	(.002)	(.002)	(.001)	(.001)
Exper. days	.0004	.0003	.0007	.0008	.0004	.0003
	(.0005)	(.0006)	(.0008)	(.0008)	(.0005)	(.0006)
First contr.	-.812***		.29		-.823***	
	(.156)				(.156)	
Controls	YES	YES	YES	YES	YES	YES
N	5708	4213	5708	4213	5708	4213
Pseudo R^2	0.079	0.074	0.079	0.74	0.087	0.072

We analyze first the estimation of the multinomial logit in the left panel. This is divided in two parts, where the first is on the choice between *flat* contracts and *PPV* (first and second column) and the second is on the choice between *flat* contracts and *carnet* (third and fourth column). In the first column of each couple we report the estimates of the regression on the full sample of contracts while in the second we restrict the sample excluding the first contract bought by the user.

We start the discussion highlighting the difference between the first two columns, and the second couple of columns. Most of the coefficient in the latter case are not significantly different from zero meaning that users are not influenced by these variables in their choice between *flat* and *carnet*. Analyzing the choice between *flat* and *PPV*, two variables are worth noting before concentrating to the ones of interest. The gender variable has a significant negative effect on contract choice, meaning a negative association between the likelihood of buying a *flat* contract and being female. Hence, women are more likely to opt for *PPV* and *carnet* than men. Moreover, the dummy for the first contract is also highly significant meaning that at their first choice users are more likely to opt for a flat contract .

Moving to our main variable of interest, the distance, we find that a quadratic relation best fit the data. The minimum of this function, depending on the estimation, ranges from 1.4km to 1.6km which is inside the spectrum of distances we have in our sample (from 0 to 8km). The interpretation for this quadratic relation is given by the two effects highlighted discussing the model. Starting from the close nearby of the gym, as distance increases more *rational* users are less willing to buy *flat* contracts and their share declines with respect to *PPV*. At the same time, the effect on choices determined by *naive* and *sophisticated* users is not strong enough to invert this trend. It is only after the threshold we identify that the second effect dominates and the share of flat contracts increases. As we discussed in section 1.2, at this point we cannot distinguish between the *flat* contracts bought by *sophisticated* users (to have a commitment to attend) and the missing *PPV* of naive users who plan to attend *ex-ante* and do not show up *ex-post*. Since the carnet choice between *carnet* and *flat* contract is not affected by this factor, we aggregate the two categories and we estimate a logit model, reported in the right panel of the table. Coefficient are stable across samples and models so we compute the effect of distance using the logit estimates. We find that the change in the probability of opting for the *PPV* contract with respect to the *flat* is negative and equal to -1.6% at the the third kilometer (estimation of the full sample) and it reduces to -1.4% using the restricted sample[12].

***Ex-post* price per visit and cost factors.** The findings obtained so far have to be matched with the *ex-post* performance in the gym, so we discuss now the role of foreseeable costs in determining the actual per visit price payed by users to attend. The first point we must

[12]The computation is made on a male subject, with average age at the average temperature, studying engineering (most common choice) and not receiving financial aid. The same parameters in the case of a female user leads to higher estimated effects, slightly above 2% in absolute value.

address however, is the presence of incorrect choices.

In section 1.2 we have derived predictions on contract choice and we have defined the concept of loss in our context. To evaluate contractual choices, we compute the cost of attending under all contracts available in the menu (given observed attendance) and we compare them with the actual expense. More precisely, we compute the difference between the best alternative and the actual expense and we divide it by the latter. We multiply this measure by one hundred to have a percentage measure of the sub-optimality of observed choice. For the sake of brevity in the following we label it as the loss associated to the contract.

Table 1.3 reports the fraction of contracts which end up with a loss and the size of that loss. A relevant number of contracts is characterized by losses and the average loss

Table 1.3: Summary statistics on losses by contract type

Type of contract	**Fraction of loss (%)**	**Size of loss (%)**
Carnet	41.3	37.47
One month	28.2	37.34
Two or three months	46.7	34.64
Academic flat	46.9	41.94
Flat 1 euro promotion	44.5	35.89
Overall	**32.8**	**36.72**

amounts to 36.7% of the price of the contract which corresponds to 8% of the average monthly budget of an average user. Moreover, the vast majority of losses derives from low actual attendance. In the case of *flat* contracts this is clearly the only possible source of losses but it is interesting to notice that is also the case for *carnets*. Users incurring in a loss with this type of contract are, in 92% of the cases, attending too few, making the *pay as you go* a cheaper contract.

Possible explanations for these facts are: i) users do not have self control problems but a majority of them has faced unexpected shocks affecting sport activity; ii) the population is heterogeneous and a subgroup of user is overestimating the future number of visits or is buying *flat* contracts and *carnet* to commit to attend, iii) a mix of i) and ii). In the following we will discuss which of these alternative explanations can be considered the most consistent with the data.

As pointed out above, a population of users without self control problems must be able to anticipate correctly future attendance and no systematic relation should exist *ex-post* between foreseeable costs and the probability of incurring in a loss. On the other side, in a population where agents do have self control problems, both in the case where expectations are biased (case of *naivete*) and where expectations are correct (case of *sophistication)*, then this relation should emerge.

To verify empirically this prediction we estimate a reduced form model for the amount of the loss implied by the per visit cost of contracts. As stated above, we compute a relative measure of sub optimality (this is to avoid the spurious effect of having higher losses due to a higher price of the contract) and this variable ranges from 0 (when the contract has the lowest per visit cost meaning that the user chose correctly) to 100 that is when the user buys a contract and never attend, which occurs very rarely in our sample. If we consider the overall contract sample, out of 7414 contracts only 66 have zero attendance (0.009% of the cases). Since 67% of the contracts are not sub optimal we have an inflation of zero observations in our sample and we estimate a left censored tobit model pooling the contracts and controlling for the characteristics of the users . Given that users typically buy more than one contract (on average they buy 3 contracts), we run a cluster robust regression (clustering is made at user level). Letting $y_{k,i}$ be the loss in contract k signed by user i we estimate the following tobit model:

$$y_{k,i} = max(0, y^*_{k,i})$$

where

$$y^*_{k,i} = C_{k,i}\beta_1 + P_i\beta_2 + S_k\beta_3 + \varepsilon_i \qquad (1.11)$$

Vector $C_{k,i}$ contains the observable characteristics of contracts: a dummy that takes value 1 if the contract is the first in the contractual history of the user, a dummy that takes value 1 if the contract is *flat* and 0 if it is a *carnet* and finally the contract length. The vector P_i contains characteristics of the user like gender, age, distance from the gym, a dummy for financial aid, academic field of study and indicators of academic performance. Finally, S_k contains variables related to exogenous conditions as weather (fraction of rain days in the contract) and mean temperature during the contract. Table 1.4 reports the results of the estimation. In columns (A) and (B) all contracts are included in the regression and the inexperience that users have at the very beginning of their contractual history is captured by the dummy for the first contract. Since this can not be enough to control for the fact that at the first experience, expectations are not well formed, in columns (C) and (D) we report the estimates excluding all first contracts[13]. The first columns of both couples reports the

[13]Users generally start with a short trial period, given by the possibility to exploit two free entrances, before being asked to pay to attend. Even if these two entrances are informative about the value of a visit, it can be argued that it is not enough to understand the issue of implementing a planned activity. Hence, we run both the regressions taking the first contract as the period where users learn about costs and benefits related to exercise in the gym.

marginal effect computed at the mean of the regressor for the uncensored dependent while the second column of both couples reports the marginal effect on the probability of being censored and uncensored.

Starting from the characteristics of the contracts, the estimates show that the first contract is associated with higher losses. This is in line with the idea that users at this stage do not know exactly their willingness to exercise and have difficulties to foresee their future attendance. As an example, on average, the loss of a three months contract increases by 4,50 euro in absolute term if the user is at her first contract. The length of the contract is also a determinant of higher losses and can be interpreted as the effect of uncertainty in forecasting the number of visits in a longer time span. As an example, the effect of 30 more days on the amount of the expected loss is to increase it by 0.9 percentage points. Interestingly the amount of the loss is not significantly affected by choosing a *flat* contract meaning that errors are made also by those who choose a *carnet*.

Even if the cost of reaching the gym is known at the beginning, distance has a non negligible influence on the loss. On average a user who is 3 kilometer far from the gym is experiencing higher losses by 2.7 percentage points with respect to a user who is in the close nearby of it. This finding is reinforced by a probit estimation we run on the probability of incurring in a loss. From this regression we see that the effect of an additional kilometer is to increase the probability of incurring in a loss by 2.4 percentage points. We also find a significant gender effect with males incurring in lower losses. Students who receive financial aid tend to incur in lower losses by 2.4 percentage points. Looking at the academic background dummies we find that some are related to lower losses with respect to the baseline given by humanities. Users are grouped in the following fields: humanities (literature, psychology, sociology, foreign languages and law), social science (political science and economics), mathematical science (engineer, mathematics, physics, chemistry and biology), medicine and physical education. The last group and students with a mathematical background are less likely to incur in a loss. The first finding is not surprising since physical instructors on average have a very high monthly attendance and coherently choose flat contracts. The second group contains engineers, mathematicians, physicist and other natural science related fields. This relation may suggest a higher ability in computing the implicit price of the different options at the moment of choosing the contract. Being a good student in terms of grades is not affecting the loss while being a diligent student, which means to meet academic deadlines and to cope with the assigned workload, is associated to a lower expected loss.

Finally, we observe a strong effect of the variables related to the temperature. More precisely, higher losses are associated to contracts in the cold season and in the summer. The polynomial in the mean contract temperature has its minimum at 12.5°C which is the

Table 1.4: Tobit regression on the percentage loss in the contract

Dep. variable: amount of the percentage loss				
	(A)	(B)	(C)	(D)
First contract (d)	3.771***	0.094***		
	(0.672)	(0.016)		
Contract days	0.031***	0.001***	0.020**	0.001**
	(0.007)	(0.000)	(0.008)	(0.000)
Flat (d)	-1.271	-0.032	-1.307	-0.033
	(1.506)	(0.038)	(1.656)	(0.041)
Gender (d)	-2.427***	-0.061***	-2.631***	-0.066***
	(0.734)	(0.019)	(0.916)	(0.023)
Age	-0.052	-0.001	0.008	0.000
	(0.153)	(0.004)	(0.181)	(0.005)
Distance	0.849***	0.021***	0.933***	0.023***
	(0.270)	(0.007)	(0.312)	(0.008)
Financial aid (d)	-1.419**	-0.036**	-1.222	-0.031
	(0.715)	(0.018)	(0.860)	(0.022)
Medicine (d)	0.245	0.006	1.056	0.026
	(1.193)	(0.030)	(1.493)	(0.037)
Social science (d)	1.446	0.036	0.959	0.024
	(1.087)	(0.027)	(1.321)	(0.033)
Mathematical science (d)	-1.978**	-0.050**	-1.845*	-0.046*
	(0.837)	(0.021)	(1.017)	(0.025)
Physical science (d)	-7.782***	-0.193***	-6.636**	-0.161**
	(2.600)	(0.058)	(3.271)	(0.072)
Grade avg.	-0.641	-0.016	-0.846	-0.021
	(0.727)	(0.018)	(0.872)	(0.022)
Credits	-2.044**	-0.052**	-1.938**	-0.049**
	(0.859)	(0.022)	(0.967)	(0.024)
Rel. rain days	3.908	0.099	3.047	0.076
	(2.810)	(0.071)	(3.300)	(0.083)
Mean temperature	-0.647***	-0.016***	-0.727***	-0.018***
	(0.193)	(0.005)	(0.212)	(0.005)
Mean temperature^2	0.026***	0.001***	0.027***	0.001***
	(0.007)	(0.000)	(0.007)	(0.000)
N	4337	4337	3149	3149
pseudo R^2	.00988	.00988	.00732	.00732

Marginal effects; Robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

average temperature in April in Bologna. Periods of the year with higher and lower mean temperature are characterized by higher losses. As an example, if we consider a cold month, like December, when the mean temperature is 4°C, losses increase by 1.8 percentage points. The same apply for the summer months. In June, when the mean temperature rises to 25°C, losses increase by 3.7 percentage points. This finding is coherent with the presence of costs related to exercise which are not fully anticipated by users at the moment of signing the contract but arise at the moment of going to the gym to exercise. At the same time, the relative number of rain days is not affecting the expected loss even if we see that it has a non negligible effect on the number of monthly visits[14].

The findings reported on the effect of distance matches the predictions we had on a heterogeneous population of users. Indeed, we observe that *ex-ante* this cost factor do not reduce the likelihood of signing flat contracts and *ex-post* is associated with losses. This is not compatible with the predictions of a rational sorting model without self control problems. Given that, the next step of the analysis is the identification of types of users in the population.

1.4.1 Identification of types

So far we have shown that the observed behaviour is not compatible with an homogeneous population of *rational* users. This is also the final claim of DVM. Given this observation, we scrutinize the two alternatives proposed by a self control model to rationalize losses. The question we want to answer to is: besides *rational* users, are those who incur in losses *naive* or *sophisticated*?

To discriminate among these two possibilities we use the information we gather on expectations about future attendance. It is known that eliciting expectation is not a trivial issue and can be subjected to some critiques. In laboratory experiments however, especially when the elicitation is made in a incentive compatible way, data on expectations are widely accepted.

[14]This claim builds on a regression on the monthly attendance of users as a function of the exogenous conditions. More precisely, we have estimated the following regression by OLS:

$$monthly\,attendance = \beta_0 + \beta_1 mean\,temperature + \beta_2 mean\,temperature^2 + \beta_3 rel.raindays + controls$$

The relative rain days (measured as the fraction of days in the contract with a rain precipitation higher than 2mm per square meter) in the contract have a positive effect on the number of visits. This is counterintuitive since costs of attending should increase with rain. A sound explanation for this finding is that alternative occupation, that represent a temptation for the user, become less attractive when the weather is bad lowering the self control cost of exercising (for more on this see Nardotto 2011).

Elicitation of beliefs (in our case concerning the expected attendance) is complicated because, as we know from experimental economics, expectations cannot be elicited by making a payment conditional to actual behavior since this would induce *ex-post* compliance. Another difficulty comes from the elicitation technique. This element is critical since the quality of the measure is sensitive to the way the question is posed to subjects. There is a vast literature on this issue and we do not discuss it here. Considering all these elements, the strategy we opted for was to ask different questions in the way suggested by the literature and then check the consistency of the measures we get (more information on the elicitation technique is in appendix 1.9.2).

We can rely then on this additional information for the 50% of the population from 2009 to 2010, that is a sample of 351 users[15]. Given elicited beliefs, we first investigate the forecast error of users which is defined as the difference between expected and actual attendance divided by the expected attendance. The average forecast error in the population is positive and amounts to 30% which means that, on average, a user who expects to visit the gym ten times in a month actually attends seven times.

Coherently with the presence of users who are not perfectly aware of their self control problems, expectations are upward biased. To divide users between optimistic, pessimistic and users with correct expectations we must define a threshold in the forecast error. We reasonably consider an absolute value of 25% which implies that a user who expects to attend twice a week (roughly eight times in a month) ends up missing this target half of the weeks (if she exercises once in each week) or in one out four (if she does not exercise at all in that week). We claim that this band is large enough to account for unexpected contingencies affecting the sport activity and users who are above the positive threshold are optimistic while those below the negative threshold are pessimistic. In figure 1.1 we report the histogram of the forecast error and the two threshold we set to define pessimistic and optimistic users. Users who correctly anticipate future attendance are 44.9% of the total population, users with optimistic expectations are 51.1% and pessimistic users are only 4%. The last figure is interesting since it shows that agents can also be pessimistic about future behaviour when this depends on their ability to stick to a plan. However, based on this figure, it would be difficult to support a claim against the focus on overconfidence since

[15] A typical cohort in the gym is around 700 users. The number of users who participated to our activities is 405 (the activities have been made available in two periods of the year: from 1st September 2009 to 16th March 2010 and from 1st May 2010 to 31st July 2010.) and we consider 351 of them since we drop those who bought specific contracts and those who were not active in the gym at the moment of answering to our questions. We agreed with the gym that every user who bought *PPV* or any type of contract from September 2009 onward was eligible to participate to the activities so some users filled the survey even if they were not active in the gym. In this case we can use some information in the survey but not the one on expectations on future behavior.

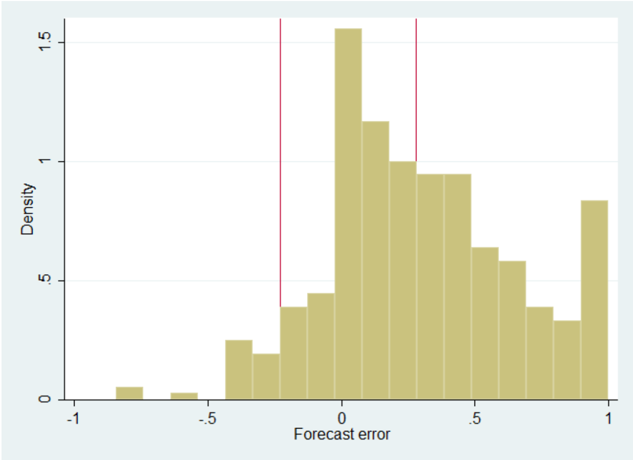


Figure 1.1: Forecast error in the population

4% is a very small fraction of the population when compared to the share of users with too optimistic beliefs.

The next step is to match the information on expectations with the cost of attending. This allow us to identify in the population the categories we have defined in section 1.2. In this analysis we focus mainly on users with *flat* contracts and *carnet* since only three users were attending buying *PPV*. The model implies that *rational* users have correct expectations on future attendance and select contracts such that on average they do not incur in losses; conversely, *naive* users have optimistic beliefs and can end up with losses depending on the realization of costs; finally, *sophisticated* users have correct beliefs and due to self control costs they incur in losses. As table 1.5 shows, in our sample the fraction of *rational* users is 40.6% while *sophisticated* users amounts to only 4.3% of the population. The percentage of *naive* users is 51.1%. Part of them incur in losses (23.1% of the sample) while the remaining part are overconfident but their attendance is high enough to avoid losses[16]. In this group we also find the three users who attend the gym buying *PPV*. Without a commitment device like the *flat* contract, they expect to attend a lot in the future but end up with very few visits, as predicted by the model.

Expectations about future attendance and foreseeable costs. The discussion of the role of foreseeable costs made so far has shown that users on one hand sort in the menu without taking into account properly the self control cost of distance and on the other hand they experience *ex-post* losses related to this cost factor. Another direct proof for this bias comes

[16]This categorization is performed on 325 users who buy contracts that can be evaluated in terms of losses. In the rest of the paper, when evaluating the precision of expectations on future attendance and not the loss experienced by the users, we use the whole sample of 351 users.

Table 1.5: Types of users

	No ex-post loss	**Ex-post loss**
Expectations' group	Rel. freq. (%)	Rel. freq. (%)
Pessimistic	2.2%	1.8%
Correct beliefs	40.6% *(rational)*	4.3% *(sophisticated)*
Optimistic	28.0% *(naive)*	23.1% *(naive)*

from the analysis of expectations. Table 1.6 reports the estimated coefficients of two regressions respectively on the probability of having correct beliefs and on the error in forecasting future attendance. To performe the analysis we exclude pessimistic users and we focus on those who have correct beliefs and on optimistic users. Explanatory variables are the distance and its square, mean temperature and its square, the dummy for the first contract, the contract length and the usual controls (gender, age, academic etc.).

Table 1.6: Regressions on forecast error and probability of having correct beliefs on future attendance

Dep. var:	**Prob. of having correct beliefs**		**Forecast error**	
Distance	0.155*	(0.065)	-0.089*	(0.064)
Distance^2	-0.044**	(0.023)	0.021**	(0.042)
Mean temperature	0.122***	(0.000)	-0.049**	(0.015)
Mean temperature^2	-0.004***	(0.001)	0.001**	(0.029)
Contract days	<-0.000	(0.521)	<-0.000	(0.583)
First contract (d)	0.062	(0.326)	-0.011	(0.792)
Academic controls	YES		YES	
Demog. controls	YES		YES	
$N = 292$, pseudo R^2=0.08359, R^2=0.0845				

Marginal effects; p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We highlight first a regressor which contributes to strength our elicitation method. The variable for the length of the contract is not only not significant but it is practically zero.

Asking different questions to elicit beliefs (attendance in the next 30 days and until the end of the contract) and taking the best answer, we have been able to avoid the problem of heterogeneity in contract length and to obtain an estimate for the next future which is comparable across users. Moving to the main variables of interest, we find that results on the sorting behaviour are confirmed by expectations. The two polynomials in the distance have their maximum (prob of having correct beliefs) and their minimum (forecast error) at 2km from the gym and then forecast errors increase both in likelihood and in the average size. We observe that also the mean temperature has the same effect on beliefs, with a inner maximum (or minimum for the error in forecast) roughly at 15C degrees. When weather conditions are influencing the cost of attending with higher intensity they are not fully anticipated by naive users and expectations are more likely to be upward biased.

Data on expectations confirm the presence of overconfidence about future behaviour. Together with the information on contractual choices they allow us to identify types of users in the population and to discriminate between the different self control models compatible with the presence of losses. In particular, we see that the population is heterogeneous, as predicted by DVM, and losses are largely due to the presence of *naive* users rather than *sophisticated*.

However, this is not necessarily the only explanation for the documented facts. In particular, losses are compatible with two underlying scenarios: influence by the staff and risk aversion, which will be discussed in section 1.4.2.

Are *sophisticated* users sophisticated? The categorization we made so far is intuitive when we consider *rational* and *naive* users since it is based on expectations about future behaviour and the objective to select a contract which minimizes the expense. Using these two elements we identified also *Sophisticated* users who have correct beliefs about future attendance but incur in losses. The self control model rationalizes this behaviour claiming that *sophisticated* users consciously select contracts that foster attendance, such as *flat* contracts, because they provide a commitment to exercise. To find empirical evidence for this claim, we included in the survey some questions regarding the feature of contracts in terms of the incentives to exercise they may provide. A possible confirmation of the identification we made of the group of sophisticated users is to see whether this group recognizes the flat contract as a commitment device. To this end, we employ three general questions regarding *pay per visits*, *carnet* and *flat* contract; and one specific question on the contract chosen by the user. The respondent must evaluate the following sentences and provide a judgment for each of them:

D1: *Pay per visits are useful to foster attendance*

D2: *Carnets are useful to foster attendance*

D3: *Flat contracts are useful to foster attendance*

Answers: 1) No; 2) A bit; 3) Very useful

In general, we see that users correctly recognize flat contracts as commitment devices. The proportion of users who classify the *pay per visit* as very useful or a bit useful is only 15%, while for *carnet* and *flat* contracts is 57% and 83% respectively. It is interesting to notice that if we contrast our groups, then the percentage of *sophisticated* who find *flat* contracts very useful jumps to 93%.

In the survey the same question is also asked with respect to the actual contractual choice of the user:

D4: *The contract I have chosen is useful because it fosters my attendance*

Answers: 1) No; 2) A bit; 3) Very useful

It emerges clearly that groups have a different idea of the *flat* contract as a commitment device. *Sophisticated* users recognize its effect on attendance not only in general but especially in their case. If we aggregate answers 2) and 3) and run a test on the difference in the population mean of sophisticated and optimistic users, then the test does not reject the null hypothesis of a significant difference at 5% (p-value=0.0359) if the test is one-sided (the null hypothesis is that the mean of the sophisticated is higher) and at 10% (p-value=0.0719) if the test is two sided[17]. This confirms the claim that *sophisticated* users are aware of their future behaviour and select flat contracts because they foster attendance.

1.4.2 Alternative explanations for the documented facts

So far we have shown that users often do not select the contract that minimizes the expense of their consumption plan. Moreover, they select *flat* contracts or *carnet* too frequently. A simple explanation for that would be that the staff of the gym is able to make some pressure in order to convince users to select *flat* contracts instead of *PPV*.

To check whether this claim is true, we add a question in the survey asking users whether they select their contract autonomously, with some friends or with the help of the staff. We find that 12.2% of the users ask for suggestions from the staff in selecting their contract while 80% declare to choose autonomously. Hence, users who do look for suggestion are

[17]With the other possible aggregation, that is answers 1) and 2) as outcome 0 and answer 3) as outcome 1 the result of the test is the same. Moreover, a non parametric Wilcoxon test for ordinal variables rejects the null of equal median at 10% (p-value=0.0745)

not a negligible share of the population. Moreover, they can be under representative, since part of those who declare to select autonomously can be in some way influenced as well by the staff. To obtain an estimate of the relevance of the pressure of the staff at least on the former group, we proceed identifying them with a dummy variable and performing a probit regression on the probability to select a *flat* or a *carnet*. Results of the regressions are reported in table 1.14 (see appendix 1.9.3). We find that the staff exerts some pressure to the users who declare to ask for suggestions. Among this group, those who are at the first contract in the gym always select the flat contract to exercise. It is interesting to see however that, after the first contract, the estimate of the staff dummy variable, still has a positive sign but becomes non significantly different from zero (p-value=0.604). Users who are willing to accept suggestion follow the indication of the staff to select *flat* contracts, which can be considered as the default option for a user at first experience. However, when users acquire some experience, they start using other types of contracts and no clear relation emerges between the suggestion of the staff and the actual choice of the user. Since this is the result we get for the group of users who declare to ask for suggestions, it is natural to think that the rest of the population is even less affected by the pressure of the staff. Hence, it would be hard to claim that the pressure of the staff is the main explanation for the observed behaviour, especially after the first contract.

A second explanation for the observed behaviour is that users select *flat* contracts or *carnet* because they are risk averse. We test this claim with a probit regression for the probability of selecting a *flat* contract or a *carnet* where among the regressors we have the risk attitude we elicit through the payed experiment (see section 1.3 on data construction). As in the previous case we employ all the contracts bought by the users for which we have this information, under the assumption of stability of the risk attitude along time. We do not find any significant relation between the risk attitude of the user and the contractual choice. This result is confirmed also after the introduction of other variables like impatience and the financial constraint of the user.

Finally, the analysis we conducted to check whether users are indeed too optimistic about future attendance and to discriminate among types relies on the assumption that we observe their "true" expectations. The elicitation method we used, as pointed out in the previous section, can be subjected to the critique of not being salient. Given that, the estimate of future behaviour may differ with respect to the one that the agent had at the moment of signing the contract[18] because of inattention or impatience to finish the survey. To provide

[18]We do not elicit the expectation at the moment of signing the contract but through the online survey. We prefer to ask this information in a later moment (when the contract is already chosen) not to influence the decision in the gym.

evidence on the quality of the data on expectations we first check whether users, given their beliefs, are choosing a coherent contract in the menu. From this analysis we see that in 85% of the cases there is consistency between expectations and choice. Moreover, this number must be rounded upward (by 4% - 5%) considering sophisticated users who declare a correct belief and chose an apparently inconsistent contract. Another check for the quality of our data is based on two variables that we have measured: the time to fill the survey and the number of errors made by the subjects. Since the survey is made online, we are able to compute the time the user needed to complete it. Moreover, every time a user made an error in filling the survey (possibly because of inattention) we recorded this event (for example, if the user does not complete a page the page is reloaded and the error recorded). The group of users who have correct beliefs about future behavior and the rest of the users (those who are above or below the thresholds) do not have significant difference with respect to the time needed to complete the questionnaire and with respect to the number of errors. The Mann-Whitney-Wilcoxon tests are not rejecting the null of equal means (p-value=0.792 for the former and 0.4021 for the latter). Moreover, if we add the time to fill the survey in the same regression reported in table 1.11, the estimated coefficient is close to zero: -0.0067 and a test for significance does not reject the restriction to zero (p-value=0.721).

1.5 The role of experience: are types stable over time?

The analysis conducted so far shows that in the population of users we can identify types as predicted by a model of self control costs. This implies a natural question: are those types stable? In other words, we are interested in seeing whether users tend to behave in the same way along time or whether they become more able to predict future attendance through the accumulation of experience. Taking seriously the problem of self control, this is probably more interesting than to verify, as we did above, the presence of optimistic or pessimistic beliefs.

In our data, differently to DVM, there is no automatic renewal of the contract after the expiration. This means that every time a user signs a new contract she must rethink about future attendance given a greater experience about her ability to stick to her exercise plan. Hence, our strategy is first to measure experience that an agent has at a given contract as the number of days spanned by previous contracts and see whether contracts signed by more experienced users are less likely to be associated to losses. Since users who do not like the gym environment are more likely to quit, especially after the first contract, a regression on the whole sample of users is likely to produce a biased estimate of the effect of experience whenever experienced users (those who stay) are more likely to derive more utility from exercising than those who leave.

We take care of this restricting the analysis to those who do not drop from the sample and stay for at least two contracts (in the population a user signs on average three contracts). We end up with 965 subjects and 3765 contracts. We estimate the same model as in 1.11 except for the introduction of the experience among the regressors. Results are reported in table 1.7. Most of the coefficients have the same sign and magnitude as in table 1.4

Table 1.7: Tobit regression on the percentage loss in the contract

Dep. variable: amount of the percentage loss			
First contract (d)	-2.7975***	Financial aid (d)	-1.0279
	(.74322)		(.78648)
Contract days	.02938***	Medicine (d)	1.0672
	(.00802)		(1.3192)
Previous exp	-.00522**	Social science (d)	1.4915
	(.00234)		(1.195)
Flat (d)	-.70795	Math science (d)	-1.5471*
	(1.577)		(.91336)
Gender (d)	-2.173***	Physical science (d)	-6.8279**
	(.80939)		(3.2113)
Age	8.7e-05	Grade avg.	-.75123
	(.17335)		(.79205)
Distance	.69294**	Credits	-.0492*
	(.2906)		(.029)
Mean temperature	-.78563***	Rel. rain days	1.7142
	(.19508)		(2.9082)
Mean temperature^2	.02915***		
	(.0066)		
N	3752	Clusters	965
pseudo R^2	.00809		

column (A). This subgroup of the population experiences lower losses in the first contract. This is explained by the fact that once we eliminate the group of users who discovered that the gym is not attracting and left it after the expiration of the first contract, the remaining part of the population is composed by more “enthusiastic” users who at their first approach tend to attend more, incurring on average in lower losses. The estimated coefficient for experience is significant but low. This is evident if we compute the effect on the expected amount of loss due to 180 days of extra experience (which is close to the average number of contractual days of a user in the gym) which amounts to a decrease of 1 percentage points in the expected loss. This finding suggests that users do not gain too much from experience and this is compatible with the presence of stable types in the population.

In order to provide more evidence on this important result, we perform a different analysis. We generate a new variable that accounts for losses in the first two contracts and we include it as an explicative variable in a probit regression for the probability of incurring in a loss in a later contract. We perform this analysis with different specifications and results are reported in table 1.8. Column (A) and (B) refer to the model where the dependent variable

Table 1.8: Probit regression on the probability of a loss in the third or later contract

Dep. variable: probability of incurring in a loss				
	(A)	(B)	(C)	(D)
Previous losses (d)	**.27297*****	**.27994*****	**.23647*****	**.23266*****
	(.04058)	(.04257)	(.03082)	(.03112)
Regressors C	YES	YES	YES	YES
Regressors P	YES	YES	YES	YES
Regressors S	NO	YES	NO	YES
N	615	570	2149	2065
pseudo R^2	.08154	.0996	.068	.07245

is the probability of incurring in a loss in the third contract while in columns (C) and (D) we estimate the model for the probability of incurring in a loss in the third or any later contract. The three groups of regressors we report are those we defined in model 1.11 (a table with the whole estimates can be found in the appendix). As we can see, having experienced a loss in the past is a strong predictor of losses today. The magnitude of the estimate we get changes slightly with the specification of the model and ranges from 23 percentage points to 28 percentage points.

So far we did not use data on expectations to analyze the role of experience. From previous results we expect that types are stable since experience has a small effect in reducing losses and users who experienced losses in the past are more likely to incur in losses in the future. A definite clue about this claim comes from the results of the three regressions reported in table 1.9 (columns A-C). In the first two columns (A and B) we report estimated coefficients of two linear regressions that relate the forecast error and experience. As we said above, we define the forecast error as the difference between the expected number of visits and the actual one, divided by the former. This measure is bounded above to one which is the case of a user who declares some positive number of visits but then never attend.

As we see in column (A), more experienced users have more correct expectation on future attendance. The estimated coefficient is such that 180 days of experience imply a 6 percentage points reduction in the forecast error. However, this result is mostly driven by very naive and inexperienced users who have a forecast error equal to one and are very likely to quit their activity after the expiration of the contract. We exclude this group of users to perform the same regression as in (A) and we report the estimates in (B). In this case the

Table 1.9: Regressions estimates

Dep. Var.	Forecast error		Prob. of having correct expectations	
	(A)	(B)	(C)	(D)
First contract	-.05555	-.01098	.0047	.01464
	(.04966)	(.04561)	(.0684)	(.06941)
Experience days	-.00033**	-.00018	6.4e-05	5.3e-05
	(.00015)	(.00014)	(.00021)	(.00021)
Gender	.02388	.01261	.00877	.02238
	(.03893)	(.03549)	(.05355)	(.05437)
Correct recall				.22616***
				(.05181)
N	351	326	351	351
R^2	.01504	.00727		
pseudo R^2			.0003	.03783

coefficient is lower (the effect of experience on the forecast error is reduced to 2%) and is not significantly different from zero. Finally, in column (C) we report the estimates of a probit regression on the probability of being in the group of users with correct expectations. In this case outliers, i.e. users with very wrong expectations, are less a concern and we get an estimate which is very close to zero and not significant, confirming the claim that experience plays a little role in improving the forecasts of the users.

We conclude the section analyzing a last but important element: the recall of past performance. So far we have collected evidence on the presence of types observing that behaviour does not change much along time (the probability of incurring in a loss is not much affected by experience and users who incur in losses are likely to do so in the future) and expectations (which do not improve sensitively with experience). Hence, coherently with stable types, it is reasonable to expect that users who have wrong expectations about future beliefs are likely to have also some bias in evaluating past performance, while users with correct expectations also have correct beliefs with respect to the past. Following this reasoning, we put together information on expectations on the future and recall of the past.

In the survey we asked users to recall their past attendance in the gym[19] and to state

[19]To elicit beliefs on past attendance we employ two different questions on the performance in the month of activity and in the overall gym history. Depending on the contractual history of the subject, one question or the other has been selected. For details on the elicitation method see appendix 1.9.2.

whether they take it into account when selecting a new contract[20]. For each user we match the categories based on expectations about the future (pessimistic, optimistic and correct forecast) with the same categories relative to the past behaviour[21]. In table 1.10 we report the relative frequencies we observe in the population. First, the statistics clearly show that a

Table 1.10: Groups of users based on expectations and recall

		Recall of past attendance			
		Pessimistic Rel. Freq. (No.)	Correct recall Rel. Freq. (No.)	Optimistic Rel. Freq. (No.)	Total Rel. Freq. (No.)
Expected future attendance	Pessimistic	0.28% (1)	2.28% (8)	2.28% (8)	4.84% (17)
	Correct beliefs	1.42% (5)	29.91% (105)	13.39% (47)	44.73% (157)
	Optimistic	0.57% (2)	22.22% (78)	27.64% (97)	50.43% (177)
	Total	2.28% (8)	54.42% (191)	43.30% (152)	100.00% (351)

bias exists also in the perception of the past. More than 43% of the users overestimate their past performance while 54% recall correctly. Interestingly, 30% of the users have correct beliefs both on the future and on the past attendance while 28% are both overestimating their past performance and have an optimistic view of future attendance. This evident persistence in beliefs emerges clearly in the last probit regression we run for the probability of having correct expectations on future attendance. Estimated coefficients are reported in column (D) of table 1.9. From this regression we see that a strong predictor of correct beliefs is given by having correct recall of the past performance.

From table 1.10 we also see that a relevant share of users, which amounts to 22.22% of the total, have a correct perception of their past attendance but are still too optimistic about the future. This is interesting because it suggests an interpretation that has to do with the update of beliefs. From the data we can argue that *naive* users either do not update their beliefs about the future remaining too optimistic even if they recognize the past performance (78 out of 177, that is 44%) or (with the exception of the negligible minority of users

[20]Past performance is recognized by users as an important determinant of choice. In the survey we have ask whether they evaluate the contract options available in the menu considering past performance and from the answers we see that 58% of the population declares to take into account this element in their process of choice.

[21]We employing the same threshold of 25% as we did in the case of expectations on future attendance.

with a pessimistic recall) even reinforce their wrong belief on future attendance through an optimistic recall of the past one (97 out of 177, that is 55% of *naive* users).

1.6 The relationships between types and traits.

In order to form an expectation about future behaviour and to translate it into a contract that minimizes the total expense or some other objective (such as reaching a minimum number of visits) an agent must think carefully about her future activities and willingness to exercise. This process undoubtedly needs some mental skills and mental resources[22]. Moreover, other than cognitive skills, personality traits can play a big role in determining the behaviour of a user. In order to shed some light on the relationship between cognitive abilities, personality traits and economic outcome, we have elicited these traits in the same population we have analyzed in the previous sections.

The trait elicitation is described in section 1.3 and in the appendix. Briefly, we measure cognitive skills in two ways: through a symbolic test (which is a proxy for IQ); and with a set of questions designed to elicit the ability of the subject in making simple calculations. We label the latter as numeracy skills. With respect to personality traits, we refer to the psychological literature and employ standard questions to elicit “neurocitism” and “conscientiousness”. These traits are in the “Big five” list of cognitive traits which are in turn built on traits that belong to them (see section 1.3).

As a first step we run a Mann-Whitney-Wilcoxon test looking at each trait. With this test we compare the mean scores of two groups: users with correct expectations on future attendance and the group of optimistic and pessimistic users. Among the list of traits we are interested in, the only one which is significantly different in the two groups is the numeracy index which is higher in the group with correct expectations. In this case we find that the test rejects the null of equal median between the two groups at 10% (p-value=0.0699).

As a second step we perform a regression analysis that has as dependent variable firstly the forecast error (columns A-D) and secondly the probability of having correct expectation on future behaviour (columns E-H). We test different specifications, reported in table 1.11 . In column (A) and (E) we report the estimates of regressions on the cognitive skills. In columns (B) and (F) we add to the previous specification the result of the test on impulsivity since this can bias the estimates when it is left in the error term. We focus on personality

[22]Our experience in the gym during the planning phase of the study, showed us how much this is true. Very often users think for minutes about the best contract to select and many of them consult the calendar that the staff provides at the reception of the gym. Moreover, the vast majority exploits the possibility of the free trial entrances and, during this trial period and before selecting the contract, ask for a copy of the contract menu.

Table 1.11: Linear regression on the forecast error and probit regression on the probability of correct beliefs

Dep. Variable	**Forecast error**				**Prob. of correct beliefs**			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
First contract	-.02002	-.00966	-.04779	-.01838	-.00725	-.01471	.02213	.00343
	(.0543)	(.05422)	(.05414)	(.05473)	(.07548)	(.07578)	(.07441)	(.07726)
Experience days	-.00011	-7.9e-05	-.00017	-6.3e-05	-.00012	-.00014	-4.2e-05	-.00014
	(.00017)	(.00017)	(.00017)	(.00017)	(.00024)	(.00024)	(.00023)	(.00024)
Symbolic test	-.00102	-.00191		-.00158	-.00212	-.00147		-.00229
	(.00286)	(.00288)		(.00292)	(.004)	(.00405)		(.00414)
Numeracy	-.06815***	-.0948***		-.09145***	.08097**	.10162**		.09925**
	(.02542)	(.02829)		(.02829)	(.03602)	(.04087)		(.04125)
Impulsivity		.04571**	.00368	.0407*		-.03373	-.00181	-.03417
		(.02178)	(.01933)	(.02196)		(.03073)	(.02646)	(.03118)
Neurocitism			.00477	.00418			.00481	.00554
			(.00545)	(.0054)			(.00743)	(.00754)
Conscientiousness			-.01642**	-.01401*			.0197**	.01957*
			(.00715)	(.00714)			(.00984)	(.01008)
Internal loc			.00808	.01155			-.00979	-.01054
			(.00825)	(.00924)			(.01138)	(.01319)
External loc			-.00137	.00137			.00839	.00772
			(.00635)	(.00728)			(.00874)	(.01039)
N	300	300	312	300	300	300	312	300
R^2	.02988	.0442	.02867	.06269				
pseudo R^2					.01311	.01604	.01233	.02698

traits only in columns (C) and (G) and finally we consider models with both cognitive abilities and personality traits in columns (D) and (H).

The picture we get is consistent among different specifications. Numeracy skills are the most important. They are negatively associated with the forecast error and positively with having a correct forecast, while the proxy for IQ (symbolic test) does not display any correlation with the dependent variable. This finding confirms the negative correlation between a scientific academic background and losses we have found estimating model 1.11 (estimates are reported in table 1.4) and implies that mathematical skills help users to select contracts accordingly to their actual behaviour. Among non cognitive traits, neurocitism do not correlate with the probability of having a correct forecast while conscientiousness does. This is not surprising since conscientiousness is defined as the degree to which a person complains with plans (externally imposed as norms and duties but also internally imposed). The signs of the coefficients we get from the estimation are stable among different specifications and are negatively associated with a higher forecast error and positively with the probability of having correct beliefs. Finally, internal locus of control (the degree to which a person believes her own actions to be determinant to outcomes) and external locus of control (the degree to which a person believes external elements beyond own control to be determinant to outcomes), we find sound coefficients (they are of the same size with opposite sign) but no one is significantly different from zero. Again, we do not find any significant coefficients for experience.

1.7 Conclusions

In this paper we empirically asses the self control model as a robust explanation for the observed behaviour in our market. Here, agents are likely to incur in dynamically inconsistent behaviour due to optimistic beliefs on future consumption. Moreover, showing that optimistic beliefs are not reduced by experience, we argue that the observed persistence along time of contracts characterized by a high per visit cost cannot imputed to random factors but instead to the fact that users can have stable biased beliefs. This is consistent with the presence of stable types of consumers in the population. Finally, we find that users' types are related to cognitive abilities and traits. In particular, numeracy skills are crucial to form correct expectations and to select the best contract option from the available alternatives.

This analysis has the merit of providing a deep examination of the choice process behind observed behaviour in our market. We achieve this goal through a sound implementation of different methodologies of economic analysis which goes in the direction suggested by Manski (2004) to complement data on choice and consumption with data on beliefs. The major weakness of the study is in the representativeness of the subject pool we analyze since it is composed by students and by the fact that beliefs are elicited in a non-incentive compatible way.

The stability of types in the population provides stronger motivations to the recent literature on general models of behaviour that can explain what we observe in this and in many other contexts. See, for example, Benabou and Tirole (2004), Fudemberg and Levine (2011) and Lowestein and O'Donoghue (2005). Moreover, it motivates the possibility to provide to agents tools aimed to help in decisions and in subsequent behaviour. We explore this possibility in a related paper, see Calzolari and Nardotto (2011).

1.8 Variables

Table 1.12: List of variables

Age, gender	Age and gender of the user at the moment of signing the contract.
Distance, distance^2	Linear distance (and distance squared) from the address of the user and gym.
Mean temperature, mean temperature^2	Mean temperature experienced during the contract. Data provided by the regional environmental authority.
Rel. rain days	Relative number of rain days (more then 4mm of rain) experienced during the contract. Data provided by the regional environmental authority.
First contract	Dummy variable for the first contract in the user contractual history.
Tot experience days	Number of days spanned by the contracts signed by the user before the one of interest.
Flat	Dummy variable for a flat contract.
Medicine	Dummy variable that takes value 1 if the student is enrolled in the school of medicine.
Humanities	Dummy variable that takes value 1 if the student is enrolled in the school of Italian literature, foreign languages, history, geography or law.
Social Science	Dummy variable that takes value 1 if the student is enrolled in the school of management and business, economics or political science.
Mathematical Science	Dummy variable that takes value 1 if the student is enrolled in the school of engineering, statistics, biology, chemistry or mathematics.
Physical Science	Dummy variable that takes value 1 if the student is enrolled in the ISEF (school for physical instructors).
Financial aid	Dummy variable that takes value 1 if the student receives financial aid from the university.
Grade avg.	Average grade of the student in his academic history
Credits	Total of academic credits earned by the student in his academic history

1.9 Appendix

1.9.1 Summary statistics

This section collects summary statistics on the population under analysis. Table 1.13 reports the statistics of interest.

Table 1.13: Summary statistics

Gender	**Rel. Freq. (%)**	**Distance**	**Rel. Freq. (%)**
Male	50.85	0 km - 0.5 km	31.33
Female	49.15	0.5 km - 1 km	22.76
Academic background	**Rel. Freq. (%)**	1 km - 1.5 km	12.89
Humanities	35.9	1.5 km - 2 km	10.81
Medicine	8.5	2 km - 3 km	11.29
Social science	14.4	More than 4 km	10.92
Mathematical science	33.7	**Age category**	**Rel. Freq. (%)**
Physical science	0.9	21 or less	29.1
Erasmus	6.0	22 to 25	46.1
Not categorized	0.6	26 to 29	15.5
Experience classes	**Rel. Freq. (%)**	30 or more	9.3
First contract (0 days)	26.58		
1 - 30 days	11.49		
30 to 90 days	20.15		
90 - 360 days	32.63		
More than 360 days	9.15		

The population of the gym is equally divided between males and females. Since they are university students the vast majority of them has an age that ranges from 19 to 25 (75% of the sample). The users are located in the nearby of the facility. Only 10% is more than 4km far from the gym. The facility is located in the center of the city and the vast majority of the users reach it by foot or by bicycle. With respect to the academic are of study, the taxonomy we followed is:

Humanities: literature, foreign languages, law, sociology, psychology.

Mathematical science: mathematics, physics, engineering, biology, chemistry

Social science: economics and business, political science

Medicine: medicine, pharmacy

Erasmus: students that are visiting the University of Bologna through the Erasmus program or similar programs

Not categorized: schools that are not specific to a field (e.g. Magistero and SISS are school intended to prepare teachers of the high schools).

1.9.2 Surveys and experiments

This section is devoted to examine the data coming from the survey and the experiments in more details.

Beliefs elicitation To gather this information on expected future attendance we employ different questions in the survey depending on the type of contract signed by the user. These cases are possible:

- *PPV* or *carnet*

D 1: How many times will you attend the gym in the next 30 days?

D 2: Probability table on the next 30 days (see below)

- *Flat* contract

D 1: How many times will you attend from until the end of the contract?

D 2: Probability table on the next 30 days (see below)

The first question in both cases requires a punctual estimate of future activity in the gym. The second question requires to put probabilities to different possible outcomes. In figure 1.2 we report a screenshot of the table. We ask directly probability of events and before the table there are instructions and examples to instruct the user. Even if the user is asked to associate probabilities to events, in the examples we present some possible answers both in terms of probabilities and in terms of frequencies. This is suggested by the literature on beliefs elicitation in surveys, as summarized by Manski (2004). The answer is not valid (and the user cannot proceed in the questionnaire) until the amount of probability associated to the possible events is different from 100 percentage points.

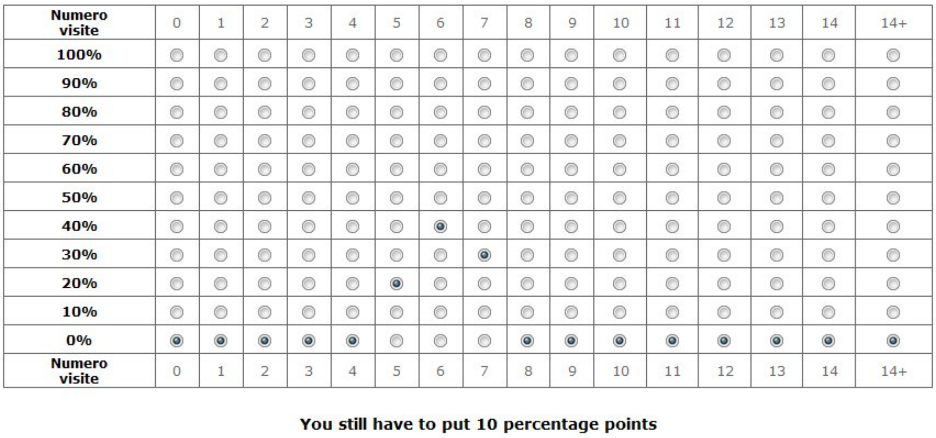


Figure 1.2: Probability table

Perception of past attendance With respect to the perception of past attendance, this is asked to all of them in two ways: the first question asks what is their general past attendance, the second is more specific and ask them what was their attendance in the last month.

Survey on Cognitive and Non Cognitive traits The cognitive abilities we focus on are a general measure of IQ and a measure of numeracy skills. The former test is the same as in Lang, Weiss, Stocker, and Rosenbladt (2007) which has been proved by the authors to be highly correlated with IQ. The test is designed to be run online and it consists in memorizing symbols and inputting them in a small table for 30 seconds. The user has one trial stage to practice the test and then she has to do the three main stages. In figures 1.3 and 1.4 we reported two screenshots of this test.

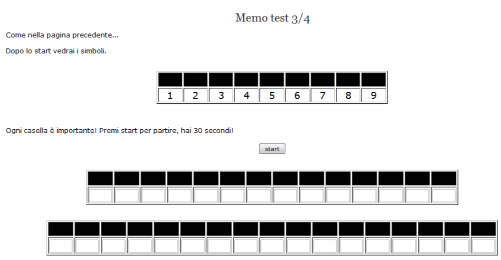


Figure 1.3: Before starting the countdown

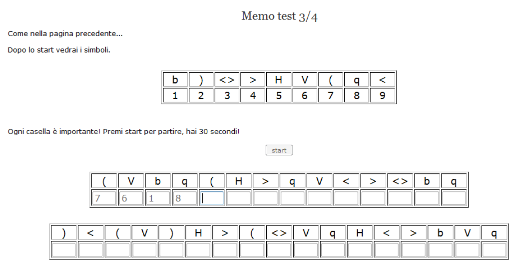


Figure 1.4: After starting the countdown

As we can see from the figures, the user can start the test pressing the start button. After the start she has 30 seconds to fill the cells with the number corresponding to the symbol displayed. When a number is selected the following cell is automatically activated so the user is focused only on remembering the numbers the inputting them matching the symbols. When the time elapses the user is redirected to the next stage.

The program is written in Javascript so it is run locally by the PC. Since it is also very simple, so no slowdown is expected to occur during the test. If the user shuts down the window during the test she has to restart and we record it in our database. Overall, 14 users interrupted the test and they did it at the very beginning (that is during the trial stage) and the following time they run it completely. The rest of the users did the test at the first time without interruptions.

The aggregate measure we construct is the sum of the performance in the three stages and the results we got in the online implementation is comparable with exactly the same test run in the laboratory of the School of Economics in Bologna. The subject pool in this case was composed by undergraduate students taking a course in econometrics. The only difference we observe is that in the population that run the test online we observe an higher non response rate (users who did not input any number and waited that the 30 second elapsed). Out of 312 users who terminated the test, 11 decided not to do it while in the other case we have only one non response. If we exclude these observations, in both cases we do not reject a test for normality (as we expect from a test related to IQ). Moreover, the mean value of the test in the online population is 34.058 while in the laboratory is 33.924 and Mann-Whitney-Wilcoxon test does not rejects the null hypothesis of equal means between the two samples (p-value=0.3858). This is an indirect confirmation of the quality the data collected online with respect to the one collected in the lab.

The numeracy trait is elicited through a battery of questions:

A. If in a population the probability of contracting a disease is 10 percent, how many people out of 1000 will get it?

B. A shop is selling a sofa at half of the original price. If the current price is 150 euro how much is the sofa without the discount?

C. A car shop is selling a second hand car at 6000 euro which is 2/3 of the original price of the car. How much was the car when it was new?

D. Suppose you have 2000 euro in your bank account. The interest rate that the bank is applying to your savings is 10% per year. How much will you have in two years from now?

The impulsivity test is taken from Frederick (2005) and the battery of questions is the same he proposes:

A. A bat and a ball cost 1,10€in total. The bat costs 1,00€more than the ball. How much does the ball cost? ___ cents

B. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ___ minutes

C. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ___ days

To elicit non cognitive traits we employ a standard set of question taken from the German SOEP as in Dohmen, Falk, Huffman, and Sunde (2010). To elicit the following traits *compliance* we ask:

Procrastination: When you have a task to do that you don't like, how much do you procrastinate? (1 to 11 scale)

Internal LOC 1: How my life is depend on me. (1 to 7 scale from I completely don't agree to I completely agree)

Internal LOC 2: One has to work hard to succeed. (1 to 7 scale from I completely don't agree to I completely agree)

Internal LOC 3: If I run up against difficulties in life, I doubt my own abilities. (1 to 7 scale from I completely don't agree to I completely agree)

External LOC 1: What a person achieve in life is is above all a matter of fate or luck. (1 to 7 scale from I completely don't agree to I completely agree)

External LOC 2: I frequently have the experience that other people have a controlling influence over my life. (1 to 7 scale from I completely don't agree to I completely agree)

External LOC 3: Inborn abilities are more important than any efforts one can make. (1 to 7 scale from I completely don't agree to I completely agree)

Conscientiousness 1: I consider myself as someone who does a thorough job. (1 to 7 scale from I completely don't agree to I completely agree)

Conscientiousness 2: I consider myself as someone who does things effectively and efficiently. (1 to 7 scale from I completely don't agree to I completely agree)

Conscientiousness 3: I consider myself as someone who tends to be lazy. (1 to 7 scale from I completely don't agree to I completely agree)

Neurocitism 1: I consider myself as someone who worries a lot. (1 to 7 scale from I completely don't agree to I completely agree)

Neurocitism 2: I consider myself as someone who gets nervous easily. (1 to 7 scale from I completely don't agree to I completely agree)

Neurocitism 3: I consider myself as someone who is relaxed, handle stress well. (1 to 7 scale from I completely don't agree to I completely agree)

1.9.3 Regression tables

In this table we report regression results of the probit model for the probability of selecting a *flat* contract or a *carnet*.

Table 1.14: Probit regression on the probability of choosing a flat contract

dep. var. probability of selecting a flat contract or a carnet					
gender	.49408***	.49023***	.49199***	.54859***	.48852***
	(.18104)	(.17931)	(.17906)	(.16953)	(.18006)
risk attitude	-.02778	-.03425	-.03405		-.0251
	(.03378)	(.03448)	(.03517)		(.03528)
impatience		.00835	.00868		
		(.01417)	(.01341)		
budget constrain			6.1e-05		
			(.00065)		
staff dummy				.40142	.2364
				(.43427)	(.45559)
first contract				.08238	-.00679
				(.13985)	(.15099)
N	868	868	868	1035	868
pseudo R^2	.03473	.03627	.0363	.04831	.03676

risk attitude: is the number of rows in the table before switching to the lottery. The higher is the value the more a user is risk averse.

impatience: is the number of rows in the table before switching to the lower prize at a former closer date. The higher is the value the more a user is patient.

We perform the analysis on all the contracts signed by users who filled our survey since only for them we have this information and we are left with 1035 contracts.

Chapter 2

Nudging with information: a randomized field experiment on reminders and feedback

abstract

Can people be helped to stick to their plans with a little help of information? We provide theoretical and empirical analysis of the effects of reminders and feedback on investment activities that contemplate up-front costs and delayed benefits, such as education and healthy behavior. With a randomized field experiment we show a strong and significant increase of physical exercise by users in a gym induced by simple reminders. We also study individuals' decisions to acquire feedback on their past activity and related effects on attendance. We show that limited attention has a role in explaining our results which however also require an additional behavioral bias that we identify in strategies of mental accounting. We think these results are important since they show that virtuous behavior, such as healthy life style, can be induced with no monetary incentives thus significantly saving on public finances: providing incentives with information is effective and cheap.

KEY WORDS: randomized field experiment; limited memory; inattention; reminders, feedback; mental accounting; sunk cost.

JEL CLASSIFICATION: D03; D11; C93.

2.1 Introduction

When individuals face activities which require immediate costs or effort and deliver benefits only in the future, they tend to under-invest and perform poorly, even as judged by themselves. For example, since education requires up-front effort and costs by pupils and parents but returns possibly higher income in the future, decisions leading to under-education are common often leading to poverty traps. Similarly, healthy life-style requires restraining from some current unhealthy behavior and practicing physical activity, with benefits that are immediately intangible and delayed. Unhealthy behavior and limited physical exercise lead to diabetes, cardiovascular diseases and obesity which the same individuals ex-post regret.

In these and other similar cases, public policies are often advocated on two grounds: first trying to alleviate the negative impacts and treating the consequences of suboptimal behavior, second to prevent poor choices by individuals. With respect to prevention, monetary incentives to induce virtuous behavior have been widely used with strong beneficial effects. For example, in the context of suboptimal physical exercise documented by DellaVigna and Malmendier (2006), Charness and Gneezy (2010) (CG hereafter) have shown how effective monetary incentives can be. In one of their studies (costing \$6000 for eighty individuals), they offered up to \$125 for at least nine visits to a gym over five weeks and observed that after treatment these individuals continued to attend significantly more than what they did before. As for education, the program studied by Angrist and Levy (2009), which cost \$650,000 to increase certification rate at school, turned out to be successful although mainly for girls.[1]

Given the enormous direct and indirect social costs of poor individual behavior on activities such as education and health, it is not surprising that all these studies have proved that the large amount of money put at stake in these programs was indeed worthily employed.[2]

In this paper we instead investigate whether desirable and comparable effects can be obtained motivating individuals with no monetary incentives. In particular, we concentrate on the possibility that providing information to individuals with reminders and feedback on past activities may affect their behavior as desired and explore the underlying mechanisms leading to those incentives. We believe this is important also because providing monetary incentives, although effective, can be very costly for the public finances as the previous figures illustrate and for some countries even not affordable. Providing incentives with the help of information may be instead very cheap especially nowadays with mass information

[1]Fryer (2010) has studied very effective incentive experiments in public schools in four US cities for the 2007-2008 and 2008-2009 school years, distributing a total of \$6.3 million on incentives for inputs rather than outcomes in education. Similar successful programs in the New York City school system has awarded \$600 for each passing grade, the Baltimore City Public School District has paid a bonus of up to \$110 to improve scores on state graduation exams and similar programs in the US award up to \$500 for each exam passed.

[2]In the US, obesity may be responsible for almost \$40 billion of increased medical spending through 2006 and the medical costs of obesity could have risen to \$147 billion per year by 2008.

and telecommunication technologies.

To this end we provide an empirical and theoretical analysis of the role of information on investment activities. We first provide a parsimonious theoretical framework based on a few behavioral assumptions, namely limited memory and a form of mental accounting related to sunk costs, that deliver testable implications on the role of reminders and feedback on investment activities. We then provide related empirical evidence based on a randomized field experiment on a particular investment activity, exercising in a gym, which has been the subject of many recent studies. In this environment, we have monitored contract subscription and attendance to a gym (a sport facility managed by the sport union at the University of Bologna, Italy), before, during and after a treatment which contemplated delivering information to some individuals. In particular, treated users received a weekly E-mail reminding the possibility to exercise in the gym and also to obtain feedback on their own past activity through a personalized web page accessible with personal credentials. Users in the control group instead did not receive any reminder and could not access any personalized web page. Our data on contracts and attendance span the entire population of the gym from January 2008 to August 2010 (967 student users where present in the gym in the Ac. year 2008-2009 and 568 new users in the following year[3]). The experiment is run from September 1st 2009 to March 15th 2010 with 243 users participating (55% of the active population in this time window). Part of these users are not considered in the analysis since they quit the gym activity after the first contract. We focus instead on two samples: first, the one with those who have contract before, during and after the treatment (86 users); second, the enlarged sample of those who purchase contracts before and during the treatment (146 users)[4].

We have unveiled a remarkably strong effect of information which is capable to significantly increase monthly attendance comparing before and during treatment. In particular, splitting the population as in CG between regular and non-regular users (on the basis of mean monthly attendance), we verify that treated users with low attendance exercised 1.95 visit more per-month during treatment as compared with controls, an increase of 25%. This effect is remarkably strong also because our low attendance users are quite active at the baseline exercising more than 7 time per-month which would put them in the group of regular users of CG. After removing the treatment, i.e. when E-mail ceased to be sent, treated and controls behavior converged but still showing larger attendance by treated (a statisti-

[3]Figures on the Ac. year 2008-2009 refer not only to new users joining the gym in this academic year but also those who where already exercising with pre-existing contracts. We start our inquiry from January 2008 since the previous recording system was not reliable and has been replaced by a new one at the end of 2007.

[4]As it is shown in section 2.6, results obtained with the first sample are confirmed considering the second one.

cally significant 20% more comparing before and after treatment for low attendance users), which could be the effect of habit formation as also reported in CG in the case of monetary incentives. We also document with a probit regression that reminders have a strong and highly significant impact in increasing the probability of attending in the next 12-36 hours after receiving the E-mail.[5] Furthermore, by increasing attendance reminders also reduce the probability that individuals "give up" and quit the gym forever.

Why do simple reminders increase attendance and help forming good habits? Indeed, our theoretical model clarifies that fully rational individuals and even present-biased users should not be affected by reminders which instead might have an effect on individuals with limited memory and or using some form of mental accounting based on sunk subscription expenses. We show that considering the period before treatment (all users), the daily probability of attendance within monthly contracts significantly decreases from the first to the fourth week and this is true any time a user subscribes a flat contract. We use our theoretical analysis to show that this finding is not consistent with users uniquely affected by limited memory and we also exploit different reactions to reminders by individuals affected by limited memory and using mental accounting. Overall, we provide evidence that reminders do induce more physical exercise since they increase individuals' attention but also because they rekindle the initial expense to subscribe and attend the gym.

We finally, investigate why individuals decided to acquire feedback on their past activity (available in their personalized web page during treatment) and what effect this had on attendance, on top of reminders. We studied the probability to acquire feedbacks using individual cognitive and non-cognitive traits obtained with online survey and experiments administered to users in our experiment. The probability to acquire feedback is higher for individuals with higher general cognitive abilities but low ability to make simple calculations. The probability is also higher for those who are more conscientious (defined in the psychological literature as the degree to which a person is willing to comply with rules, norms and standards) and patient (measured by time preferences as in Holt and Laury, 2002). Hence, more skilled users do not look for feedback probably because they are able to recall the exact number of their past visits and to compute the implicit cost of their contract. At the same time, cognitively endowed users are more likely to acquire information probably because they receive a gratification in checking that they "did it right". We have also verified that for those individuals acquiring feedback, mental accounting is a strong drive for exercise. Overall, feedback seem to have an effect in inducing more attendance

[5]In all our regressions on attendance we also control for several individual characteristics (such as age, gender and the physical distance between the place of residence and the location of the gym), seasonality and local weather conditions (daily frequency).

which sums up with that of reminders.

Our results are related to the vast theoretical and empirical literature on investment goods. CG were interested in verifying whether after a period of intense exercise induced by monetary incentives healthy habit persisted over time and showed that this is the case for non-regular users (attending very low, i.e. 3 times per week) when observed up to two and half months after treatment. These results have been confirmed by Acland and Levy (2011) who also used monetary incentives and showed habits indeed last long (over eight months after treatment), although decaying over time and being significantly diminished by long vacations. Our results are comparable when considering the effect we document during treatment and observing that our non-regular users are very active at the baseline (7 times per week). What is more, we were able to obtain these comparable effects at no cost and this has important implications for public policy.[6] Another argument made by opponents of monetary incentives is the risk of crowding out of intrinsic motivation which CG documented for regular users. Our findings instead show no such effects, probably because nudging individuals "softly" with information allows them to go on with their pre-treatment decisions. This is another relevant policy implication because targeting monetary incentives to the "right group" may be difficult, whilst information seem to self select as desired.

This paper is also related to a recent and growing literature on limited memory. Karlan et al. (2009) test a theory of limited attention by randomly reminding account holders to make savings deposits and show that reminders increase savings balances by about 6%. Stango and Zinman (2010) and Zwane et al. (2009) have shown that surveys may affect saving and investment behaviors because they act as a shock to attention. With these papers we share the effectiveness of repeated reminders but we differentiate with our analysis on alternative motivations and the role of feedbacks.

Thaler (1999) has analyzed the role of mental accounting which is relevant also in our experiment. Gourville and Soman (1998) with a qualitative analysis have reported a decreasing path for exercise in a small pool of 33 users in a gym, but they used semester-long contracts which may have been affected by seasonality trends. There is also a large psychological literature documenting with real and though experiments that individual seem to be affected by sunk costs (see Arkes 1996 among others). Baliga and Ely (2011) explain the sunk cost effect as a consequence of a memory bias when a decision-maker does not remember the reasons he began a project and the initial sunk cost may thus bring additional

[6] The once and for all cost of setup of the server technology for our dedicated web site is approximately \$1000 and thousand of users could have been addressed.

information about future utility. Our results could be consistent with this idea if our gym users do not remember the benefits they we able to infer at the date of subscription and the sunk cost of subscription brings back this information. Finally feedback have been shown to be effective also in reducing electricity consumption (Sexton et al. 1987).

The paper proceeds as follows. Section 2 develops a simple theoretical model. Section 3 illustrates the data and our experiment. Section 4 presents the main results on the effects of reminders. The ensuing Section 5 illustrates what motivates the decision to acquire feedbacks and its effects. Section 6 concludes.

2.2 Model

In any period $t \in [1, T]$ an individual may attend to an "investment" activity (e.g. exercise in a gym) which involves an immediate cost c_t and an one-period delayed benefit (of better health and fit) $b > c_t$. The cost is stochastic IID with distribution function F and strictly positive density f over the support $[0, \bar{c}]$. At $t = 0$ the individual may purchase a flat rate *subscription* with up-front expense L_T which allows to freely attend in any of the following T periods. Alternatively, she can decide to attend with *pay-per-visit* at unitary price p. The discount factor is normalized to one for simplicity and, for the time being, we assume no cash constraints and risk neutrality. The individual may be present-biased in which case all future payoffs are discounted by the same factor $\beta \leq 1$ and she may be partially aware of this bias so that the perceived bias is $\hat{\beta} \in [\beta, 1]$. As in O'Donoghue and Rabin (1999) the consumer is naive if $\hat{\beta} = 1$ (she thinks herself not having any present-bias), if $\hat{\beta} = \beta$ she is sophisticated (being fully aware of her bias) and she is partially sophisticated if $1 > \hat{\beta} > \beta$.[7]

At any date $t > 0$ the agent observes the cost realization and decides whether to attend, $d_t = 1$, in which case she pays p if she did not subscribe in $t = 0$ or not to attend, $d_t = 0$. Hence, in any t the individual attends if $\beta b \geq c$ when subscribing and if $\beta b \geq c + p$ otherwise. Anticipating a present-bias $\hat{\beta}$ she prefers to subscribe if

$$-L_T + T \times \int_0^{\beta b} (\beta b - c) dF(c) \geq T \times \int_0^{\beta b - p} (\beta b - p - c) dF(c).$$

Notice that an individual not affect by present-bias (i.e. $\beta = 1$) who attends independently of the payment method (i.e. if $\bar{c} \leq b - p$), would thus subscribe whenever

$$A_0 = -L_T + Tp \geq 0, \tag{2.1}$$

[7]We assume that β and $\hat{\beta}$ are time invariant, a result confirmed in our data by Nardotto (2010). For a theoretical model in which the individual learns about her present bias see Ali (2011).

i.e. if she expects a sufficiently high attendance so that by subscribing at date $t = 0$ she expects to save A_0. An analogous expression for monetary saving will be further explored below.

The individual may be *inattentive* and, at any date t, with probability λ the possibility to attend is not salient with respect to alternative activities even if her current net payoff for attending is positive. Furthermore, for a subscribing individual the up-front payment L_T may continue to be salient over the T periods even if it is sunk after $t = 0$. To formalize this possibility and as suggested in the *mental accounting* literature (Thaler, 2004), at date t the individual's preference also contemplates a "transaction" utility $v(A_t)$ weakly increasing and concave with $v(0) = 0$, where A_t is some date-specific measure of mental account. When the individual subscribes, she may open a mental account that considers the up-front payment L_T and may contrast this payment with the alternative of attending with a pay-per-visit. At time of subscription $t = 0$, the expression of this mental account is reported in (2.1), at later dates it could be, for example,

$$A_t = -L_T + pD_t \tag{2.2}$$

where $D_t = \sum_{i=1}^{t-1} d_i$ is past attendance. Since salience of the up-front payment L_T may decrease over time, we posit that $v(A_t)$ depreciates at a factor $\delta \in [0, 1]$.

At any date t an individual is inattentive with probability λ, in which case $d_t = 0$. With probability $1 - \lambda$ she is instead attentive and attends if

$$\beta b - c + \delta^t \Delta v(A_t) \geq 0. \tag{2.3}$$

where

$$\Delta v(A_t) \equiv v(A_t + p) - v(A_t)$$

is the mental accounting effect of attending at t. One possibility is that if $v(A_t) = 0$ for any $A_t \geq 0$, then mental accounting has an effect uniquely when the account is "in the red", i.e. $A_t < 0$.

Reminders and feedback. As in our experimental treatment, any τ periods the individual receives a message that reminds her of the possibility to attend (a Pavlovian stimulus). When the individual receives the reminder at date t (i.e. $r_t = 1$), the individual becomes fully attentive, she observes the cost and decides d_t. However, with no reminders (i.e. $r_t = 0$), the probability of being inattentive in the following period $t' \geq t$ increases back to $\lambda\rho(t' - t)$ where the "retention function" $\rho(\tilde{t} - t) = 1$ for some $0 < \tilde{t} < t + \tau$ and $\rho'(.) > 0$, $\rho(0) = 0$.[8] The stimulus of the reminder may also revamp the salience of the sunk cost L_T.

[8]The decaying effect of the memory stimulus is supported by a large empirical psychological literature

We thus similarly posit that the mental accounting depreciating factor becomes $\delta\rho(t'-t)$. Hence, upon receiving a reminder at date t, the probability of attendance at any date $t' \geq t$ is

$$\Pr(d_{t'} = 1 | r_t = 1) = [1 - \lambda\rho(t'-t)] \times F[\beta b + \delta^t \rho(t'-t)^t \Delta v(A_t)]. \qquad (2.4)$$

At any date t an attentive individual has also the possibility to acquire a *feedback* on past activity which specifies total past attendance since subscription $D_t = \sum_{i=1}^{t-1} d_i$, and the implicit single-visit expense associated with the current subscription in case she attends with the same frequency $D_t/(t-1)$ of the first t periods.

Testable implications. We now state some testable implications which we will consider in the ensuing empirical analysis. The change in the probability of attending at date t' having received a reminder at date $t \leq t'$ is

$$Pr(d_{t'} = 1 | r_t = 1) - \Pr(d_{t'} = 1 | r_t = 0) =$$

$$\lambda[1 - \rho(t'-t)] \times F[\beta b + \delta^t \Delta v(A_t)] + [1 - \lambda\rho(t'-t)] \times \int_{\beta b + \delta^t \Delta v(A_t)}^{\beta b + \delta^t \rho(t'-t) \Delta v(A_t)} dF(c)$$

The first term is due to limited memory (it vanishes when $\lambda = 0$) and the second to mental accounting (it vanishes when $\delta = 0$).

(Implication 1) "*In a population with inattentive individuals (i.e.* $\lambda > 0$) and /or *with individuals with mental accounting (i.e.* $\delta > 0$*), reminders increase attendance in future periods. Instead, attendance of fully attentive individuals with no mental accounting is unaffected by reminders."*

Even if reminders do increase attendance, the previous implication does not allow to disentangle the two motives, i.e. inattention and mental accounting. To address this issue we first notice the following which can be seen from the definition of the probability to attend (2.4) by setting $\rho(t'-t) = 1$.

(Implication 2) *"With no reminders, the probability to attend at any date t is constant in t for inattentive individuals with no mental accounting (i.e.* $\lambda > 0, \delta = 0$*) and may decrease in t for attentive individuals with mental accounting (i.e.* $\lambda = 0, \delta > 0$*)."*

When $\lambda = 0, \delta > 0$ the probability to attend at t becomes

$$\Pr(d_t = 1) = F[\beta b + \delta^t \Delta v(A_t)]$$

and, if $\Delta v(A_t)$ is either decreasing (as for example with a mental balance as in (2.2)) or constant in t, then the larger is t the smaller is salience of the mental account and the probability of attending is indeed decreasing in t due to mental accounting.

---

suggesting that a good fit for the function $\rho(.)$, also known as the "retention function", is a power function with an asymptote to some stable probability of attention, here λ. See Rubin and Wenzel (1996) among others.

Also the effect of reminders is different with limited attention or mental accounting.

(Implication 3) *"For an inattentive individual with no mental accounting (i.e. $\lambda > 0, \delta = 0$), a reminder at t increases (the probability of) attendance in future periods $t' \leq T$ at a per-period rate which is non-decreasing in t'."*

When $\lambda > 0$ and $\delta = 0$ and the individual receives a reminder at date t, the probability of attending at any of the future dates $t' \leq \tau$ is $[1 - \lambda\rho(dt)] \times F(\beta b)$ where $dt = t' - t$, so that the increase of (expected) attendance induced by the reminder (before receiving the next reminder) is

$$\lambda F(\beta b) \sum_{dt=0}^{\tau} [1 - \rho(dt)]$$

which is time invariant. Here we have assumed the stimulus of the reminder decays and vanishes within τ periods. This may not be the case with high frequency of reminders (i.e. τ is small) or more persistent stimulus. In these cases we should observe that receiving a reminder increases attendance over time. This is different in case of mental accounting. Consider, for simplicity, the extreme case of a full attentive individual (i.e. $\lambda = 0$) so that

$$\Pr(d_{t'} = 1 | r_t = 1) - \Pr(d_{t'} = 1 | r_t = 0) = [1 - \rho(dt)] \times \int_{\beta b + \delta^{t'} \Delta v(A_{t'})}^{\beta b + \delta^{t'} \rho(dt) \Delta v(A_{t'})} dF(c).$$

Now $\Delta v(A_t)$ may be constant or decreasing over time, as with the mental accounting balance in (2.2) where one checks period after period if past attendance D_t "justifies" the initial expense L_T. The cross derivative of the previous expression with respect to t' and $\rho(dt)$ is negative: the larger is t' the smaller is the effect of the stimulus $\rho(dt)$ of the reminder.

(Implication 4) "*For an individual with mental accounting (i.e. $\delta > 0$), a reminder at t increases (the probability of) attendance in future periods $t' \leq T$ at a per-period rate which* may *decrease in t'."*

There are also types of balance for mental accounting which instead imply an effect of reminders which increases over time. This is the case, for example, when the individual also considers future attendance in the balance and this is determined extrapolating from past attendance. Following a sequence of periods with no attendance, A_t sharply reduces and $\Delta v(A_t)$ increases as well as the probability of attending.

Before concluding this section we notice that although mental accounting is intuitively related to the possible regret faced by an individual after subscribing, one can show that a direct application of regret theory first developed by Loomes and Sugden (1982) does not allow to identify a specific effect based on this different behavioral assumption.

2.3 The field experiment

This study is conducted in a gymnasium owned by a sport association for university students in Bologna (Italy). This gym is not strictly reserved to students but they are the vast majority of costumers (80% of the total) since the association has the objective to promote sport activities among university students. Our analysis focuses on the students sub-sample. The contracts offered to users by the gym are customary in this industry. The menu of contracts contains subscription flat-rate contracts of different durations: one, two, three months and for the entire academic year (respectively sold at: 45€, 85€, 118€ and 270€); a pay-per-visit ticket (at 6€) and carnet of visits (10 or 20 entrances, respectively at 50€ and 90€). Fees are roughly 30% lower than average market fees (because of subsidies from the university), but are still relevant for a student: the price for a monthly contract amounts to 22% of the typical monthly disposable income of a student (200€, net of the cost of board and lodging). Differently from other gyms (DellaVigna and Malmendier, 2006, for example), in our case there is no automatic renewal of contracts so that users who discover they do not like the activity, are very likely to quit immediately after the first contract.

For each student purchasing any right to enter the gym in the period January 2008 - August 2010, we observe all choices in the menu of contracts and actual attendance recorded at the gym entrance with an electronic key.[9]

Starting from September 2009 we conducted our field inquiry. We recruited students advertising remunerated activities to be performed on a dedicated web-site by sending E-mail to the gym mailing list composed by all students who ever purchased any right to enter and distributing posters and fliers at the gym. Remunerated activities were available to all participants just once and before proceeding to any further step.[10] More than half (55%) of the population of the gym registered at our dedicated web site thus leading to 243 participants.

At the moment of registration on the site, each user had been assigned to a treated group

[9]Some contracts allow students to both attend all courses organized by the gym and also use the gym facilities, others discounted contracts only allow to use facilities. The performance of the gym and its staff is monitored by the sport association which pays staff salaries and guarantees the staff's incentives to monitor entrance and the use of the electronic key.

[10]These activities consisted in (i) a survey with information on sport at the gym and in general, and also students' life, (ii) some cognitive and non-cognitive tests, (iii) an experiment on risk and time preferences (Holt and Laury, 2002), also performed online. All participants were paid 5€ for completing activities (i)-(ii). In activity (iii) one out of five participants received a payment based on actual choices which ranged from 10 to 80€, with an average of 25€. The technology we adopted for the online implementation of these activities is php/MySQL. In interactive sections we have designed a dynamic interface in Javascript. Results related to these activities are reported in Nardotto (2010).

or to a control group (8 subjects out of 10 have been assigned to the treated group). Each treated user received a *weekly E-mail* containing no individual information but reminding the possibility to attend the gym and to browse a personalized web page displaying (i) the number of visits made since the beginning of the current contract and (ii) the implicit final expense of each single entry at the gym expecting the same past attendance until the end of the contract (the price of the alternative pay-per-visit was also displayed).[11] Reminders have been sent once a week in different days (generally on Sunday or Wednesday, see the appendix for the exact deliver dates and the text of the message) and they neither reported information on other users nor the list of recipients.

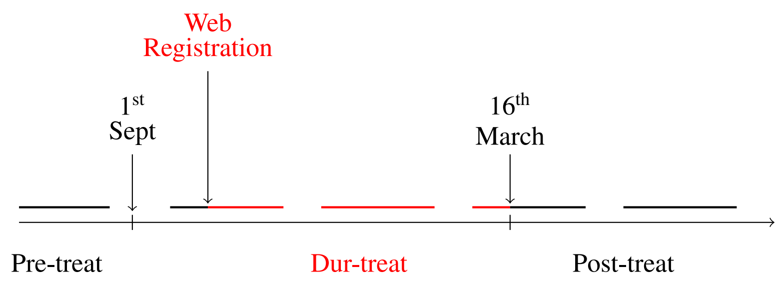
Users in the control group did not receive any reminder and could not access the personalized web page (they could only participate the activities described in note 2.3 but just once and for all).

The timing of the experiment has been the following. In the first week of September 2009 our dedicated web-site became accessible and users could register. The closing date of the site was unknown to the gym users and set at 2010 March 16$^{\text{th}}$, well before the end of the academic year (in Italy end of July) so that we could collect data on users' performance also after the end of the treatment. Since from July 31$^{\text{st}}$ the gym closes for one month summer break, our data span from beginning of September 2009 to end of July 2010. We also have data on individual contract choices and actual attendance for the entire population also in the previous academic year (September 2008-July 2009). Since, as we will explain later, a large part of the population is quitting the activity immediately after the first contract, and we are interested in comparing performance before and after the treatment, we increase the sample size of the control group adding those user of the previous academic year who have a long contractual record (i.e. they subscribe in equivalent period translated one year before). The inclusion of these users in the counterfactual for our treatment is based on the following grounds: first, pre-treatment attendance of these users is the very close to the one of the treated group; second, we see that users of the academic year 2009-2010 with long contractual record are very likely to participate in the field experiment so we consider users of the previous year with the same characteristic as a good counterfactual.

Figure 2.1 illustrates the timeline of the treatment for a user who subscribed contracts before, during and after treatment. Each line represents the time span of a given subscription

[11] Participants could access the personalized web-page at any time logging in with individual credentials set-up at registration. Treated users had also the possibility to use a calendar tool to plan days of future attendance and check ex-post actually attendance. This tool received little attention (only 15 users). They could also use a search engine to find the least costly solution to attend the gym based on expected attendance but also this tool received little attention. Finally, they could unsubscribe the weekly E-mail and only two users did so.

Figure 2.1: Timeline of the experiment



of this user: bold lines refer to days of the subscription in which the user did not receive reminders, light lines refer to periods in which she received the weekly E-mails.

Summarizing, among the pool of 243 individuals registering our web-site, 103 users purchased at least a subscription before and during the treatment period, and 52 of these also purchased a subscription after the treatment has been lifted[12]. As we said above, add to the control group users of the previous academic year who have purchased contracts in the same periods (before, during and after treatment) one year ahead. Hence, the different sample dimension, both in terms of users and in terms of contracts is reported in table 2.1.

Table 2.1: Individuals registering the web-site by contract subscriptions

	Total users (contracts)	Treated
With contracts before, during and after treatment	86 (595)	43
With contracts at least before & during treatment	151 (771)	79
Total registered on the web-site	243	

2.4 Reminders increase attendance

We now compare monthly attendance of treated and control users in the three periods, before, during and after the treatment. Furthermore, we first restrict the analysis on the sub-sample of those 86 users who actually purchased a subscription in *each* of the three periods

[12]The difference between the number of users who registered into the web site and bought contracts before, during and after the treatment is due to two reasons. First, more than 40% of users sign one contract in the gym and then quit the activity (some of them to do some other activity promoted by the University Sport Association); second, since in between the before treatment period and the during treatment period there is the end of the academic year and the consequent summer break, after the summer break many students do not appear again in the gym.

(before, during and after) so that referring to those "*non-quitters*" we can investigate to effects of reminders at the intensive margin in a strict sense. Then, we will analyze whether reminders had also an effect on the decision to quit the gym after treatment, thus verifying the effect at the extensive margin. As we will discuss in the sequel (Section 2.6), our results on reminders and attendance are confirmed with the larger sample of users which contemplates also those who subscribed contracts only before- and during- but not after-treatment.

Results on attendance for non-quitters are reported in table 2.2 which is divided in two parts. The upper part reports average attendance of both groups, treated and controls. In

Table 2.2: Monthly attendance of treated and control users

		Before treatment	During treatment	After treatment	Change Pre-Dur.	Change Pre-Post	N. users
Treated group		9.41	9.51	9.45	0.10	0.04	43
Controls		9.52	8.55	8.66	-0.97	-0.86	43
Diff. treat - controls (pooled)		-0.11	0.96	0.79	**1.07*** (0.0525)	0.9 (0.1220)	
Treated	High att.	11.14	10.35	10	-0.79	-1.14	23
	Low att.	7.42	8.54	8.8	**1.12**	**1.38**	20
Controls	High att.	11.21	10.08	9.78	-1.13	-1.43	22
	Low att.	7.75	6.95	7.48	**-0.8**	**-0.27**	21
Diff. treat - controls (low attendance)		-0.33	1.59	1.32	**1.92**** (0.0306)	1.65 (0.1185)	

the lower part both groups are split between high- and low-attendance users where this distinction is obtained using the threshold of the average attendance in the period before-treatment.

Before treatment, controls and treated users display a very similar behavior in terms of monthly attendance. On average they attend the gym 9.5 times per month and statistical tests do not reject the null of equal mean (Wilcoxon test, p-value=0.57, two sided t-test, p-value=0.84). Notice that these users are quite active as compared with previous studies. We then compare the change of behavior during treatment and with respect to pre-treatment, of both treated and controls and observe that treated attendance increases and instead that of controls decreases. This shows a difference of attendance of 1 visit per month in the treatment period. The decline of attendance of the control group during treatment is due to seasonality since months during treatment are characterized by cold winter weather which is associated with lower attendance (we precisely study the effect of temperature below). However, this is not the case for treated users, who instead increase their attendance during treatment. The different behavior of treated and controls (fourth column lines two and three) is statistically significant (Wilcoxon test, p-value=0.0651, one-sided t-test, p-value=0.0525).

When reminders ceased to be sent, i.e. after treatment, the two groups converged towards more similar monthly attendance. A difference in attendance is still observed with treated again attending more than controls (the third column), but it is not statistically significant (Wilcoxon test, p-value=0.28, two sided t-test, p-value=0.24). With this respect however notice that we are currently pooling all subscriptions from the day after we ceased treatment (March 16$^{\text{th}}$) to the closure day of the gym (end of July), i.e. a long period of four and half months. It is then possible that treatment had an effect in building habits as reported in CG, thus explaining the difference which still is observed after treatment between controls and treated, but this habit fade out in the following months thus explains why this difference is not significant. In fact, if we consider a shorter period after treatment of two and half months (as in CG), then the (non reported) difference comparing during and after treatment for treated and controls becomes higher and significant (one sided t-test).

Independently of treatment and periods, on average our user are very active showing more than two visits to the gym per-week. This rate of attendance is significantly higher than in all other studies with and without students (see the Introduction) and thus the effect of reminders that we are documenting is even more striking. For example, CG reported that their monetary incentives had an effect on non-regular users but had no significant effect on regular users who were defined as those attending more than 1.6-2 times per week before treatment, actually even less than our low attendance users.

If we split our sample between low and high attendance users, and concentrate on the former group the effect of our reminders is even stronger and more striking. During treatment and as compared with before treatment, low attendance treated users show an increase of 1.92 visits per month with respect to low attendance controls, roughly an increase of 25% in monthly visits. This difference is significant despite the reduction in the sample size (Wilcoxon test p-value=0.054, one sided t-test, p-value=0.0153, two-sided t-test, p-value=0.03). Hence, as in CG our results on the effect of reminders are mainly driven by low attendance users, although the effect is significant also for the entire population, as shown before.

It is also instructive to compare the effect of reminders on low attendance users before and after treatment, thus excluding the during treatment period (the last column of table 1, bold numbers). The difference in attendance between treated and controls is again statistically significant and amounts to 1.65 more visits per month, thus an increase of more than 20% of attendance which we can attribute to habits persisting even after treatment. This figure is clearly less than what observed in CG who were mainly interested in habit formation observed after treatment and not during-treatment, since they used monetary incentives that obviously affected treated attendance. The difference between our and their figures can be attributed to two facts. First, we use a longer period after treatment (4.5 instead of 2.5-3 months). Second, our low attendance users are much more active than their non-regular users (respectively around 7 and 3 visits per month) which implies that it is much more difficult to induce a persistent change of habits to our users.

The effect of our reminders on attendance are confirmed by a linear regression performed on the contracts purchased by (non-quitting) users. The dependent variable is the number of monthly visits in each contract and the main regressors are the dummy for being treated, the dummy that identifies the contracts that belong to the treatment period (labeled "During") and the interaction between the two. The model also includes characteristics of the user like age, gender, academic background, a dummy for receiving financial aid from the university and variables which are important in determining the monthly attendance, i.e. the distance from home to the gym and the average temperature experienced during the current subscription. Table 2.3 illustrates the estimated coefficients for this model. Since the unit of observation the contract, we employ a cluster robust regression (clusters associated to users). The first column reports the estimated coefficient of the regression performed on the whole sample while the second is restricted to low attendance users.

Estimated coefficients confirm the previous findings. Reminders have a substantial effect on the number of monthly visits to the gym which is stronger in contracts of low attendance users. Indeed, the dummy referring to treated users is not significant by its own but it is instead positive and significantly different from zero when interacted with the dummy of the period in which the treatment was active. It is also interesting to notice that the (linear) distance between dwelling place of users and the gym location reduces monthly visits and that temperature has a hump shaped effect on attendance with an attendance maximizing temperature at 14 degree Celsius which is exactly the mean annual temperature of the city of Bologna where the gym is located (Nardotto 2011b further exploits these results).

As a final check of the effect of reminders on attendance, we estimate the probability of attendance at any calendar date t for user u with the following model

$$\Pr(attendance\ in\ day_{t,u}) = \Phi(reminder_t, D_t, W_t, S_u, C) \tag{2.5}$$

where $reminder_t$ is a dummy equal to one when we consider a day t which is treated by a reminder, D_t is a vector containing monthly dummies and dummies for the day of the week, W_t is a vector of local weather conditions of the day (temperature and rain), S_u is a vector of user's characteristics such as gender, age, distance from home to the gym and academic variables. In order to consider users facing the same incentives to attend as defined by the

Table 2.3: Linear regression for treatment effect on monthly visits

Dep. Var: Monthly visits (clusters on users)				
	Whole sample		Low attendance users	
	coeff.	std err	coeff	std err
Treated	-0.075	(0.558)	0.587	(0.376)
During period	-0.933*	(0.559)	-0.864	(0.609)
Treated * During	1.207*	(0.655)	1.768**	(0.682)
Distance	-0.424**	(0.206)	-0.370***	(0.116)
Mean temperature	0.258***	(0.093)	0.298***	(0.096)
Mean temperature^2	-0.010***	(0.003)	-0.011***	(0.003)
Gender	-0.396	(0.654)	0.190	(0.446)
Age	0.021	(0.149)	0.117	(0.104)
Fin. aid	0.066	(0.691)	-0.354	(0.416)
Academic controls	YES		YES	
Constant	8.406**	(3.657)	3.273	(2.466)
N. of contracts	595		358	
R^2	0.0924		0.1295	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

inequality (2.3), we base our analysis on those subscribing monthly contracts, which are by far the most common choice in the gym's menu (63% of all purchases in the gym, see table 2.11 in the appendix). Since we do not exactly observe when our E-mail is actually read (which depends on how frequently users check their E-mail), we consider a day as treated if we sent the E-mail the day before or the same day but before 12:00 a.m. (the gym is open from 7:00 to 22:00). The strong and highly significant effect of reminders is reported in table 2.4.[13]

All these results confirm that reminders do affect attendance inducing a significant increase in month visits. The theoretical model developed in Section 2.2 clarifies that (Implication 1) fully rational individual and even present-biased users should not be affected by reminders. On the contrary, individuals with limited attention and / or using mental

[13]This analysis shows that reminders have an impact within 36 hours after we sent out our E-mails. However, depending on the technical equipment used by individual to check their E-mail account (e.g. PC or smartphones), the time window in which reminders produce their strongest effects may be different and possibly even shorter. We are currently exploring this possibility.

Table 2.4: Immediate effect of reminders on attendance

Dep. var: probability of a visit				
	Probit (Clusters)		LPB (I D)	
Reminder (d)	**0.075*****	(0.023)	**0.086*****	(0.019)
Monthly	YES		YES	
Weather	YES		YES	
Demographics	YES		YES	
N	25437		25437	
pseudo R^2	.07224			
R^2			0.1144	

Marginal effects, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

accounting should be induced to attend more by reminders (Implications 3 and 4) and our empirical findings illustrate that the population of individuals is indeed affected by either or both of these two biases. Before proceeding and trying to disentangle these two possibilities, we conclude this section illustrating the effect of reminder on users' decision to quit or not the gym.

So far we have analyzed a subgroup of the overall population exercising before, during and after treatment (i.e. non-quitting the gym). However, the composition of the treated group may be affected by the treatment itself (extensive margin). In particular, the treatment may discourage those who do not perform well, or on the contrary it may foster attendance making users more likely to continue exercising after the expiration of a given contract thus renewing a new contract. In an unreported regression we show that the probability of quitting the gym thus not renewing any contract during the treatment period is reduced by being treated with our reminders although the coefficient is not significant. Hence, if anything, reminders induce some users who otherwise would have quit the gym, to instead remain and continue exercising, thus increasing attendance also at the extensive margin.[14]

2.5 Why do reminders increase attendance?

As discussed in section 2.2 (Implication 2), if users are only affected by limited attention then, independently of reminders and in a given subscription, the probability to attend at any date t should be constant over time. Instead, individuals using mental accounts may

[14]This regression also shows that the probability of quitting is higher and significant for users at the first contract and for being male.

show attendance varying over the contract, increasing or decreasing in t, depending on the mental account they use. We begin by directly checking the daily probability of attendance over monthly contracts and we consider first monthly contracts signed before the start of the treatment in order separate the effect of treatment on attendance.[15] In figure 2.2 we plot the

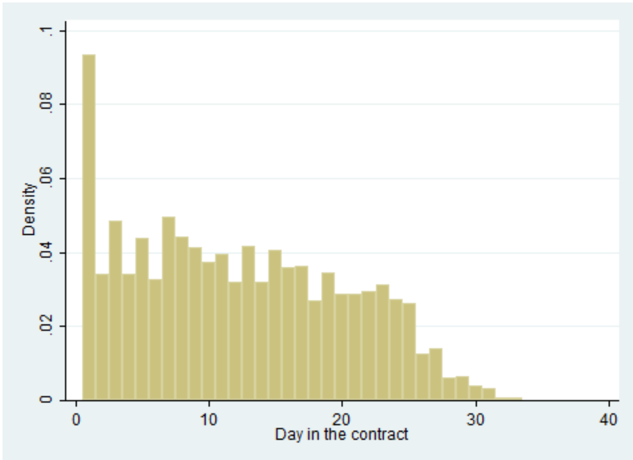


Figure 2.2: Probability of a visits at each contract day for monthly contracts.

observed density of attendance in each of the 31 days of a contract and simple inspection shows that there is a clear decreasing path over the month.

The figure also shows some spikes in attendance which are expected and correspond to the first day (when users buy a contract they are physically at the gym and immediately profit of the right to attend), and every seven days (since the week day of purchase and of first attendance tends to be the most preferred day in the week). Notice also that the day after a spike the probability of visit drops since although our users are quite active, it is rare that they attends two days consecutively.

Now, to model attendance during the contract we estimate the following linear probability model for observing a visit in day t in contract j signed by user u,

$$\begin{aligned} P(Visit\ in\ day_{t,j,u}) = & \gamma_2 2^{nd} week + \gamma_3 3^{rd} week + \gamma_4 4^{th} week + \\ & + \gamma_5 Remaining\ days + \beta_1 D_t + \beta_2 W_t + \beta_3 S_u \end{aligned} \tag{2.6}$$

The main explanatory variables of interest are the dummy variables for the four weeks in the contract. More precisely, we divide the span of the contract in five parts: the first week, from the 1st to the 7th day, the second from the 8th to the 14th, the third from the 15th to the 21st, the

[15] Again we refer to monthly flat contracts also because longer flat contracts may be affected by strong seasonality and by a long sequence of unpredictable and adverse contingencies which naturally decrease attendance over time.

fourth from the 22nd to the 28th and finally a dummy for the remaining days.[16] We use the first week as the baseline for these dummies. As before, we control for weather conditions with W_t, weekly and monthly seasonality with D_t, and individual characteristics with S_u. The analysis is performed on two samples: the first, labeled "pooled" refers to attendance before treatment and pools all users, the second, labeled "treated" refers to attendance of only treated users and during treatment. All regressions are performed with individual fix effects.

Results of the estimation are reported in the first two columns of table 2.5. The esti-

Table 2.5: Regression for the probability to attend in day t

Dep. Var: Probability to observe a visit in t	Before treatment pooled	During treatment treated	Before treatment pooled	During treatment treated
Red account			0.022	0.074***
			(0.014)	(0.020)
Week 2	-0.027**	-0.034*	-0.027*	-0.032*
	(0.014)	(0.019)	(0.014)	(0.019)
Week 3	-0.051***	-0.056***	-0.043**	-0.027
	(0.014)	(0.020)	(0.017)	(0.022)
Week 4	-0.088***	-0.084***	-0.076***	-0.041**
	(0.015)	(0.022)	(0.020)	(0.025)
Remaining days	-0.122***	-0.128***	-0.068***	-0.082**
	(0.025)	(0.037)	(0.025)	(0.039)
D_t	YES	YES	YES	YES
W_t	YES	YES	YES	YES
S_t	YES	YES	YES	YES
N	8854	4104	8854	4104
R^2	.12338	.17737	.12449	.18406

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

mated coefficients of the first column (before treatment condition) capture the decreasing

[16]We divide the month on a seven day basis to have the first (preferred) day in each of the weeks. We also include the $Remainingdays$ dummy since some contracts, due to calendar effects, are slightly longer than others.

path identified in figure 2.2: attendance reduces along the month and at an increasing rated, reaching a reduction of almost 9% in the last fourth week as compared with the first one. With standard t-test on the week dummy variables we reject the null hypothesis of joint equality to 0 (p-value=<.00001)[17]. Attendance within a subscription is indeed a decreasing function of time and this finding is not compatible with limited attention only, as derived in Implication 2. This result is further strengthened by estimated coefficients in column 2 of the Table where the same regression is performed on the sample of treated users during the treatment. Even if these users are made more attentive by reminders, the declining trend in attendance is also present for these users and there is no evidence for a reduction with respect to non-treated.[18]

In the last two columns of the table we provide evidence for a possibly more robust explanation for the effect of reminders. As we have illustrated in the model section, there are possibly many different accounting rules that individuals may have in mind.

Here we try to test directly the simplest rule illustrated in equation (2.2) according to which the account is "in the red" and negative as long as past attendance was not enough to justify the decision to purchase the flat contract, i.e. $A_t = -L_T + pD_t$. This account is weakly decreasing over time and should then imply a mental accounting effect which is decreasing over time along the time-span of the contract. Hence, we introduce in the econometric model a dummy variable, $Red\ account$, which takes value 1 if the account is in the red, i.e. $A_t < 0$, and value 0 if it is positive[19].

Starting form the third column, we observe that the inclusion of $Red\ account$ only slightly reduces the size of the weekly trend which is still present in the data. The estimated coefficient is positive as we expect but it is not significantly different from zero. We interpret this finding as the presence of a greater incentive to attend when the account is in the red which is however not so widespread in the population to emerge perhaps also because some users may employ types of mental accounting that are more sophisticated than the one we are currently considering. Alternatively, it could be that being inattentive, users do not react perfectly to the current status of the account. This is consistent with the fact that results change dramatically when we consider treated users during the period of treatment

[17]Single t-test performed on couples of coefficients lead to the following results: first, $H_0 : \gamma_2 = \gamma_3$ is not rejected (p-value=0.101), second, $H_0 : \gamma_2 = \gamma_4$ is strongly rejected (p-value= <0.0001), third: $H_0 : \gamma_3 = \gamma_4$ is rejected at 5%, (p-value=0.013) and fourth: $H_0 : \gamma_2 = \gamma_5$ is strongly rejected, (p-value=0.0083).

[18]In an extended regression (not reported) with both samples and interactions terms between weeks and a dummy for treatment period, we do not reject the null hypothesis that imposes a zero restriction on all these interactions terms.

[19]Notice that the gym sets independently the values of L and p and so it determines exogenously the pivotal visit at which the account of the user goes from red to positive.

with estimates are reported in the last column. Here the effect of the mental account is the dominant force in the first weeks of the month when users are reminded to attend and strive to reach the goal of going positive in their account. In the data we observe that more than 75% of the switch (from red to black account, respectively negative and positive) are made in week 2 and week 3 (36.7% and 40.4% respectively) and the change in magnitude of the weekly dummy between column 2 and column 4 (both on treated users in the during period) we testifies the importance of the mental account in determining attendance behavior. The effect of going in the black account in the second and third week is captured by the $Red\ account$ dummy and the weeks after, the third and the fourth (but also the remaining days), display a strongly reduced effect. Week 3 is roughly half of its value in column 2 (it goes from -0.056 to -0.027) and the same holds for the Week 4.

The complementarity of remainders and mental accounting in inducing attendance is further confirmed by a set of regressions, reported in table 2.6 on the effect of remainders. The regression performed is the same as in the previous case when we estimated the effect

Table 2.6: Regression for the effect of reminder during red or positive account

Dep. Var: Probability to attend in day t				
	Red Account		Positive account	
	Low attendance	High attendance	Low attendance	High attendance
Reminder	0.088***	0.059	-0.001	0.085*
	(0.026)	(0.038)	(0.041)	(0.044)
Monthly	YES	YES	YES	YES
Weather	YES	YES	YES	YES
Demographics	YES	YES	YES	YES
N	8689	6244	3689	5089
R^2	.15878	.20417	.13052	.16656
pseudo R^2				

Marginal effects; Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of reminders on the daily probability of attendance. Here the sample is split dividing the days when the account is in the red (first two columns) and the ones in which the account is in the black (last two columns). In the first column of each couple we consider the low attendance users while in the second the high attendance ones. As previously reported the effect of reminders is larger and strongly significant for low attendance users than for high ones. On top of this, we find that the effect of reminders on low attendance users vanishes when they reach the goal of going positive in their account while it is stable in case of high users, which is predicted by our model. Since low attendance users are characterized by high cost and or low benefits of attendance, an important contribution to their utility to attend may come from the mental accounting component. Because of that, when they are in red, the arrival of the remainder determines a revamp of this component in their net utility which determines a strong propensity to visit the gym. On the contrary, when the account is positive, this additional component may be very small and the propensity to attend in the next day is not affected. This is not the case for high attendance users who continue to respond similarly to the reminder independently of the account since they are more motivated by their smaller costs and or larger benefits of exercising.

The mental accounting we analyzed so far is only one of the many possible alternatives and has been empirically tested through the inclusion of a dummy variable which takes value 1 when the account is in red and 0 when the account is in black. In the following table we explore the same myopic account but we allow for a non linear response to the account status A_t. This means that we compute for each day A_t as the difference between L_T and the price for the past visits pD_t and we include it in regression. This allows to estimate the shape of $\Delta v(\cdot)$ which should be decreasing as the account becomes more positive so affecting at a decreasing rate the probability to attend. Result are reported in table 2.7.

There are now two remarkable facts. First, this accounting method is now strongly significant, independently of treatment, with a second order polynomial effect which is decreasing in attendance up to a value of +7.8 and +11 (which means to visit the gym more than 15 to 18 times since the first 6 visits are in the red part of the account). Hence, users are more likely to exercise when the account is in red and this effect fades out as the number of visits increases (notice that the second branch of the parabola is not relevant since it is associated to a very high number of monthly visits). Second, the negative path of attendance over the month which we have documented above is now entirely explained by the mental account since the coefficients of the week dummies now become positive (and significant) partly incorporating the positive constant of the second order polynomial[20].

[20]An estimation of a linear formulation of this account confirms this finding. However, we think that the better fit of the second order polynomial has a relevant economic interpretation since it is associated to the fact that the change of sign in the account, from red to black, matters for at least for part of the population. Hence, the parabola better capture the non linearity between the red part of the account and the black one.

Table 2.7: Regression for the probability to attend in day t

Dep. Var: Probability to observe a visit in t				
	Before treatment pooled		During treatment treated	
Red account	-0.033***	(0.0026)	-0.033***	(0.0036)
Red account^2	0.0021***	(0.0003)	0.0015***	(0.0005)
Week 2	0.069***	(0.0162)	0.0455**	(0.0222)
Week 3	0.114***	(0.0197)	0.0899***	(0.027)
Week 4	0.12***	(0.0227)	0.111***	(0.0326)
Remaining days	0.146***	(0.0289)	0.113***	(0.043)
D_t	YES	YES	YES	YES
W_t	YES	YES	YES	YES
S_t	YES	YES	YES	YES
N	8854		4104	
R^2	0.126		0.183	

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Feedback on past attendance

In order to understand the characteristics of individuals who acquire feedback and then compare their attendance before and during the treatment, here we consider all the 151 users who purchased monthly contracts at least before and during treatment but not necessarily after treatment.

Table 2.8 illustrates regressions for the probability to acquire feedback, performed only on the treated users who were the only users with the possibility to access to the personal web pages. Regressors are several individual characteristics that we obtained with our survey and online experiments (see Section 2.3). The variable "Symbol digit test" refers to the score of a fast test due to Lang, Weiss, Stocker, and Rosenbladt (2007) that has been employed also in other papers in economics (e.g. Dohmen, Falk, Huffman, and Sunde, 2010) and that we implemented in our web-site to measure computational speed (known to be strongly correlated with IQ). The "Numeracy" variable refers to another cognitive skill and is obtained as a score on a number of questions designed to elicit the ability of the subject to make simple calculations (Nardotto 2011 provides more details and shows how this score is strongly related to users' ability to correctly forecast future attendance). "Conscientiousness" is one of the "big five" non-cognitive personality traits defined in the psychological

Table 2.8: Probit regression for the probability to acquire feedback

	(A)	(B)	(C)	(D)
Symbol digit test	0.133**	0.141**		0.204**
	(0.067)	(0.067)		(0.083)
Numeracy	-0.050	-0.068		-0.188**
	(0.068)	(0.074)		(0.090)
Experience days		0.002*		0.002*
		(0.001)		(0.001)
Gender		-0.072		-0.022
		(0.117)		(0.128)
Age		0.015		-0.001
		(0.024)		(0.027)
Conscientiousness			0.137**	0.059
			(0.063)	(0.071)
Neurocitism			0.070	-0.066
			(0.059)	(0.077)
Monthly visits (BT)		-0.003		-0.005
		(0.010)		(0.010)
Patience				0.091***
				(0.031)
N	79	79	79	79
pseudo R^2	.04358	.10635	.09798	.28248

Marginal effects; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

literature to measure the degree to which a person is willing to comply with rules, norms and standards. "Neuroticism" is another "big five" non-cognitive personality trait measuring emotional stability (which in many contexts has been proved to affect economic choices) and in our case it could be a proxy for the attitude towards information acquisition (in particular bad news). Finally, "Patience" is a standard measure time preferences obtained with Holt and Laury (2002) procedure. "Monthly visits" are the average monthly visits in the period before the treatment. Other regressors, not included in these table, are the demographic characteristics already used in previous regressions, other measures of traits and the measure of disposable income obtained in our survey.

These regressions show that the probability to acquire feedback is larger for individuals with higher cognitive abilities but low numeracy skills. The probability is also higher for those who are more conscientious and patient. The first two elements are interesting, and their effect is strong. More skilled users, in terms of numeracy, do not look for feedback probably because they are already able to compute the implicit cost of their contract which was available in the feedback page and to recall the exact number of past visits. At the same time, cognitively endowed users are more likely to acquire information probably because they value it or simply because they are more curious (for a discussion on the relation between personality traits and cognitive skills, see Borghans, Duckworth, Heckman, and Bas, 2008). Experience seems to affect the decision to acquire feedback. The effect on the probability of 90 days of experience is to increase it by 18%. This coefficient is stable across specification but always borderly significant. Personality traits are borderly significant but the sign, at least for conscientiousness, is the one we expect: users who have a better score are more likely to acquire feedback. Finally, more patient users are more likely to acquire information.

We now concentrate on the group of individuals acquiring feedback (26 users in the group of 79 treated). First, the following picture illustrates the density (across all users) of

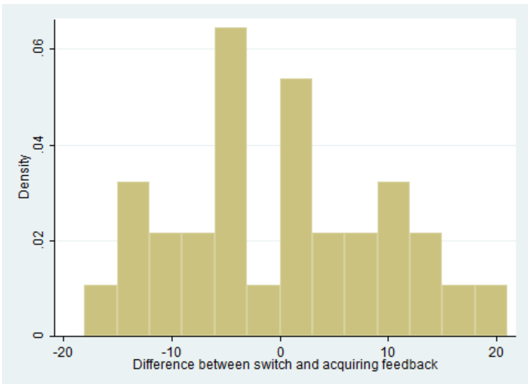


Figure 2.3: Day of the month in which the feedback is acquired with respect to the switch from red to green

this decision to access the feedback web-page in relation with the days before and after the day in which the account defined as $A_t = -L_T + pD_t$ changes sign becoming positive. Notice that users using feedback inspect the web-page not just once but repeatedly for many days. More importantly, we observe two spikes at 3 to 6 days before the switching day of the account and the second spike just after the switch. This suggests that those who use feedbacks like observing the confirmation of having performed well at the gym and look for this information in two precise moments: when they are closed to pass the pivotal visit that makes the account positive and right after this visit to the gym.

Now we estimate the same regression illustrated at the end of the previous section, with a linear probability model of attending in a given day on the three week dummies (see Table 2.9) but we restrict to the individuals who do acquire feedbacks.

Table 2.9: Regression on the effect of being in the red account

Dep. Var: probability to observe a visit in day t		
	Coefficient	Std. Error
Red account	0.083*	(0.044)
Week 2	-0.046	(0.038)
Week 3	-0.029	(0.047)
Week 4	-0.021	(0.059)
Remaining	-0.075	(0.075)
N	1195	
R^2	.16746	

Marginal effects; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Remarkably, for these individuals the coefficients of the week dummies are no more significantly different from zero, whilst the dummy for the mental accounting continues to be significant, positive and of the same magnitude. Hence, it seems that for individuals acquiring feedbacks, the main driver of the decision to attend is strictly related to mental accounting.

We conclude this section by investigating an interesting effect of feedbacks. Increased attendance induced by reminders has clearly an effect on whether a given subscription turns out to be ex-post an optimal choice in the sense that it minimizes the subscriber expense among the options available in the menus of the gym. This measure of cost optimality has been extensively used in the literature to verify whether individuals are present-biased (see for example DellaVigna and Malmendier, 2006 and Nardotto, 2011). As a measure of (sub-)optimality we thus define a monetary loss

$$L = \frac{\hat{S}(D_T) - \underset{S \in Menu}{min} \; S(D_T)}{\hat{S}(D_T)}$$

where $\hat{S}(D_T)$ and $S(D_T)$ are the costs respectively for the contract actually subscribed and for any (combination of) contract available in the menu offered by the gym and these costs are in general functions of total attendance D_T actually observed at the end of actual subscription. We then estimate a probit model (with clusters at the user level) for the probability that the contract k of user i generates a loss $l_{k,i}$ with

$$l_{k,i} = max(0, l^*_{k,i})$$

where

$$l^*_{k,i} = T_i + D_k + TD_{k,i} + C_{k,i}\beta_1 + P_i\beta_2 + S_k\beta_3 + \varepsilon_i \tag{2.7}$$

and $T_i,\ D_k,\ TD_{k,i}$ are respectively dummy for being treated, dummy for contract in the period during treatment and interaction, $C_{k,i}$ is a vector of users and contract's characteristics (e.g. whether it is a flat rate, experience of the user at the moment of subscribing and others), P_i are user's characteristics (gender, age, academic background and others), S_k are other contract specific factors including weather and distance from home.

Table 2.10 reports the estimated coefficients. Since reminders do increase attendance

Table 2.10: Probit regression on the probability of incurring in a loss

Dep. Variable: Prob. of incurring in a loss				
	Full sample		**Info-active + controls**	
	Coeff.	Std. Error	Coeff.	Std Error
During treatment (d)	0.143***	(0.051)	0.158***	(0.056)
Treated (d)	-0.037	(0.044)	-0.074	(0.055)
Treated x during (d)	**-0.102****	(0.044)	**-0.127*****	(0.047)
Weather	YES		YES	
Academic	YES		YES	
Demographics	YES		YES	
N	771		462	
pseudo R^2	0.08168		0.10513	

Marginal effects; Robust standard errors in parentheses

and higher attendance reduces the loss of a contract, the probability of incurring a loss is reduced by the interaction of the treatment and the during-treatment period dummies. These regressions performed on the larger sample of 146 users who buy contracts before the treatment and during the treatment periods, confirm the previous findings on the restricted users with contracts in all periods. What is interesting to notice is that in the case of individuals acquiring feedback (column two) this interaction is even stronger and more significant. Although our web-site also allowed users to calculate the least cost contract (or sequence of contracts) for given time horizon and expected attendance, we rather think that more plausible explanations are that either individuals attending more and thus incurring in lower losses sort into the group of those acquiring feedback, or feedback further revamp mental accounting and induce higher attendance (or both these two effects are active).

2.7 Concluding remarks

In this paper we investigated the effects of reminding consumption opportunities to individuals who are likely to be inattentive and using mental accounting strategies. We have tested this possibility on a sample of subscribers in a gymnasium, finding that reminders strongly increase attendance in the gym for treated users receiving weekly E-mail, especially for those users with relatively low attendance before treatment and that reminders have an effect few hours after being received. We have also clarified that inattention alone cannot explain the change of behavior we observe and that an at least concurrent (if not predominant) channel through which reminders foster attendance is mental account. Individuals indeed respond in their attendance decision to a mental account that contemplates the upfront subscription payment to attend the gym, actual attendance and the expense associated with alternative means to attend. Finally, we have studied the identity of individuals who deliberately decided to acquire feedback on past individual performance and relate this to a further increase in attendance and lower monetary expense to attend.

With our filed experiment and simply using reminders and feedback we were able to obtain results comparable to those that have been obtained in similar environments but with significant monetary incentives. A word of caution must be warned as usual when extrapolating our results to different environments. However, we think that they are important in terms of policy implications since they show that good habits can be induced with a very cheap but still effective instrument, information, also very much limiting the risk of crowding out intrinsic motivations.

Our analysis has been limited to individuals who, independently of our treatments, decided to enter the market, the gym, and at least to try to exercise. An interesting avenue for future research is to see if and how reminders may also induce individuals to actually start (and then continue) with desirable and healthy activities.

A simplistic extrapolation in terms of policy of our results is that we could try to incentivize individuals on many different and desirable investment activities by simply sending them as many reminders as the activities. We expect however that the problem of limited attention would emerge again although, in this case, with respect to the many reminders competing for attention. It would be interesting to verify whether this is the case and what is the optimal number of reminders which limits the risk of information overload.

2.8 Appendix

2.8.1 Tables

Contract menu and actual purchases.

Table 2.11: Attendance by contract category and prices (year 2010)

Type of contract	**Price (€)**	No.	Rel. freq.
Single entrance	6,00	548	8.8%
Carnet (10 visits)	50,00	122	2%
Carnet (20 visits)	90,00	62	1%
One month	45,00	3918	63.1%
Two months	85,00	308	5.0%
Three months	118,00	518	8.3%
Academic flat	270,00	58	0.9%
Flat 1 euro promotion	–	281	4.5%
Specific contracts	–	384	6.2%

The text of the E-mail (English version).

Dear student,
do you want to know your actual use of GymCus?
Go to your personal webpage at [http: address] (send an E-mail to cusb.gym@unibo.it if you forgot your login credentials).
In the webpage "Information on your activities at the gym: learn how to use the GymCus!" you can:

1- Check your weekly attendance and associated actual price for single entrance

2- Plan and check your visits to the gym

3- Find the least costly solution that better fits your needs (mostly useful when you are about to subscribe a new contract)

4- Unsubscribe this E-mail service

Use these information services for a more convenient use of your GymCus!
Regards, cusb.gym@unibo.it

Chapter 3

On the relationship between risk aversion and impatience with cognitive abilities and personality traits

Abstract

In this paper we investigate the relationships between individual characteristics of cognition and personality and two fundamental determinants of economic decision: risk aversion and impatience. To investigate this link we run an online experiment on a sample of students enrolled at the University of Bologna. Experiments where incentivized through monetary rewards and incentive compatible. Cognitive skills have been measured through an established test and personality traits have been elicited through a standard questionnaire employed in both psychology and economics to measure the Big five personality traits. We find that risk aversion and impatience are negatively associated with higher cognitive skills. Moreover, subjects with higher numeracy skills are more likely to behave as risk neutral. Personality traits are the main predictors of time preference and in particular neurocitism and conscientiousness. The former is the degree to which a person is emotionally stable when facing stressful situations, the latter is the degree to which a person sticks to plans both internally and externally imposed. Finally, we show that personality facets like impulsivity strongly affects behaviour in the experiment and much can be gained in the analysis taking into consideration these factors.

Keywords: Cognitive ability, personality traits, risk aversion, patience, online experiment
JEL codes: D12, D81

3.1 Introduction

It has been recognized, both by economist and psychologist, that economic outcomes are related to cognition and personality. Cognitive skills are fundamental to interpret, collect and elaborate information and so to make economic decisions. At the same time, personality traits may also influence the mental processes related to choice, determining relevant consequences in economic decisions. For example, the emotional reaction to uncertainty of an individual is likely to affect his attitude towards risk. At the time the same factor behind this reaction may influence also the response for example to the possibility to acquire information or the way information is processed.

So far, both in the economic and psychology literature, most of researchers investigated these relationships analyzing the correlations between both cognitive skills and personality traits and labor outcomes or other economically relevant life outcomes (examples are the probability of being unemployed, earnings, mortality, unhealthy habits etc.). In this paper we study the relationships between risk aversion and impatience with cognitive abilities and personality traits. We measure directly these preference characteristics through two payed experiments and we elicit cognitive skills and personality traits of the participants.

The experiments are performed online and the subject pool of participants is composed by 300 students at the University of Bologna. The design of the experiments is close to the one of Holt and Laury (2002) on risk preference. We elicit personality traits following the established categorization in psychology of the Big five personality traits. This theory of personality reduces the multitude of facets that characterize human behaviour to five factors (openness to experience, conscientiousness, extraversion, agreeableness and neurocitism) which can be considered at the highest level of generalization. We consider two types of cognitive skills. The first is the ability to process and recall information which is measured through a test developed by Lang, Weiss, Stocker, and Rosenbladt (2007). The second are numeracy skills (in our case we focused on the ability in making simple calculations: to compute sums, divisions and ratios) which are measured by a small sample of questions.

This paper is close to the one by Dohmen, Falk, Huffman, and Sunde (2010) (hereafter DFHS). In this respect we find that the relationships highlighted in this paper between cognitive skills and both risk attitude and patience are confirmed. In particular, higher cognitive skills are associated to lower risk aversion and greater patience. Moreover, two new results come out from our study. The first is related to the role of numeracy skills, which proved to be important in determining the attitude towards risk of subjects. In particular, we find a positive and strong correlation between numeracy skills and risk neutral behaviour. This result is robust to the inclusion of the academic background of subjects among the explanatory variables and confirms the findings of other studies such as Benjamin, Brown, and

Shapiro (2005). The second result regards the role of personality traits in affecting the time preferences of subjects. We find neurocitism (that is defined as the degree to which a person is emotionally unstable and experiences the world as threatening) as an important predictor of patience. Coherently with recent studies such as Anderson, Burks, DeYoung, and Rustichini (2011), we find that more emotionally stable subjects are also more patience. Another trait which seems to have a role in affecting time preferences is conscientiousness, which is defined as the degree to which a person is willing to comply with rules and standards, both internally and externally imposed. In this case we find a positive effect of this trait and the willingness to postpone gratifications. However, a statistically significant relation emerges in some tests but not in general specifications of our econometric estimations.

The second contribution is methodological. On one side we show that results coming from an online experiment are comparable with those obtained in a more controlled environment (in the case of DFHS subjects received the visit of a trained experimenter in their home). This is in line with the few studies that compare the behaviour of participants in online and laboratory experiments and provides support for the reliability of online experiments. On the other side we show the importance of including, both in the case of laboratory and online experiment, measures of impulsivity in the battery of test proposed to subjects. In our case, as in every experiment, subjects may be not be very attentive to their task (even if the experiment is payed and incentive compatible), so it may be worthy to control for disturbing factors like impulsivity in completing the activity. Including into our test a short number of questions proposed by Frederick (2005), designed to measure the degree of cognitive reflection and impulsivity of subjects, we show that noise in the data can be effectively reduced once controlling for this element in the analysis.

The paper is organized as follows: is section 3.2 we present the data. In particular, we discuss how measures of risk attitude, patience personality traits and demographics are obtained. In section 3.3 we report the empirical analysis and in section 3.4 we perform some robustness analysis on the previous results. Section 3.5 concludes.

3.2 Data description

Data were collected as part of a study on choice behaviour in a subscription market (see Nardotto 2011b). The subject pool is composed by students at the University of Bologna and experiments have been run from September 2009 to the end of July 2010. The age of these students ranges from 19 years old to 34 years old and the vast majority of them is between 19 and 27. Participants went through a survey on demographics, the tests on cognitive abilities and personality traits and, as a last step, they run the payed experiments on risk and time preferences[1].

3.2.1 Measures of cognitive ability

The test we adopt to measure cognitive abilities borrows from a submodule of the nonverbal section from the Wechsler Adult Intelligence Scale (WAIS) which is the most established test in psychology to evaluate the cognitive skills and is composed by different sections designed to measure different aspects of cognition. In the test we use, subjects have to match symbols accordingly to a mapping and have a time limit. The purpose of the test is to evaluate how fast subjects are in processing and recalling information. We opt for this particular test because it is well suited to be performed online and Lang, Weiss, Stocker, and Rosenbladt (2007) proved it to be highly correlated with general IQ measures. The test consists in the following: subjects are told about the content of the activity which consists in being faced with nine unfamiliar symbols each of them paired with a digit from one to nine. Beneath the box with the symbols and the digits, subjects find a start button and below two rows of covered symbols that become visible only at the moment of start. Once the subjects press the start button, they have ninety seconds to fill as much entries as they can. Whenever a number is inputted, the cursor goes automatically to the next symbol. Two figures of the screen are reported in appendix 3.6.1. The test is made of three stages with increasing difficulty and is preceded by a trial stage to show to the subject the functioning of the test.

The program is written in Javascript and it is run locally by the computer of the subject so no slowdown can be caused by the connection to the remote server. Moreover, since it is also a very simple program, no slowdown is expected to occur during the test. If a subject interrupted the test (closing the window, shutting down the computer etc.) we recorded this event in our database. In this way, when the subject connected again to the web site she had to start again from the beginning. Overall, 14 subjects interrupted the test but this happened always at the very beginning (during the trial stage) and at the second connection they run the entire test properly. The rest of the subjects did the test at the first time without interruptions.

The distribution we get for the score in this test is reported in figure 3.1. As we see from the graph, a subject did not run the test properly but waited until the time was up. In the analysis we exclude this observation from the sample[2]. We run the same cognitive

[1]Every step has been implemented online and the technology we adopted is php/MySQL. For the interactive sections (symbolic test and risk and time experiments) we have designed a dynamic interface in Javascript.

[2]We see also that this subject performed very poorly in the impulsivity test. We did not condition the payment for participation to any performance indicator, and is very likely that this subject did not pay much

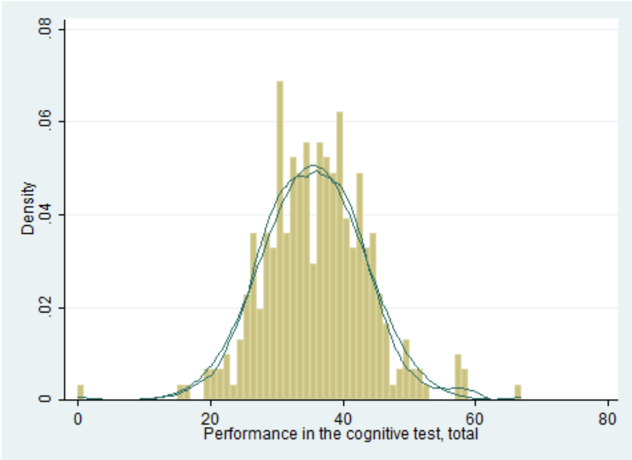


Figure 3.1: Distribution of scores in the cognitive ability test

test in a class of undergraduate student of the university of Bologna and a Kolmogorov-Smirnov test for the equality of the two distributions do not reject the null hypothesis that the two samples come from the same population (p-value=0.325). Hence, performing the test online rather than in a laboratory do not seem to have an effect on the performance of subjects.

The second measure of cognitive ability we employ is based on a short test for numeracy skills. It consists in four questions designed to evaluate the attitude of subjects in doing simple computations of ratios, sums and proportions. Questions and the distribution of scores are reported in appendix 3.6.1.

3.2.2 Measures of personality traits

The modern theory of personality identifies five factors at the broadest level of abstraction: openness to experience, conscientiousness, extraversion, agreeableness and neurocitism. All these factors are categories to which different facets (hence considered at lower lever in this hierarchical scale) belong. For example, neurocitism, defined as the degree to which a subject is emotionally stable, has as its facets, among the others, traits like anxiety and vulnerability. Hence under this approach, to obtain a measure for each trait, it is necessary to elicit the different facets that belong the broader factor.

In our case we focus on two of these five traits, conscientiousness and neurocitism, since we consider them as the most relevant in the choice problem we propose to the subjects. Conscientiousness is defined as the degree to which a person is sticky to externally or internally imposed plans and the tendency to act consciously. Neurocitism, often labeled

attention to the activities we proposed.

emotional stability, is the tendency to experience negative emotional states such as anxiety and anger. Other than these two, we elicit two other traits which can be relevant in our case: internal and external locus of control. The former is the degree to which a person believes that life outcomes are determined by own actions and abilities while the external locus of control is the degree to which she believes they are determined by external elements beyond her own control such as fate or luck.

To measure these personality traits we employ standard questions proposed by the psychological literature and in particular the same questions employed in the German Socio-Economic Panel (SOEP). The survey approach to elicitation of traits is obviously the common choice in large scale studies where direct elicitation by professional psychologists is not feasible. For a discussion on the strength and the weaknesses of this approach see Borghans, Duckworth, Heckman, and Bas (2008).

3.2.3 Experiments on risk aversion and impatience

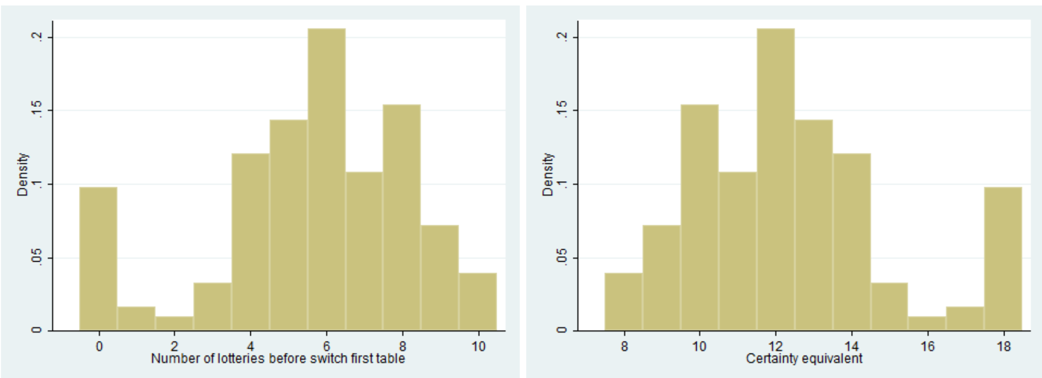
The experiments to elicit the attitude towards risk and patience are payed and incentive compatible. Every subject does both of them starting from the one on risk attitude. The stages are the following: first, subjects read the instructions of the risk experiment; second, they run a tutorial which has the same features of the payed stages and at the end of the tutorial stage, it is performed a simulation of a final payment. Third, in order to check whether the subjects have understood the content of the experiment, they must fill a questionnaire and cannot proceed to the payed stage until the provided answers are all correct. Fourth, the payed stages on risk preference are run. After this stage, the subjects follow the same procedure for the time experiment, so they read the instructions, run the tutorial, answer to the questions and finally do the payed stage. When the time experiment is finished, subjects go to the web page where the payment is displayed and automatically the experiment ends[3]. Instructions of the experiment are reported in appendix 3.6.3.

The design of both experiments is very close to Holt and Laury (2002). In the experiment on the attitude towards risk, subjects face two tables with 10 rows each. Each row is made by a couple of alternative payments: option A and option B. The former is an amount of money that is granted for sure (e.g. 20 euro); whereas the latter is a lottery that have a small prize (e.g. 5 euro) and big prize (e.g. 35). The two payments of option B are granted with a probability 1/2. All the rows are displayed from the beginning of the stage and the

[3]Subjects are part of a field experiment so they are registered in our database and can login in the experiment web site only through their account. Before seeing the actual displayed in the screen, the access of the subject to the web page of the experiment is disabled so nobody could re-start the experiment in order to get a higher payment.

sequence is such that option B remains constant while option A becomes at each row (moving downward) less attractive since it is reduced at each step by a fixed amount. Because the valued of the option B is constant and the one of option A is decreasing, a subject with monotonic preference should switch from A to B at a certain point of the table and then continue to select B. However, it is common in the literature to see cases in which subjects select again option A after the first switch to B. This makes difficult to infer the degree of risk aversion of these subject who in general amount to 8%-12% of the subject pool. In our case we design the experiment in a way that lessen this issue. First, we allow the subject to go back to a previous row to change the selected option at any point in the table. Second, when option B is selected, a pop-up window appears in the screen saying that in the subsequent rows option A guarantees a lower payment. When the pop up disappears, subjects can complete automatically the table pressing a button that becomes active besides the table or to continue to fill it manually. In our case, with this two elements, we do not observe any switch back to A after a switch to B.

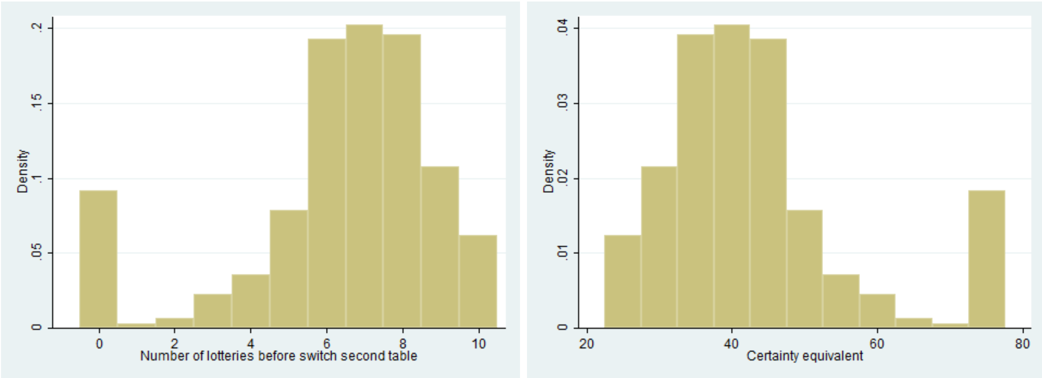
Switching points associated to risk neutrality are placed differently in the two tables. In the former a risk neutral agent should switch at row 7 while in the second table at 6 or 7. Payments differs among the tables. In the first, option B has a small prize of 5 euro and a large one of 20 euro (this implies an expected value of 12.5 euro) while the certain amount starts from 18 euro and is lowered at each row by 1 euro (hence, in the last row it reaches a value of 9 euro). In the second table, option B has a small prize of 20 euro and a large one of 80 euro (this implies an expected value of 50 euro) while the certain amount starts from 75 euro and is lowered at each row by 5 euro (hence, in the last row it has a value of 30 euro).



First table switching row | First table certainty equivalent

Figure 3.2: First table switching row and corresponding certainty equivalent

Results for the switching rows in the two risk experiments are reported in figures 3.2, 3.3. As we see from the graphs a non negligible share of subjects switches immediately



Second table switching row Second table certainty equivalent

Figure 3.3: Switching row and corresponding certainty equivalent, second table

in both tables. Without considering this group, which is treated separately in the following, few subjects are very risk loving. If we split the population between risk loving subjects (switches between 1 and 4), risk neutral (switches between 5 and 7) and risk averse (switches between 8 and 10) then the shares are respectively 19.39%, 50.72% and 29.35% in the first table with lower payment and 7.55%, 52.16% and 40.29% in the second table with higher payment. This seems to suggest that users tend to be risk neutral and risk averse, expecially when payments increase[4]

The experiment on impatience is similar to the previous one. In this case each row has 2 options, A and B, where the former is 20 euro of telephone credit which is transfered in 2 days, while the second is a 40 euro telephone credit transfered in a later point in time[5]. Moving down in the table, which in this case has 20 rows, the time span between the small amount and the large amount of credit increases since latter is made more far in time. In the first row the subject has to wait 7 days which are increased constantly by 4 days per row until 51 and then by 10 days per row. In the last row to obtain the large prize the subject has to wait for 131 days.

As for the risk experiment, once the subject makes the first switch, a pop up window

[4]From our survey we know that the average monthly income (net of expences) of the population is 200 euro. Hence, the expected value of the lottery amounts to 6.25% of the monthly income in the first table and to 25% in the second table.

[5]Subject have been informed about the nature of the payment (telephone credit) in the instructions. When the experiment took place, the vast majority of consumers in the mobile telephone market in Italy has a *pay as you go* contract so our payment valuable for our subject pool. Those who could not receive the payment had the possibility to contact the experimenter and ask for a different type of payment. Nobody however did use this option.

appears saying that in the following rows the large payment is made even more far and automatic completion button appears. To lessen the issue of transaction costs of obtaining the payment at different dates, which can be a serious concern in this experiment, the telephone credit has been transfered electronically by the experimenter so no cost is born by subjects to receive their prize at any date. Results for the switching row in the time experiment are reported in figure and 3.4. The modal row is the last one and is associated with a annual discount rate lower or equal to 5.8%.

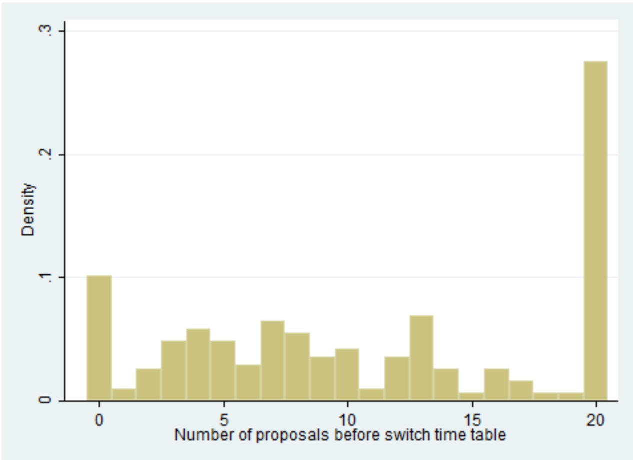


Figure 3.4: Number of rows without switch in the time experiement

Considering the data obtained in the three tables, an important issue regards the immediate switches made by a subgroup of subjects. They selects to switch at the first row in both tables of the risk experiment and are very impatience in the risk experiment, switching at the beginning of the table. We explain this behaviour through the possibility to quit the experiments quickly through the automatic completion button. This explanation is supported by the fact that this group of subjects perform poorly in the impulsivity test (see section robustness for more details) but interestingly not in the symbolic test, which cannot be skipped since the time duration of the stages is fixed. Hence, for this group of subjects, the observed behaviour in the experiment is likely not to reflect their actual characteristics. This is an important issue that researcher who run experiment online must consider as a potential source of noise in the data. Because of this, in the analysis we focus on the rest of the population

3.3 Result

Our findings are in line with results emerged in the literature. We find however other results that have not been pointed out so far. First, as in Dohmen, Falk, Huffman, and Sunde

(2010), we find a negative relation between cognitive ability and risk aversion; second, we find a positive relation between cognitive ability and patience. Other than that, we also show that cognitive skills are positively associated with being risk neutral and in particular numeracy skills. Moreover, we find a relation between some personality traits, namely conscientiousness and neurocitism, and impatience.

3.3.1 Relationship between risk aversion and impatience with cognitive skills

We start by examining the relationship between risk aversion and cognitive skills. Table 3.1 reports the estimated coefficients of an interval regression[6] of the switching row on the measures of cognitive skills. The measures of cognitive skills we consider are the score in the symbolic test[7] and score in the numeracy test. Column (A) to (C) refers to the first

Table 3.1: Estimation of the relation between risk aversion and cognitive skills

Dependent variable	Number of rows before switching to option B					
	First table			Second table		
	(A)	(B)	C)	(D)	(E)	(F)
Symbol digit test	-.257**		-.241*	-.183*		-.174*
	(0.031)		(0.051)	(0.076)		(0.098)
Numeracy		-.198	-.157		-.120	-.091
		(0.283)	(0.405)		(0.339)	(0.475)
Constant	6.76***	7.34***	7.25***	7.53***	7.9***	7.81***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
log pseudo-likelihood	-564.98	-571.51	-564.51	-534.56	-537.23	-534.36
N	275	276	275	277	278	277

p-values in parentheses, robust standard errors

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

table of the risk experiment where option A range from 4% of the average monthly income of subjects to 10% and the certainty equivalent of the lottery is 6,25% of this income. Columns (D) to (E) to the second table where payoffs range from 10% to 40% and the certainty equivalent of the lottery is 25% of the average monthly income.

[6]We follow Dohmen, Falk, Huffman, and Sunde (2010) estimating this generalization of the tobit model which allows for left and right-censored observation with known intervals.

[7]We standardize this measure so it has mean equal to zero and standard deviation equal so one.

Both cognitive measure are negatively correlated with the degree of risk aversion represented by the number of rows where option A is selected. Subjects with higher cognitive skills tend to have switch to the risky option before meaning that they are willing to bear more risk. Coefficients however are significant at 5 and 10 percent only for the first measure of cognitive skills. This suggest that cognitive skills are associated with less risk aversion, as pointed out in the literature, but numeracy skills are less important.

The role of numeracy skills emerges in table 3.2 which reports the estimates of a probit regression for a switch to option B in the rows associated with risk neutrality[8]. The positive coefficients of our general measure of cognitive skills, captured by the digit score test, suggest that risk neutral agents tend to be more skilled but only in table 1 column (A) the coefficient is significantly different from zero. The main result we find regards the importance of numeracy skills. Coefficients are strongly significant and robust to the inclusion of the other trait in the specification. The two tables confirm the findings of previous research on the relationship between cognitive skills and risk aversion and add an important element to the picture since we identify in the numeracy skills a determinant of the risk attitude of economic agents.

Table 3.2: Estimation of the relation between risk neutrality and cognitive skills

Dependent variable	Probability of switching at 6th or 7th row					
	First table			Second table		
	(A)	(B)	C)	(D)	(E)	(F)
Symbol digit test	.0686**		.0583*	.0358		.0244
	(0.021)		(0.052)	(0.246)		(0.437)
Numeracy		.109***	.0983**		.129***	.124***
		(0.006)	(0.014)		(0.002)	(0.003)
N	275	276	275	277	278	277
Pseudo R^2	.01507	.02175	.0323	.00356	.02611	.02766

Marginal effects; p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[8]To perform this analysis we create a dummy variable that takes value 1 when the subject switches to option B at rows 6 or 7. In the second table row 6 is the one where the expected value of the two options are the same (50 euro) while in row 7 the value of option A decreases to 45 and the lottery does not change. Hence, at row 6, a risk neutral agent is indifferent between the two options while in row 7 should select option B. In table 1, the expected value of option B is 12.5 euro while the value of of option A is 13 in row 6 and 12 in row 7. A strict rule in this case would be to consider as risk neutral only subjects who switch at row 7 but we consider such a criterion too restrictive and we categorize as risk neutral also subjects who switch at row 6.

The empirical analysis for the relationship between impatience and cognitive skills is similar to the previous. We estimate an interval regression model for the switch to option A on both our measures of cognitive ability. Table 3.3 reports the result of the estimation. Estimated coefficients have the expected sign but the score in the symbolic digit test is not

Table 3.3: Estimation of the relation between patience and cognitive skills

Dependent variable	Number of rows before switching to option A		
	(A)	(B)	C)
Symbol digit test	-0.184		-0.338
	(0.737)		(0.545)
Numeracy		1.292*	1.360*
		(0.095)	(0.084)
Constant	13.911***	9.946***	9.713***
	(0.000)	(0.000)	(0.000)
log pseudo likelihood	-760.8	-760.598	-758.946
N	274	275	274

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

significant. It is however strongly significant the numeracy score in both the specifications of columns (B) and (C).

3.3.2 Relationship between risk aversion and impatience with personality traits

Personality traits have been recognized by the psychological literature as relevant in predicting socioeconomics outcomes such as academic achievement and job performance. A growing economic literature is focusing on this relationships and notable examples are Heckman and Rubinstein (2001) and Heckman, Stixrud, and Urzua (2006). Our results confirms this relationship in particular with respect to the link between patience and personality traits. Table 3.4 reports estimated coefficient of an interval regression between the switching row and the personality traits we are interested in. Neurocitism is defined as the tendency to experience negative emotional states such as anxiety and anger. The measure we have is reversed and the higher the score the more stable a subject is. The estimated coefficient of neurocitism is positive and strongly significant, meaning that more emotionally stable subjects are also more patient in the payed experiment. An important trait determining choice in our experiment is impulsivity. As for neurocitism the scale in which the trait is

Table 3.4: Relationship between patience and personality traits

Dependent variable	Number of rows before switching to option A								
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
Neurocitism	1.047^{**}					1.412^{***}	1.200^{**}	1.880^{***}	1.436^{**}
	(0.039)					(0.009)	(0.028)	(0.002)	(0.018)
Conscient.		0.672				1.005^{*}	1.047^{*}	1.339^{**}	1.340^{**}
		(0.229)				(0.092)	(0.077)	(0.047)	(0.042)
Internal loc			-0.276			-0.769	-0.809	-1.300^{*}	-1.232^{*}
			(0.615)			(0.212)	(0.181)	(0.074)	(0.074)
External loc				0.377		0.604	0.643	0.452	0.562
				(0.474)		(0.281)	(0.244)	(0.467)	(0.355)
Impulsivity					1.354^{***}		1.282^{***}		2.430^{***}
					(0.005)		(0.008)		(0.000)
N	275	275	275	275	275	275	275	306	306

Marginal effects; p-values in parentheses, robust standard errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

measured is reversed in the sense that the higher the score, the less impulsive is the subject. Columns (E) and (G) report the estimated coefficient for this trait. As we said at the end of section 1.3, we focus here on the restricted subject pool of participants who did not switch immediately at the first row in the risk experiment and in the time experiment. Hence, these coefficients, which are always strongly significant, show the importance of this personality trait in determining economic choices, especially those involving intertemporal trade-offs. Conscientiousness is defined as the degree to which a person is sticky to externally or internally imposed plans and the tendency to act consciously. Also in this case we find a positive coefficient which is weakly significant in specifications (F) and (G). The last two columns reports estimated coefficients for the entire sample of subjects. In this case the results are in line with previous findings and the significance of conscientiousness increases.

3.4 Robustness

In this section we perform some robustness checks to verify whether the relationships we identified are robust to the inclusion of elements that have been shown to affect traits, both cognitive and personality traits. This in turn may affect the relationship between these primitives of cognition and personality and the economic preference of subjects.

The first one is age. As summarized by Borghans, Duckworth, Heckman, and Bas (2008), traits tend to evolve over time at a slow rate, so in our case, where the age spectrum is limited, we do not expect to find a significant effect of age on the traits and so on risk attitude and impatience. The second one is the gender which has been shown to be related positive to patience, see for example DFHS. In our case the dummy variable takes value 1 when the subject is male. The third is disposable income. In particular, each subject has been asked to indicate his monthly disposable income net costs of board and lodging and the distribution in the population is reported in figure 3.5. The graph of the variable is reported in figure 3.5. Income is introduced in the regression model in a discrete way. We divide the population between low income, medium income and high income subjects where the former category is given by the first 25 percentile of the population and the latter by the last 25 percentile. This allows to better capture the different groups in the population of budget constrained and high income subjects. The last control is the academic background of subjects. Since this may be relevant to determine skills, as in the case of numeracy skills, we control also for this element which we can observe[9].

[9]In case of numeracy skills this is particularly evident in the data. Subject with a mathematical background display a positive correlation with the numeracy index equal to 0.3 while students in literature and other schools which belong to the humanities display on average a correlation of -0.21. However, due to space limitations we

Table 3.5: Robustness checks

Dep. var.	Number of rows before switching to option B				Probability of switching at 6th or 7th row				Number of rows before switching to option A		
	First table		Second table		First table		Second table				
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(L)	(M)
Symbol digit test	-0.268**	-0.323**	-0.175*	-0.229**	0.054*	0.037	0.027	0.033	-0.276		-0.496
	(0.029)	(0.011)	(0.095)	(0.037)	(0.075)	(0.239)	(0.394)	(0.321)	(0.615)		(0.367)
Numeracy	-0.173	-0.298	-0.060	-0.110	0.087**	0.055	0.128***	0.122***	1.472**		0.787
	(0.366)	(0.132)	(0.640)	(0.403)	(0.032)	(0.213)	(0.003)	(0.009)	(0.036)		(0.289)
Mid income	-0.325	-0.369	-0.172	-0.178	-0.001	-0.010	0.050	0.036	0.260	0.019	0.000
	(0.209)	(0.155)	(0.432)	(0.417)	(0.991)	(0.877)	(0.465)	(0.603)	(0.826)	(0.987)	(1.000)
High income	-0.456	-0.407	-0.267	-0.253	-0.058	-0.072	-0.108	-0.121	-3.112**	-2.979**	-2.965**
	(0.159)	(0.201)	(0.431)	(0.441)	(0.450)	(0.347)	(0.173)	(0.129)	(0.026)	(0.017)	(0.031)
Male	0.252	0.211	-0.253	-0.298	0.103*	0.056	-0.038	-0.067	-1.643	-1.817*	-1.887*
	(0.318)	(0.416)	(0.236)	(0.182)	(0.079)	(0.369)	(0.547)	(0.307)	(0.127)	(0.099)	(0.082)
Conscient.		0.221*		0.197*		-0.034		-0.081**		0.865	0.875
		(0.067)		(0.098)		(0.268)		(0.013)		(0.134)	(0.106)
Neurocitism		0.033		-0.078		0.033		-0.022		1.224**	1.294**
		(0.784)		(0.506)		(0.295)		(0.493)		(0.023)	(0.017)
External loc		0.018		0.006		-0.027		0.022		0.579	0.507
		(0.891)		(0.956)		(0.393)		(0.510)		(0.291)	(0.351)
Internal loc		-0.078		0.134		0.051		0.044		-0.521	-0.572
		(0.598)		(0.283)		(0.138)		(0.215)		(0.403)	(0.346)
Impulsivity		0.225*		0.094		0.062**		0.029		1.378***	1.235**
		(0.065)		(0.342)		(0.042)		(0.367)		(0.006)	(0.022)
Academic background, age	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	275	275	277	277	275	275	277	277	274	275	274
pseudo R^2					.04296	.07748	.03806	.0594			
log pseudo-likelihood	-562.68	-559.51	-533.14	-529.67					-754.73	-750.14	-747.48

Marginal effects; p-values in parentheses, (d) for discrete change of dummy variable from 0 to 1, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

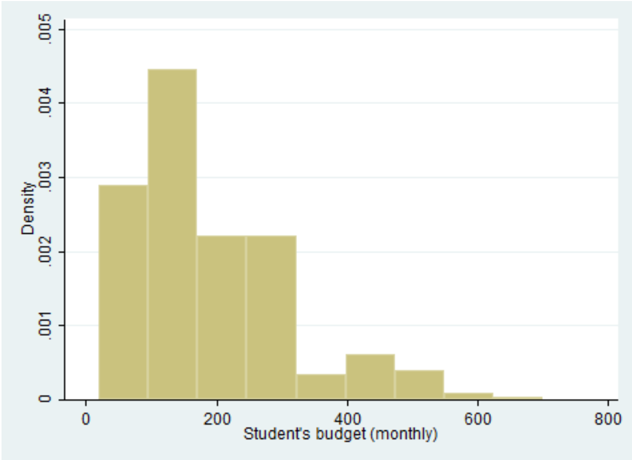


Figure 3.5: Disposable income distribution in euro

Results are reported in table 3.5. Starting from the risk attitude, we find that the relation between cognitive skills and risk aversion is still significant and negative, both in the first table with lower payoff and in the second with higher payoff. Age has very little effect and this is probably due to the restricted range of age in our sample (DFHS instead find a positive and highly significant U-shaped relationship). Income seems to be negatively correlated to risk aversion but the estimated coefficient is never significantly different from zero (this seems however to confirm the result in DFHS). The gender dummy do not have a clear effect since in the first table (low payoff) the sign is positive, meaning that male subjects are more risk averse while in the second it is reversed. However, in no one of the estimation the coefficient is significantly different from zero, which is also the case in DFHS.

The inclusion of the disposable income, age and gender do not change our results on the relationship between numeracy skills and the risk neutral choice. Here the numeracy skills of individuals confirm to be the most important element to predict a switch precisely in correspondence to the row in which the expected value of the lottery is equal to certain amount. This confirms the idea that, on top of cognitive skills a key factor for observing a risk neutral choice, is the ability to compute expected values.

Finally, the relation on patience qualitatively as the one we estimated without the new variables. Numeracy skills are positively related to patience while cognitive skills, captured by our symbolic test are negatively related to patience. Neurocitism remains the main explanatory factors, with a positive and significant relationship with patience, which confirms the findings of other paper in the literature, see for example Anderson, Burks, DeYoung,

do not report the estimated coefficients of the academic dummies.

and Rustichini (2011). In our sample, male subjects tend to be less patience with respect to female but the coefficient is borderly significant. Surprisingly, subjects with higher disposable income tend to be less patience and prefer to have immediately the telephone credit. This result is not in line with previous studies such as DFHS.

Summarizing, the relationship highlighted in section 3.3 are robust to the inclusion of additional controls, confirming the idea that cognitive abilities and personality traits are determinant of the attitude toward risk and impatience.

3.5 Conclusion

In this paper we have investigated the relationship between both cognitive abilities and personality traits with fundamental elements of economic choice as risk aversion and impatience. We have confirmed some of the results already emerged in the literature on the negative relationship between cognitive skills and risk aversion and on the positive relationship between these skills and impatience. Other than that we have produced evidence for a qualitative interpretation of cognitive skills since we have shown how numeracy skills do affect risk taking behaviour in the direction of increasing the likelihood of risk neutral behaviour. With respect to personality trait, we have shown the importance of emotional stability in determining the degree of patience of subjects.

The paper contributes also to the experimental economic literature since it shows that online experiments are a valid instrument for economic analysis and can be used in place of laboratory ones which suffer the limit of a limited number of participants. At the same time we provide evidence that impulsivity of participants in performing the assigned task may be a factor that largely affects outcomes in experiments and much can be gained reducing it or, when not possible, controlling for it in the empirical analysis.

Results coming from this study have important policy implications, from the perspective of improving human capital and economic development, since it is proved the role of cognitive skills on risk attitude and patience. Entrepreneurship for example requires the willingness to take calculated risks and patience to develop projects and to innovate. Hence, investments oriented to the the accumulation of skills can have high returns in terms of higher economic activity. Our findings are also important for contract design and screening since cognitive skills and personality traits, which are more easy to observe than risk and time preferences, can be useful information to design optimal incentives in contracts.

3.6 Appendix

3.6.1 Cognitive and numeracy test

The test we employ to measure cognitive abilities in the same as in Lang, Weiss, Stocker, and Rosenbladt (2007) which has been proved by the authors to be highly correlated with IQ. The test is designed to be run online and it consists in memorizing symbols and inputting them in a small table for 30 seconds. The user has one trial stage to practice the test and then she has to do the three main stages. In figures 3.6 and 3.7 we reported two screenshots of this test.

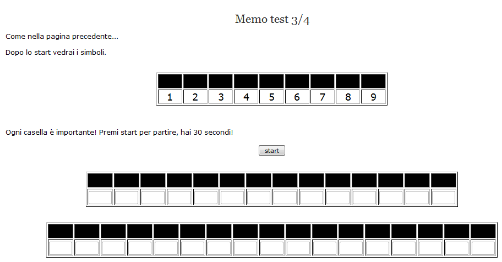


Figure 3.6: Before starting the countdown

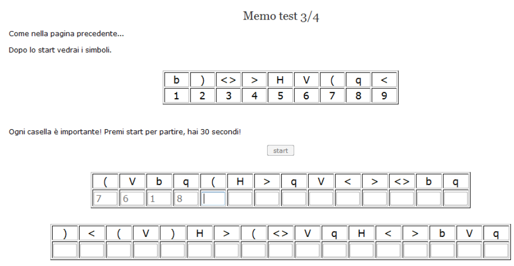


Figure 3.7: After starting the countdown

As we can see from the figures, the user can start the test pressing the start button. After the start she has 30 seconds to fill the cells with the number corresponding to the symbol displayed. When a number is selected the following cell is automatically activated so the user is focused only on remembering the numbers the inputting them matching the symbols. When the time elapses the user is redirected to the next stage.

The aggregate measure we construct is the sum of the performance in the three stages and the results we got in the online implementation is comparable with exactly the same test run in the laboratory of the School of Economics in Bologna. The subject pool in this case was composed by undergraduate students taking a course in econometrics. The only difference we observe is that in the population that run the test online we observe an higher non response rate (users who did not input any number and waited that the 30 second elapsed). Out of 312 users who terminated the test, 11 decided not to do it while in the other case we have only one non response. If we exclude these observations, in both cases we do not reject a test for normality (as we expect from a test related to IQ). Moreover, the mean value of the test in the online population is 34.058 while in the laboratory is 33.924 and Mann-Whitney-Wilcoxon test does not rejects the null hypothesis of equal means between the two samples (p-value=0.3858). This is an indirect confirmation of the quality the data collected online with respect to the one collected in the lab.

The numeracy trait is elicited through a battery of questions:

A. If in a population the probability of contracting a disease is 10 percent, how many people out of 1000 will get it?

B. A shop is selling a sofa at half of the original price. If the current price is 150 euro how much is the sofa without the discount?

C. A car shop is selling a second hand car at 6000 euro which is 2/3 of the original price of the car. How much was the car when it was new?

D. Suppose you have 2000 euro in your bank account. The interest rate that the bank is applying to your savings is 10% per year. How much will you have in two years from now?

The distribution of scores is reported in figure 3.8.

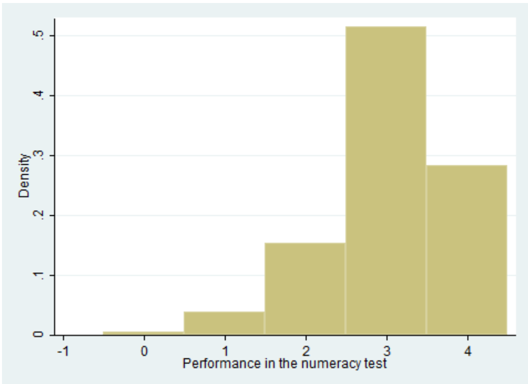


Figure 3.8: Distribution of scores in the numeracy test

3.6.2 Estimation results on the relation between risk aversion and personality traits

We report here the results of the estimations on the relation between risk aversion and personality traits. Coefficients are almost all not significantly different from zero. This is in line with Dohmen, Falk, Huffman, and Sunde (2010).

Table 3.6: Relation between risk aversion and personality traits

Dependent variable	Number of rows before switching to option B											
	First table						Second table					
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(L)	(M)	(L)
Neurocitism	0.030				0.073	0.058	-0.064				-0.081	-0.079
	(0.800)				(0.556)	(0.635)	(0.564)				(0.493)	(0.499)
Conscientiousness		0.134			0.176	0.181		0.214**			0.189	0.189
		(0.247)			(0.147)	(0.137)		(0.043)			(0.108)	(0.109)
Internal loc			-0.061		-0.124	-0.127			0.120		0.101	0.101
			(0.651)		(0.411)	(0.403)			(0.281)		(0.416)	(0.414)
External loc				0.032	0.035	0.040				0.011	0.048	0.048
				(0.791)	(0.788)	(0.764)				(0.916)	(0.671)	(0.676)
Impulsivity						0.103						-0.012
						(0.333)						(0.894)
N	276	276	276	276	276	276	278	278	278	278	278	278

Marginal effects; p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Relation between risk neutrality and personality traits

Dependent variable	Probability of switching at 6th or 7th row											
	First table						Second table					
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(L)	(M)	(L)
Neurocitism	0.067**				0.043	0.031	0.005				0.001	-0.009
	(0.021)				(0.165)	(0.316)	(0.859)				(0.980)	(0.774)
Conscientiousness		-0.011			-0.032	-0.030		-0.060**			-0.066**	-0.065**
		(0.695)			(0.287)	(0.324)		(0.046)			(0.034)	(0.041)
Internal loc			-0.061		0.052	0.051			0.003		0.027	0.026
			(0.651)		(0.117)	(0.115)			(0.925)		(0.431)	(0.439)
External loc				0.032	-0.039	-0.036				0.010	0.009	0.013
				(0.791)	(0.212)	(0.251)				(0.729)	(0.766)	(0.691)
Impulsivity						0.091***						0.070**
						(0.001)						(0.013)
N	276	276	276	276	276	276	276	276	276	276	276	276
pseudo R^2	.01537	.00041			.03067	.06305	8.2e-05	.01082	2.3e-05	.00031	.01256	.0293

Marginal effects; p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6.3 Instructions

Since the experiment is performed online and readers may be not willing to read long paragraphs, we divided the instructions in different pages with a specific content and let the reader to move forward and backward in the navigation. At the end of the instructions we report the questions which every subject must answer correctly to proceed to the trial stage. Instructions are translated from Italian.

Welcome page Welcome, this activity is aimed to study some aspects of economic decision.

- We guarantee your privacy and that the objective of the study is only scientific research
- Participating to the activity, you have the possibility to receive an amount of money or telephone credit as a payment for you participation
- This payment will depend upon your choice in the activity. For more details see the following pages.
- The activity lasts more or less 20 minutes. Hence, we ask you to start only if you have the time to finish, otherwise you can do it a more suitable moment.
- The activity is divided in two parts and can be done only once. The first part involves money and the second telephone credit of similar value. The instructions of the first part are in the next pages while those of the second will be presented after the end of the first part.

Page 1
Instructions
To perform this activity you must read the following instructions.
It is very important to read them carefully since the payment will depend on your decisions and it is not possible to redo the activity or going back to a previous stage during the activity.
It may be useful to print the instructions. However, do not worry, it is not too complicated!

Introduction The activity is aimed to study some aspects of economic decision.
Briefly, you will be asked to fill some tables (plus two trial ones that will help you to familiarize).
In each row of these tables you will have to select the option you prefer among two, labeled option A and option B.

ATTENTION! Read the instructions carefully. At the end you will take a short test to verify if you understood the content of the activity.

Page 2

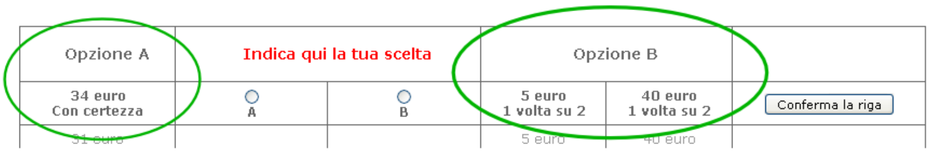
INSTRUCTIONS – AN EXAMPLE

An example of choice between the two options is the following. Suppose you are asked to choose what you prefer between: having 10€ with certainty (option A) or having 15 euro is the result of a coin flip is head versus 1€ if the result of coin flip is tail (option B).
You are asked to answer to this kind of question of the basis of your preferences.
It is important that you will select the option on the basis of what you prefer because, at the end of the activity, if you will be selected for receiving the payment, one of your choices (randomly selected) will be your actual payment.
Select on the basis of what you prefer since in case of being payed you will determine your payment with your decisions.
In the next screen you can see graphically who options will be displayed. . .

Page 3

INSTRUCTIONS FOR THE FIRST PART OF THE ACTIVITY

An example of choice between the two options is reported below:



As you can see, Option A is placed on the left of the table and option B on the right (green circles). Moreover, notice that:

- **Option A:** is certain payment. If you select this option and this row is selected for the payment, you would receive this amount of money. In this case the amount corresponds to 34€.
- **Option B:** is an uncertain payment, and it offers you the possibility to obtain a higher amount of money with some probability and a lower amount of money with another probability. In this example option B contemplates to receive 5€ in one case out of two or 40€ in one case out of two.

Important: in all the tables option A will always be a certain amount of money while option B an uncertain amount of money. Moreover, option B will always be characterized by the same probabilities for the two outcomes: one case out of two (which is equivalent to a coin flip).

Once you have evaluated which option you prefer in the row, to make your choose is simple: in the table you have two buttons associated to the two options and one button to confirm the choice on the right of the row.

In the figure you can see these elements circled in purple and red:

Page 4

INSTRUCTIONS OF THE FIRST PART OF THE ACTIVITY

The table must be filled one row at the time, from the top to the bottom. To go from one row to the next you must make your choice and confirm it using the buttons presented in the previous page.

Important: The table is such that moving to the bottom, option B is always the same and option A decreases in the payment.

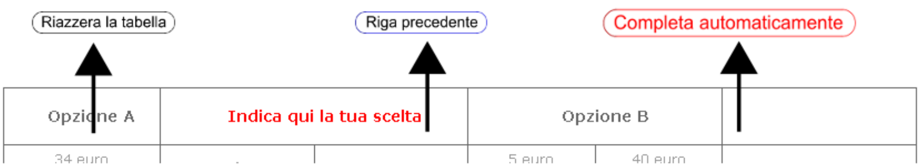
To fill the table you can also use these buttons:

- **Restart the table:** with this button you can refresh the entire table and restart.
- **Previous row:** with this button you can go back to the previous row of the table.

A third button is the **automatic completion button** which allows you to let the program filling the table and skip to the next one **when there is the possibility to do so**. If, in a certain row you select option B, and in the following rows the same option B is confronted with other options A which are worse than the one you did not select, then you have the possibility to avoid to select manually the option B in all the remaining rows but to let the computer doing it automatically and skip to the next table.

This button, when becoming available, will be announced by a pop-up window.

These buttons are placed in the upper part of the screen as the figure below shows:



Page 5

THE PAYMENT

The amount of the payment for this activity is high. Hence, 1 participants out of 5 will receive a payment.

The procedure contemplates to make an extraction by the computer of a number between 1 and 5 and to give the payment to those who will extract the number 1. We guarantee all participants not to manipulate the extraction since the funding for the project is high enough to cover the expenses. However, for transparency reasons, you have the possibility to come to the gym to perform a physical extraction there (write us an e-mail to settle the appointment).

For those who will extract the number 1, the procedure to determine the actual payment is the following (it will be simulated also in the trial tables):

- The computer will extract one of the rows of the tables at random. NOTE: the row extracted may be one row of the second part of the activity.
- If the selected option in the row is the certain amount (option A) it is not necessary another extraction
- If the selected option is the uncertain amount (option B) then another extraction will be performed to determine which of the two outcomes is the one selected for the payment. The rule in this case is that is 1 is extracted then it is the lower amount and if 2 is extracted the payment is the higher amount.
- Payment for the first part of the activity are available at the gym from the first Monday after completing the activity.

NOTE: the rule of the Monday after completing the activity is valid for this first part (hence, if the row extracted is one of this part). If instead the payment extracted is the telephone credit, then the payment will be granted automatically.

FINAL NOTE: It is not possible to do the activity twice or to go back using the navigation buttons of the browser. For this reason we suggest to do the activity carefully. If you close the browser during the activity, then you can restart connecting again. If this happens in a trial stage you will restart from the instructions. If this happens in a payed table you will restart from the beginning of that table.

Control questions

A. What is the content of your choice in the activity?

1. I have to choose among the 2 option my preferred payment
2. I have to choose at random a payment

B. Suppose to be faced by the following possibilities:
Option A: 10€ with certainty
Option B: 5€ in one case out of 2 or 20€ in one case out of 2
Suppose you select option A in this row. What happens if this row is extracted for your final payment?

1. I will receive 10€ as a payment for the activity
2. I will receive 20€ as a payment for the activity
3. I do not know yet, another extraction must be performed to know the result of option B

C. Consider the previous case and suppose now that you choice is option B. What happens in this case?

1. I will receive 10€ as a payment for the activity
2. I will receive 20€ as a payment for the activity
3. I do not know yet, another extraction must be performed to know the result of option B

D. Is it possible that two participants obtain different payments?

1. No, the payment for the activity is the same for everybody
2. Yes, because they may choose different options and also the final extraction may lead to different results
3. Yes because the payment is completely random

INSTRUCTIONS OF THE TIME EXPERIMENT

Page 1 INSTRUCTIONS OF SECOND PART OF THE ACTIVITY

You have started the second part of the activity.

What you have to do in this part is not much different from the previous since you have to fill a table choosing in each row among the two options.

What you are asked to choose between are different amount of telephone credit granted at different dates.

One example is the following: what do you prefer between 10€ of credit today or 20€ in a month?

Page 2

INSTRUCTIONS OF SECOND PART OF THE ACTIVITY

An example is in the following figure: As you can see, option A is on the left of the table

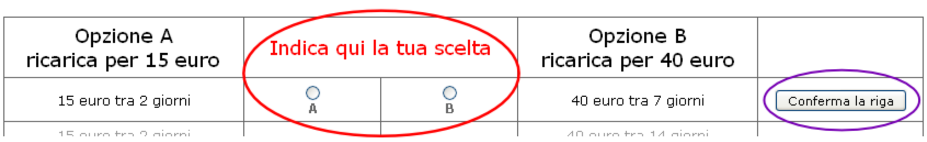
and option B is on the right. Moreover,

- **Option A:** is a lower amount of credit (here is 15€) but is granted in two days
- **Option B:** is a higher amount of credit (here is 4010€) but is granted in a longer time

IMPORTANT: in each table (the trial one and the payed one) option A will always be the same both in terms of the amount of credit and in terms of the time to wait to obtain it. Instead, option B will change at each row only in terms of the time to wait to be granted. The amount of the telephone credit remains the same.

Once you have evaluated the two options in the row you can make your choice easily, also in this case you have two buttons and the confirmation button on the right of the row.

In the figure they are circled in red and purple:



Page 4

INSTRUCTIONS OF THE SECOD PART OF THE ACTIVITY

The table must be filled one row at the time, from the top to the bottom. To go from one row to the next you must make your choice and confirm it using the buttons presented in the previous page.

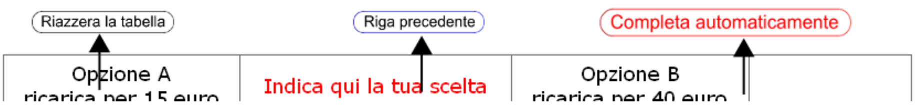
To fill the table you can also use these buttons:

- **Restart the table:** with this button you can refresh the entire table and restart.
- **Previous row:** with this button you can go back to the previous row of the table.

A third button is the **automatic completion button** which allows you to let the program filling the table and skip to the next one **when there is the possibility to do so**. If, in a certain row you select option A, and in the following rows the same option A is confronted with other options B which are worse than the one you did not select, then you have the possibility to avoid to select manually the option A in all the remaining rows but to let the computer doing it automatically and skip to the next table.

This button, when becoming available, will be announced by a pop-up window.

These buttons are placed in the upper part of the screen as the figure below shows:



Chapter 4

Teams or Tournaments? A Field Experiment on Cooperation and Competition in Academic Achievement

Abstract

This paper assesses the effect of two stylized and antithetic non-monetary incentive schemes on students' effort. We collect data from a field experiment where incentives are exogenously imposed, performance is monitored and individual characteristics are observed. Students are randomly assigned to a tournament scheme that fosters competition between coupled students, a cooperative scheme that promotes information sharing and collaboration between students and a control treatment in which students can neither compete, nor cooperate. In line with theoretical predictions, we find that competition induces higher effort with respect to cooperation and cooperation does not increase reduces effort with respect to the baseline. However, this is true only for men, while women do not seem to react to non-monetary incentives.

Keywords: education, field experiments, incentives, competition, cooperation
JEL codes: A22, C93, I20

4.1 Introduction

In the recent years a large debate has focused on the possible ways to improve schooling achievement at every level of education. The relevance of this goal is not disputable, since education contributes to the accumulation of human capital, the development of societies and it is considered as one of the main channels for the reduction of inequality. In this paper we study the effect of different incentive schemes on the schooling achievement of a sample of students enrolled in a undergraduate course at the University of Bologna (Italy). More specifically, we focus on the effect of non-monetary incentives on academic performance running a field experiment where randomized groups of students have been assigned to different incentive schemes.

The design of the field experiment is based on a theoretical model that contemplates three different incentive schemes. As a benchmark we consider the effect on effort of a piece rate reward. Then we analyze two alternatives: a tournament that fosters competition among matched students and a cooperative scheme in which they can share information and collaborate. The model suggests a weak ordering between the three: in a competitive environment individual performance should be weakly higher than in the benchmark and effort under the benchmark should be weakly higher than in the cooperative scheme. We also show that the detrimental effect of cooperative incentives on effort does not depend on the specific shape of the distribution of types in the population, while the magnitude – but not the sign – of the effects of a competitive incentive scheme depends on the shape of this distribution. To test these theoretical predictions, we randomly assign students to the treatments and we adopt a between-subjects design, i.e. each subject is only exposed to a single incentive scheme.

Data confirm the theoretical predictions in the full sample. Moreover, we show that an important difference emerges between genders: promoting competition appears to have a strong positive effect on the exerted effort only for males. In contrast, promoting cooperation reduces effort with respect to the case where students can neither compete nor cooperate, but this effect is not statistically significant for both genders. These findings are in line with the literature on how competition affects behaviour depending on gender (see for example Gneezy, Leonard, and List 2009) and provide an interesting comparison with respect to the result of Angrist and Lavy (2009) who find that monetary incentives improve performance especially on girls. We depart from this branch of the literature, complementing the results obtained through monetary incentive, by focusing on non-monetary ones since they represent a relatively cheap way to increase student's effort[1].

[1]Studies on monetary incentives proved to be successful in improving students' performance but the cost of

The paper proceeds as follows. After a brief review of the related literature (Section 1) we describe and discuss in detail our experimental design (Section 2). In Section 3, we present a simple model, and derive the theoretical predictions which will serve as a reference for the analysis of the experimental data, presented in Section 4. Section 5 concludes, and presents possible extensions of this research.

4.2 Related Literature

We provide a brief review of the literature to highlight our contribution with respect to previous theoretical and empirical work.

On the general issue of how to foster students' effort and school achievement through explicit incentive schemes, several papers explore the role of pecuniary-based incentives. Among those, Blimpo (2010) represents the closest study to our experiment. Analysing data from a field experiment in Benin with a pool of 100 secondary schools, he studies whether individual or different kind of team incentives can lead to a higher students' school performance. He considers three treatments. In the first treatment, each student obtained an individual monetary reward if and only if his or her performance exceeded a minimal threshold at the final exam. In the second treatment, participants were randomly assigned to teams of four students and each team-member received a monetary reward depending on the average team performance, if and only if all the team-members achieved a target performance level. Finally, in the third treatment, participants were randomly assigned to teams of four students but in this case only the components of the three top-performer teams were awarded with a monetary prize. Blimpo (2010) finds that the individual based incentive scheme with cut-off target is most effective for students at an intermediate performance level: at the lower tail of the skills distribution, students reduce effort, probably because they perceive the target out of reach; at the higher tail of the distribution, students know that they are able to get the prize without any extra effort, thus the average impact of such incentives is smaller. When teams are evaluated according to the average performance of the group conditionally on the achievement of a minimal performance target (2nd treatment), students across all levels of ability are positively affected: the effort exerted by the different team-mates is pushed toward the target. The tournament scheme (3rd treatment) yields the most beneficial effects: it induces all the teams to work harder as students exposed to this treatment do not have any prior information about the quality and the skills of their

inducing higher effort is not negligible. In a study conducted in the New York City school system \$600 have been awarded for each passing grade, the Baltimore City Public School District has paid up to \$110 to improve scores on state graduation exams and similar programs in the US award up to \$500 for each exam passed.

competitors in the other teams.

Recent papers consider tournaments at school with financial rewards. Kremer, Miguel, and Thornton (2009) focus their study on the evaluation of a merit scholarship programme dedicated only to female students in an elementary school in Kenya. They observe a substantial increase in the exams scores: in particular girls with low pre-test scores, who were unlikely to win a scholarship (and actually did not get it), reported positive and significant gains in terms of higher school performance. De Paola and Scoppa (2010) studied the effectiveness of monetary incentive schemes in enhancing students' performance using a randomized experiment involving undergraduates in an Italian University. Students participating in the experiment were assigned to three different groups: a high reward group, a low reward group and a control group. Rewards were assigned according to a ranking rule to the top performing students in each treated group. The authors report that financial rewards contributed to increase the students' performance: a very strong reaction emerged among high ability students who were likely to win the contest, while no significant effect was observed for low ability students that have fewer chances to win the tournament competition. Along the same lines, Leuven, Oosterbeek, and van der Klaauw B. (forth.) present results of a randomized field experiment in which freshman students at the Amsterdam University had the opportunity to earn financial rewards for passing all first year requirements. Their findings provide evidence that high ability students perform significantly better when assigned to rewarded groups. On the contrary low ability students' outcome decreases if assigned to rewarded groups. The small aggregate average effect that they observe is therefore the sum of a positive effect for high ability students and a negative off-setting effect for low ability students. These previous results highlight the importance of controlling for students' ability and individual characteristics when assessing the impact of incentive schemes on their school performance.

Among the authors who studied the effects of financial incentives, some focused focused specifically on gender differences. Angrist and Lavy (2009) evaluate the effectiveness of financial rewards on the achievement of Israeli students using a randomized experiment providing monetary awards to students who obtain the university admission. The authors show how the program led to significant effects for girls but not for boys. Differences in gender-scheme interaction emerge also from the field experiment by Angrist, Lang, and Oreopoulos (2009). In this study, researchers randomly assigned a sample of students enrolled in a Canadian university to one of three different treatments: the first group was provided with a set of support services (e.g. tutoring); the second group was offered financial rewards for good academic scores; the third one was offered a combination of support services and monetary incentives according to the academic performance. The results of the experiment show that while males did not react to any of the treatments, females improved significantly their academic performance when monetary incentives were provided.

While females appear to react more than males to monetary incentives awarded for achieving an exogenously given target, incentive schemes based on competition may yield opposite effects. Gneezy, Niederle, and Rustichini (2003) found that males are more prone to engage in competition than females and in general males' performance increases more than the females' one when subjects are exposed to a competitive setting. Similarly, Niederle and Vesterlund (2007) find that, when given the opportunity to choose between a piece-rate payment scheme or a tournament, men select the tournament twice more frequently than women, suggesting that women tend to avoid competition when they have the chance to do so. Azmat and Iriberri (2010) find that, even when the incentive scheme is based solely on the subject's performance, providing information about the *relative* performance promotes higher levels of effort among men, but not among women. We explore the role of gender, and we find that males tend to respond to incentives as predicted by the theory, while females do not.

From a theoretical standpoint, Bratti, Checchi, and Filippin (2008) proposed a model of student cooperation/competition in learning activities, showing that free riding opportunities lead to an insufficient degree of cooperation between schoolmates, which in turn decreases the overall achievement of the group. According to their analysis, a cooperative learning approach may successfully emerge when the class is homogeneous in terms of students' ability. In our study we consider an experimental design and a theoretical model where the incentive scheme is exogenous but similarly to Bratti, Checchi, and Filippin (2008) we focus on student cooperation/competition in learning activities. Our theoretical model suggests that in a competitive environment individual performance should be higher than in the cooperative environment. We also show that the detrimental effect of cooperative incentives on effort is relatively stronger for high-ability types than for low-ability types, and that the magnitude – but not the sign – of the effects of a competitive incentive scheme depend on the shape of the distribution of types in the population.

4.3 The Experimental Design

The experiment involved all the undergraduate students enrolled in the Introductory Econometrics course of the major in Management Studies at the University of Bologna, in year 2010.[2] The course lasted 10 weeks (a three-hour-lecture per week). Students participating to the experiment had to undertake 5 tests whose marks were translated into bonus points

[2]The University of Bologna is considered the oldest University in Europe and counts on average nearly 8000 enrolled students each academic year.

for the final exam. The bonus points for the final exam were equal to the average mark the student obtained in the five tests.[3]

Tests have been scheduled every two weeks and each test consisted of five multiple-choice questions to be answered in 50 minutes. Each test concerns all topics taught in the course until the last lecture before the test.

Tests were computerized[4], and were held in the computer laboratory of the School of Economics of the University of Bologna. Desks were arranged so to minimize the possibility for students to talk during the exams (see Figure 4.3 in Appendix).

The mark in each test consisted in an individual component, based on the number of correct answers in the test, and a number of extra points related to the treatment and possibly to the performance of the partner.

Our study included two treatment conditions – characterized by a competitive and by a cooperative incentive scheme, respectively – and a control treatment. In all treatments including the control, part of the incentive depended solely on individual effort. Treatments differed in how tests two, three and four were performed, while the first and the last test were identical across treatments. The first and the last tests were taken individually by each student. In contrast, in the second, third and fourth tests students in the two treatment conditions were randomly matched in couples at the beginning of each test, and had the opportunity of exchanging messages with their partner via a controlled chat program, running on their computer. In both treatment conditions, the total score in tests 2, 3, and 4 of the test depended not only on the student's individual performance (i.e. the net score), but also on the partner's performance. Table 4.1 summarizes the treatments, which are described in detail below.

Students were assigned to treatments between the first and the second test. Before starting tests 2, 3, and 4, students assigned to the two treatment conditions were asked whether they wanted to use the chat or not to communicate with the paired partner. This decision was taken simultaneously by all students. During the test, a couple of students could use the chat program only if both students declared to be willing to communicate, at the beginning of the test. If the two students chose to communicate, per each of the questions of the test they could send only one "signal" to indicate what the right answer was, and one short text message of up to 180 characters. Interactions were anonymous, as students could not know the identity of their partner. In the control treatment no interaction between students was allowed.[5]

[3]Marks in the final exam range from 0 to 30. The exam is passed with a mark equal or above 18. The bonus points ranged from 0 to nearly 4.

[4]The experiment was programmed and conducted with the software z-Tree (Fischbacher (2007)).

[5]Figure 4.6 presents a screen-shot of the graphical interface of the program used for the tests On the left-

Table 4.1: Summary of the treatments, in tests 2, 3 and 4

treatment	**extra points (rounds 2, 3, 4)**	**messages available**
CONTROL	1	no
COOPERATIVE	$\mathrm{I}(s_j^k \geq 1.5)$	yes
COMPETITIVE	$2 \cdot \mathrm{I}(s_i^k > s_j^k)$	yes

In each test, the value p^q of correct answers to each question q ranged between 0.3 and 1.2 points. Across all treatments, the number of points v_i^k a student could get by correctly answering the questions of test k was:

$$v_i^k = s_i^k \cdot \mathrm{I}\left(s_i^k \geq 1.5\right), \quad s_i^k = \sum_{q=1}^{5} p_i^{q,k}, \quad k = 1, \dots 5$$

In each test, the maximum number of points $\bar{v}$ was equal to 3. This is the individual part of the mark in the test, i.e. the component which is common across all treatments.

In the COMPETITIVE treatment, student i's mark in a test was increased by 2 extra points if she performed strictly better than the partner. The k-th test's mark $\hat{v}_i^k$ for student i under this incentive scheme is described in equation (4.1).

$$\hat{v}_i^k = v_i^k + 2 \cdot \mathrm{I}\left(s_i^k > s_j^k\right), \quad k = 2, 3, 4 \tag{4.1}$$

This provides an incentive for both matched students to compete.

Conversely in the COOPERATIVE treatment, student i's mark in a test was increased by 1 extra point if the partner's performance was sufficiently good. The k-th test's mark $\hat{v}_i^k$ for student i under this incentive scheme is presented in equation (4.2).

$$\hat{v}_i^k = v_i^k + \mathrm{I}\left(s_j^k \geq 1.5\right), \quad k = 2, 3, 4 \tag{4.2}$$

Finally, students in the CONTROL treatment received 1 extra point in tests 2, 3 and 4.[6]

Time-line of the experiment. The experiment started in February 2010, and ended in July of the same year. In the first lecture of the course, on February 25th, the full set of instructions was distributed to students and each student had two days to decide whether to

hand side of the screen students could read the question, and the multiple-choice answers. On the top-right part of the screen they could send messages to their partner, while on the bottom-right part of the screen they could read the messages possibly sent to them by the partner.

[6]This is done so that the maximum number of bonus points per team is constant across treatments.

take the partial exams or not. At this stage, students were not explicitly informed that they were taking part in an experiment and only at the very end of the course, participating students were asked to sign a consent form authorizing the treatment of data collected during the partial exams.[7]

On March 1, during a standard class, students were asked to fill in a questionnaire collecting data about some personal characteristics (age, gender, familiarity with computers, e-mail and chat programs, mother and father education). Questionnaire answers are used in the econometric analysis to control for individual-specific characteristics.[8]

On March 22nd students took the first test. Notice that at this stage students had not yet been assigned to treatments, so the grade in this first test can be used as a measure of their performance before being exposed to the treatment. Students received information about what treatment they had been assigned to only three days later, on March 25th.[9] In the same day, students were informed about their own result in the first test, and about the distribution of the first test score among participants. In this way we tried to convey common knowledge of the distribution of competences and ability in the population. Section 1.2 will show how this is relevant from the theoretical point of view.

The remaining four tests were taken approximately every two weeks, in April and May 2010 with the exception of the fifth which was administered one week after the fourth.[10] Student could benefit of the bonus points gained in the tests only if they took the final exam in June or July 2010. On March 22, before the experiment started, students were informed that the bonus points would expire after the summer.

4.4 The Model

This section describes the main features of the model we use to derive theoretical predictions and inform the experimental design. After briefly characterizing the general features of the model, we illustrate its implications in terms of expected effort under the different incentive schemes. We first describe what happens without competitive or cooperative in-

[7] The experiment was authorized by the ethics committee of the the University of Bologna *(Comitato Bioetico per la Valutazione di Protocolli di Sperimentazione)*.

[8] An overview of the answers to the questionnaire is provided in Section 4.5, and a translation of the questions is reported in Table 4.10 in Appendix.

[9] Students taking part in our experiment were then randomly assigned to two groups of about 65 people each, because the computer lab can host only up to 80 students at a time. All students assigned to the competitive treatment and half of those assigned to the control treatment were in the first group, while all students in the cooperative treatment and the remaining students of the control treatment were in the second group.

[10] This is made on purpose since the last test is taken by students individually and covers the last contents of the program as well as some of the previous ones. Hence, it will reflect the effort exerted in the previous stages.

centives (BASELINE treatment). We then characterize the optimal effort under incentives to cooperation and to competition and finally we highlight the testable predictions of the model.

General features We assume that students' abilities are in the interval $\theta \in [0,1]$ and are distributed according to a non-degenerate distribution function $F(\cdot)$. Students choose a level of effort $e_i \in [0,1]$, which determines their performance in the tests and in the final exam. The dis-utility from effort is $c(e_i)$. We further assume that $c(\cdot)$ is independent on subjects' ability θ_i, and that $c'(\cdot) > 0$ and $c''(\cdot) > 0$.

The expected score in test k is a function of ability and effort and is given by the following expression:

$$\mathrm{s}_i^k = e_i \cdot \theta_i \cdot \bar{v} \tag{4.3}$$

The utility of each student is positively affected by the score and negatively affected by the effort. We assume that students choose their level of effort two times: the first time they choose $e_{i,0}$ when the course starts, before the first test and before the assignment to the treatments; later, after having been assigned to treatments they choose the level of effort e_i that determines their performance in tests 2 to 5 and in the final exam. At this point, their expected utility is given by 3 components: the bonus points obtained in the four remaining tests to be taken – which in the two treatment conditions is the outcome of the interaction with the matched agent – the individual mark in the final exam[11] and the cost of effort. Under the assumption of risk neutrality, the expected utility at the time in which e_i is chosen is:

$$E[U_i] = \frac{1}{5}\sum_{k=2}^{5}\int_0^1 \hat{v}_i^k(\theta_i, e_i, \theta_j, e_j) \cdot f(\theta_j)\mathrm{d}\theta_j + \bar{V} \cdot e_i \cdot \theta_i - c(e_i) \tag{4.4}$$

where $\bar{V}$ is the maximum mark in the final exam.

Baseline treatment A student assigned to the baseline treatment does not interact with any other student. As a consequence, considering the four tests and the final exam, the expected utility (4.4) simplifies in:

$$U_i^{BL} = \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5} \cdot (4 \cdot e_i \cdot \theta_i \cdot \bar{v}) + \frac{3}{5} - c(e_i) \tag{4.5}$$

from this utility function we can derive the optimal effort exerted:

$$\frac{\partial U_i^{BL}}{\partial e_i} = (\bar{V} + \frac{4}{5} \cdot \bar{v}) \cdot \theta_i - c'(e_i) \tag{4.6}$$

[11] Remember that the bonus adds points on top of this mark.

Normalizing the quantity $V + \frac{4}{5} \cdot \bar{v} = 1$, we get the baseline effort:

$$c'(e_i^{BL}) = \theta_i$$

that implies

$$\frac{\partial e_i^{BL}(\theta_i)}{\partial \theta_i} > 0$$

i.e., we expect more able individuals to exert more effort in the baseline treatment with respect to less able individuals and no variation over the optimal choices of effort.

Competitive treatment To model student's behavior under the two treatments and to derive predictions, we look for the equilibrium in the bayesian-Nash games where students have private information about their own type and a common knowledge on the distribution of abilities in the population.

Under the competitive scheme, students get bonus points if their performance is better than the partner's. Equation (4.7) describes the expected utility in this case.

$$\begin{aligned} U_i^{comp} &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{3}{5} \cdot 2 \cdot Pr(e_i \cdot \theta_i > e_j \cdot \theta_j) - c(e_i) = \\ &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{6}{5} \cdot \int_0^{\theta_i \cdot \frac{e_i}{e_j}} f(\theta_j) d\theta_j - c(e_i) = \\ &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5}\left[4 \cdot e_i \cdot \theta_i \cdot \bar{v}\right] + \frac{6}{5} \cdot F\left(\theta_i \cdot \frac{e_i}{e_j}\right) - c(e_i) \end{aligned} \tag{4.7}$$

where $6/5 \cdot F(\theta_i \cdot e_i/e_j)$ is the expected number of additional points obtained in the second, third, and fourth test in case the student outperforms his partner. Hence, the expected utility can be expressed as:

$$e_i(\theta_i) \in \arg\max_{e_i} \left\{ E[U_i] = \bar{V} e_i \theta_i + \frac{1}{5}\left[4 e_i \theta_i \bar{v}\right] + \frac{6}{5} \int_{\theta_j | \theta_j e_j < \theta_i e_i} f(\theta_j) d\theta_j - c(e_i) \right\} \tag{4.8}$$

Under regularity assumption on the distribution of types in the population, it can be shown that the first order conditions are[12]:

$$\theta_i - c'(e_i) + \frac{6}{5} f(\Phi_j(e_i)) \Phi'(e_i) = 0 \tag{4.9}$$

[12] In order to have a pure strategy Nash equilibria, the distribution function of types must be non-degenerate and the mapping from type to effort must be continuous and increasing. The requirement on the distribution of types is a plausible requirement, given the heterogeneity in the population while the two on the mapping between type and effort can be proven to be true in our case. In the non-heterogeneous case, that is when the distribution of types is degenerate, it can be easily shown that no pure-strategy equilibrium exists.

where Φ_k is the mapping from the effort to the type (individual ability). Now, since $\Phi' = 1/e'$, we have the following solution for the optimal effort in the competitive treatment:

$$c'(e_i) = \theta_i + \frac{6}{5} f(\theta_i) \cdot \frac{1}{e_i'} \tag{4.10}$$

From this equation we see that the optimal effort exerted under this scheme is equal or higher than the optimal level of effort e_i^{BL} in the control treatment. The magnitude of the effect depends on the subject's ability θ_i and on the shape of the distribution $F(\cdot)$.

Cooperative treatment Under this scheme, each student has a clear incentive to share her information (in tests 2, 3 and 4) and the mark depends also on the partner's effort.
In this case the expected utility becomes:

$$\begin{aligned} U_i^{coop} &= \bar{V} \cdot e_i \cdot \theta_i + \frac{1}{5} \cdot [e_i \cdot \theta_i \cdot \bar{v}] + \\ &+ \frac{1}{5} \int_0^1 3 \cdot [\bar{v} \cdot (e_i \cdot \theta_i + e_j \cdot \theta_j - e_i \cdot \theta_i \cdot e_j \cdot \theta_j) + \\ &+ \mathrm{I}(e_i \cdot \theta_i + e_j \cdot \theta_j - e_i \cdot \theta_i \cdot e_j \cdot \theta_j > 0.5)] \cdot f(\theta_j) \mathrm{d}\theta_j - c(e_i) \end{aligned} \tag{4.11}$$

The second term in equation (4.11) represents the points obtained form the fifth test, where no interactions among student was allowed, while the third term represents the bonus obtained in tests 2, 3 and 4.
The assumption that information is shared by the students is crucial and implies that the probability of knowing the answer is given by the common knowledge of the couple. Thus, here the knowledge of the couple is the union of the knowledge of the two members and the optimal effort is given by:

$$c'(e_i^{coop}) = \theta_i - \frac{3}{5} \cdot \bar{v} \cdot \theta_i \int_0^1 \theta_j \cdot e_j \cdot f(\theta_j) \mathrm{d}\theta_j \tag{4.12}$$

The second term in the right-hand side of equation (4.12) is always non-positive, and its absolute value increases with θ_i. This shows that, since information is shared, each team member has an incentive to exploit the effort of the other lowering his own contribution. As a consequence, under the cooperative treatment, team members have an incentive to shrink their effort, and this detrimental effect of cooperation on effort is stronger for students with higher ability (θ_i).

Testable predictions To sum up, our theoretical model predicts that, given the ability θ_i, the effort exerted by student i in the three treatments is such that:

$$e_i^{coop} \le e_i^{BL} \le e_i^{comp}$$

i.e., we expect that on average students randomized into the COOPERATIVE treatment exert lower or equal effort than students randomized into the CONTROL treatment whereas students randomized into the COMPETITIVE treatment should exert more effort.[13] Conversely, at time 1, all students have the same individual incentives to increase effort and optimal effort depends only on their ability level, i.e. $e_{i,0}^{coop} = e_{i,0}^{BL} = e_{i,0}^{comp} = e_{i,0}$. Moreover, the model predicts that the detrimental effect of the cooperative scheme is stronger for high ability individuals while the same type of individuals should exert more effort with respect to the less able individuals in the baseline treatment. Note that our main testable predictions involve the differential changes in effort across treatments and ability levels. Our design allows to measure these changes, as discussed in more detail in section 4.5.1.

We also expect that students assigned to the cooperative treatment will use the chat more frequently and will use it to exchange information. Conversely, students assigned to the competitive treatment should use the chat less frequently and could potentially use it for acts of sabotage, i.e. to suggest the wrong answers. We collected data to check these aspects. Results of our inquiry are discussed in section 4.5.3.

4.5 Results

In this section we first discuss our choice of outcome measure, then present the data and discuss the results on the effect of the incentives on information sharing and on effort.

4.5.1 Measuring Effort

Our theoretical model predicts that for a given level of ability, there is a weak ordering in the effort exerted by each student i, namely $e_i^{coop} \leq e_i^{BL} \leq e_i^{comp}$. We thus expect that on average students randomized into the COOPERATIVE treatment exert lower or equal effort than students randomized into the CONTROL treatment whereas students randomized into the COMPETITIVE treatment should exert more effort.[14]

Equation (4.3) in our simple model describes the relationship between expected student performance at each test and effort, namely $s_i = \theta_i e_i$, where s_i is the net score of individual i, θ_i is a measure of individual ability and e_i is the effort exerted.

Taking logs and allowing for noise in the way in which effort generates performance, we get

$$y_i = \zeta_i + \epsilon_i \tag{4.13}$$

[13]The ordering holds if the distribution of abilities is the same in the three treatments.

[14]The ordering holds if the distribution of abilities is the same in the three treatments and this is guaranteed by randomization.

where $y_i \equiv log(s_i)$ is the log of the net score of individual i, $\zeta_i \equiv log(e_i)$ is the log of the effort exerted, while $\epsilon_i = log(\theta_i) + \varepsilon_i$ and $E[\epsilon_i] = log(\theta_i)$, i.e. we assume that only the idiosyncratic component ε averages to 0 for any i, while the error ϵ_i has a possibly non-zero mean equal to an individual specific constant.

Our experimental design provides an interesting way to measure effort under weak assumptions. Recall that we observe students' performance in similar tests both before the assignment to the treatments (test 1) and after the exposure to the treatments (test 5). Both these tests are taken individually under all treatments and cover similar topics[15]. However, by construction, the performance in the first test and the effort exerted to pass it cannot be affected by the treatments since both performance and effort are pre-determined with respect to the assignment to the different incentive schemes. Conversely, the performance in the last test should reflect changes in effort induced by the treatment. Indeed, moving from equation (4.13) and contrasting the performance in test 5 and 1, we have $y_i - y_0 = \zeta_i - \zeta_{i0} + \varepsilon_i - \varepsilon_{i0}$. It follows that $E[y_i - y_{i0}] = E[\zeta_i - \zeta_{i0}]$, i.e. by looking at the change in the logarithm of performance between the first and last test, we measure the change of the logarithm of effort net of the direct effect of any fixed individual specific factor.

Recall that all our treatment conditions have a **common** individual incentive to increase effort but differ in the incentives to compete or cooperate and only in the baseline students can neither compete, nor cooperate. Following the theoretical predictions of our simple model, we expect an increase in effort in all treatments with respect to a set up where no individual incentives are granted. Our experiment is not designed to estimate this common effect -none of our groups has no individual incentives- but to capture the differential changes induced by the different treatments. The testable prediction of our model involves the differential increase in effort under the cooperative and competitive scheme with respect to the baseline. This weak ordering holds also if we consider $log(e)$, since the logarithm is a monotonic transformation.

To test the theoretical predictions, we first contrast the distribution of effort under the three schemes and check for heterogeneity in the treatment effect over the effort distribution. We then assess the effect on the average change in $log(e)$ and run the following regression

$$E[\zeta_i - \zeta_{i0}] = \beta_0 + \beta_1 Coop + \beta_2 Comp \tag{4.14}$$

where β_0 represents the average change in $log(e)$ under the baseline, β_1 is the average differential change in $log(e)$ under the cooperative scheme with respect to the baseline, and β_2 is the average differential change in $log(e)$ under the competitive scheme with respect

[15]The last test covers a larger set of arguments which includes also those covered by the first and is more closely spaced over time with respect to the other tests.

to the baseline. The theory predicts $\beta_1 \leq 0$ and $\beta_2 \geq 0$. There is an additional prediction that $\beta_0 = 0$, i.e. no change in performance under the baseline. However, our model does not allow for learning which may occur in practice. Namely, after the first test the performance of the students in the baseline improves because they are becoming more familiar with the types of tests and the way the tests are performed in the laboratory. Allowing for learning will not affect our theoretical predictions provided that learning is constant across treatments. If learning occurs in practice, $\beta_0 > 0$.

4.5.2 Data and Descriptive Statistics

Among the 145 students attending the course, 131 applied for participation into the experiment. Our elaborations are based only on the records of the *stayers*, i.e. 114 students who participated to all 5 tests.

We exclude from the elaborations the records of 17 students who missed at least one test: 10 students assigned to the control treatment (BASELINE in what follows), 2 students assigned to the COOPERATIVE treatment and 5 students assigned to the COMPETITIVE treatment (see table 4.9 in Appendix). We shall highlight that 6 of these students were late at the 3rd test and were thus excluded from that test. The experimental program is run in z-Tree Fischbacher (2007): when the test (the experimental session) starts, additional subjects can participate only shutting down and restarting the entire session. Students were informed that not being on time for the test would result in being excluded from the test session. Out of these 17 students, 8 dropped out after the first test: all these students were assigned to the baseline treatment after test 1. When we compare stayers and dropouts in the full sample, we cannot reject the null that drop-outs had a worse performance in the first test.[16]

Once we limit the analysis to the students who participated at all tests, the samples are relatively balanced across treatments with respect to observed and pre-determined characteristics: we do not detect differences in the distribution of the score the first test (score 1) and the average score at previous exams (GPA) between any two treatments (BASELINE, COOPERATIVE, COMPETITIVE) at any conventional level of confidence (see Table 4.2). Figures 4.4 and 4.5 in Appendix report the empirical probability distribution of the pre-treatment variables (the score in the first test, and the average mark in previous exams).

[16]There are no significant differences between the subpopulation of excluded students and the stayers in observable and pre-determined characteristics among the students who where assigned to the COOPERATIVE treatment. We do not reject the null of equal means at 1% level -but we reject at 5%- in the subpopulations for the other treatments: students who participated to all tests in the BASELINE and in the COMPETITIVE treatment tend to be those who achieved a higher score in the first test (0.7 points higher than the one for those who dropped out in the BASELINE group and 0.85 points higher than the one for those who dropped out in the COMPETITIVE treatment).

Table 4.2 also reports the mean value of several other individual characteristics, obtained from subjects' answers to the questionnaire and p-values of tests aimed at detecting differences in these characteristics across treatments.[17] In general, the overall sample is well balanced across treatments. There are some exceptions: the frequency of use of e-mail is significantly higher in the BASELINE treatment than in the COMPETITIVE and in the COOPERATIVE treatments. Significant differences emerge also in terms of the education level achieved by the students' fathers (but not mothers).

To detect the role of interactions effect between the treatments and the students' ability, we consider several different proxies for student's ability and include interaction terms in a simple regression. Our favorite proxy to control for student's ability is the average mark at previous exams: students who participated in the experiment are third year students taking exams is in the last quarter of the third year; therefore, their academic history can be a reliable proxy of their academic skills. In line with the most recent empirical evidence from Italy (AlmaLaurea, 2009), also in our sample females tend to perform significantly better than males in terms of GPA (Females $= 25.2$, Males $= 24.3$, Rank-Sum Test $= P$ 0.028). We say an individual is a high ability individual if his/her score on the classification variable is above the median for that variable in the sample.

4.5.3 Communication and treatments.

Students under both treatments' schemes had two ways to communicate: they could send text messages or hints[18]. Messages and hints were limited in two ways. On the one hand students could not send any information useful to identify themselves (under the threat of exclusion from the test); on the other hand, for each of the 5 questions asked in a test, a student can send and receive only one message of both types.
Table 4.3 together with Table 4.11 in the Appendix report descriptive statistics on the use of chat by subjects. The figures suggest that almost everybody under the COOPERATIVE treatment accepted it[19], and that the average number of exchanged messages is six times higher than in the COMPETITIVE treatment.
The chat tended to be used more frequently than the hint under both schemes.
The content of conversations suggests the chat has been actually used to exchange information. Conversely, the chat was not actively used by students under the COMPETITIVE

[17] We contrasted averages across treatments by means of linear and non linear regressions.

[18] The hint consisted in a simple message informing the receiver that the sender believes a certain answer to be the right one. The sender can suggest a different answer with respect to the one actually selected in the test.

[19] At the beginning of the exam the student must input the registration number and then choose if she wants to use the chat or not.

	mean				p-values		
characteristic	**pooled**	**baseline (BL)**	**cooperative (COOP)**	**competitive (COMP)**	**BL vs. COOP**	**BL vs. COMP**	**COOP vs. COMP**
GPA	24.8	24.8	24.9	24.8	0.808	0.883	0.928
score 1	1.8	1.8	1.9	1.7	0.520	0.564	0.220
age	21.7	21.6	21.9	21.5	0.280	0.736	0.161
gender (Male)	47.4%	40.5%	51.2%	50.0%	0.346	0.418	0.915
freq. mail^{+}	46.7%	62.9%	43.2%	33.3%	0.098	0.017	0.396
freq. chat^{+}	54.3%	54.3%	51.4%	57.8%	0.803	0.785	0.602
freq. pc^{+}	43.8%	45.7%	40.5%	45.5%	0.658	0.983	0.678
father edu.$^{++}$	31.4%	42.9%	13.5%	39.4%	0.008	0.772	0.017
mother edu.$^{+++}$	29.5%	37.1%	24.3%	27.3%	0.241	0.386	0.778
risk aversion	6.0	6.0	6.1	5.8	0.964	0.497	0.515
risk averse	50.9%	48.6%	53.7%	50%	0.659	0.908	0.749
trust 1	4.9	4.7	4.9	5.0	0.567	0.513	0.744
truster (1)	39.5%	27.0%	43.9%	47.2%	0.124	0.077	0.770
trust 2	3.8	3.7	3.8	4.0	0.863	0.533	0.664
truster (2)	27.2%	21.6%	29.3%	30.6%	0.441	0.386	0.902

Table 4.2: Mean value of individual characteristics, by treatment. An individual is risk averse or truster if his answer on the scale is higher or equal to 6. $^{+}$ Binary indicator for whether chat, pc or e-mail are used more than once a day. $^{++}$ Binary indicator for whether the father as high education (equal to college or higher). $^{+++}$ Binary indicator for whether the mother as high education (college or higher). The significance of differences across treatments is estimated by means of simple linear and non linear regressions (logit) for binary indicators. P-values are reported.

Table 4.3: Use of the chat

Treatment	Acceptance of the chat	Av. num. of messages	Av. message length
Cooperative	98% of subjects	3 (out of 5)	28 words
Competitive	70% of subjects	0.5 (out of 5)	11 words

scheme: they declared to be willing to use the chat but only 0.5 messages were exchanged on average. More importantly, students did not believe in the messages of the partner[20]. Indeed, in some cases the chat has been used to deceive the partner (see Table 4.5, and Figure 4.8 in Appendix for an illustrative example).

Table 4.4 reports descriptive statistics on the number of actions taken by students under

Table 4.4: Number of actions (i.e. use of chat and use of hints) by round and treatment.

	mean	sd	median	min	max
Cooperative					
Test 2	5.12	3.36	6	0	10
Test 3	5.80	2.92	7	0	10
Test 4	6.37	2.91	6	0	10
Competitive					
Test 2	1.47	2.48	0	0	8
Test 3	1	2.51	0	0	10
Test 4	1.67	2.24	0	0	8

each treatment. Sending a text message or giving a hint are actions. Under the COOPERATIVE scheme the average number of actions tend to increase from the first test in couples (test 2) to the last (test 4), changing from nearly 5 to above 6, and the correlation between the number of actions taken in different tests is positive, between 0.34 and 0.53, and decreasing with the lag between tests. Some students under the COOPERATIVE scheme used all the available actions (5 text messages and 5 hints) and the median number of action is between 6/7: students tended to use at least one of the two available actions in each question

[20]We do not provide descriptive statistics on the extent of sabotage because these statistics would not be comparable across treatments. Indeed, given the low number of individuals that used the chat under the competitive treatment, we will not get reliable statistics for that group.

of each test and they often used both. Generally, the text message was sent before the hint, and the time lag between the text message and the hint ranges between 1 and 5 minutes in most questions and tests (see Table 4.11 in the Appendix). Conversely, under the COMPETITIVE scheme the median number of actions taken is always 0 and the average number of actions remains relatively stable slightly above 1: students tend to use both the chat and the hint for the same question and only once per test. They also tend to send the text message and the hint almost simultaneously or to send the hint before the text message (see Table 4.11 in the Appendix). The correlation between the number of actions taken in subsequent tests is weaker (between 0.17 and 0.36) and tends to increase with the lag between tests. The correlation between the exerted effort and the number of actions is negligible under both schemes.

We consider data on the couples in each test and contrast answers of the members: Table 4.5 shows that members of the couples under the COOPERATIVE scheme tend to give the same answer much more frequently than their class mates under the COMPETITIVE scheme. The difference is stable across tests and slightly higher than 25%.

Table 4.5: Proportion of cases in which the members of the couple give the same answer.

	Test 2	Test 3	Test 4
Cooperative	56.38%	77.26%	84.78%
Competitive	30.5%	52%	56.84%
Difference	25.88	25.26	27.94

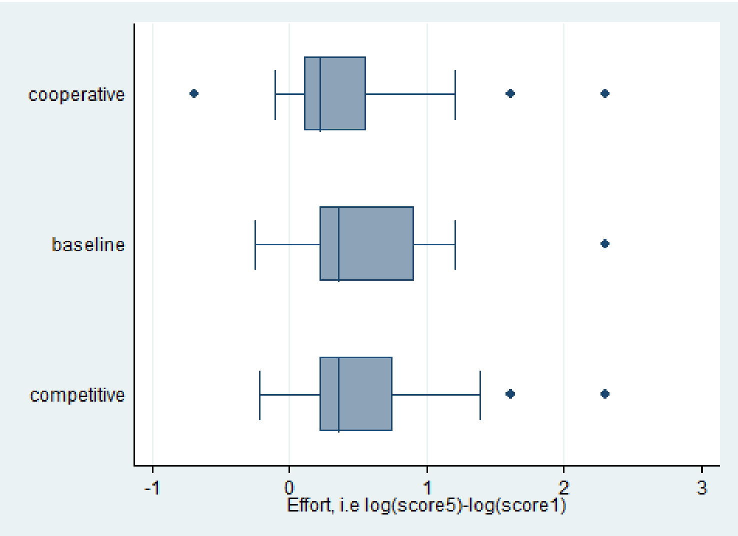
We interpret the observed pattern of information exchange across treatments as a positive response to the incentives: students understood the different mechanisms underlying the two different schemes and behaved accordingly as far as exchange of information is concerned.

4.5.4 Treatment effects

Figure 4.1 depicts the empirical distribution of effort (i.e. log(net score 5) - log(net score 1)) across treatments. The vertical blue line represents the median of the distribution, the left hinge of the box indicates the 25th percentile, and the right hinge of the box indicates the 75th percentile. Visual inspection suggests that under the COOPERATIVE treatment, subjects perform more poorly respect to the BASELINE treatment, while no sizable differences emerge between the COMPETITIVE and the BASELINE treatments.

Wilcoxon tests do not reject the hypothesis that the distribution of effort is the same

Figure 4.1: Box-plot showing the distribution of effort across treatments.



across treatments. These tests are not appropriate if we want to establish an ordering across all three treatments. Thus, we also perform a Jonckheere-Terpstra test, a non-parametric test designed to detect alternatives of ordered class differences. This test does reject the hypothesis that effort is constant across treatments versus the alternative hypothesis that effort is ordered across treatments according to our main theoretical prediction ($e_i^{coop} \leq e_i^{BL} \leq e_i^{comp}$) at 10%.
P-values of these tests are reported in Table 4.6, together with the mean level of effort in each treatment condition.

It has been pointed out in Section 4.2 that according to the experimental literature, a competitive environment may induce different effects on effort for females and for males. Consistently with these works, we find that the picture indeed changes when we split the sample by gender. Figure 4.2 reveals that the treatment effect is substantially different for male and female subjects. The detrimental effect of the COOPERATIVE treatment on effort with respect to the COMPETITIVE treatment only emerges for males, whereas for females no clear treatment effect arises.

One-sided Wilcoxon tests confirms that males' level of effort is significantly lower in the COOPERATIVE treatment than in the COMPETITIVE treatment at 10% level but no significant difference emerges with respect to the baseline. In contrast, the same test does not reject the hypothesis of equal distribution of effort between any two treatments for the female sample. These tests are not appropriate if we want to establish an ordering across all three treatments. Thus we run the the Jonckheere-Terpstra test for the subsamples of males and females: for the male sample, the test rejects at 5% the null hypothesis that effort is not ordered across treatments against the alternative hypothesis that effort is ordered according

Table 4.6: Mean level of effort, by gender and treatment.

	pooled	males	females
Mean effort			
cooperative	0.500	0.377	0.628
baseline	0.583	0.452	0.677
competitive	0.570	0.680	0.459
Wilcoxon tests (p-values)			
base. vs. coop.	0.313	0.135	0.948
base. vs. comp.	0.745	0.442	0.721
coop. vs. comp.	0.190	0.059*	0.713
Jonckheere-Terpstra tests (p-values)			
	0.088*	0.016**	0.624

Legend: One star, two stars, three stars for significant differences at 10%, 5% and 1% level respectively.

to what predicted by the theory; no effect is detected for females. P-values of these tests are reported in Table 4.6.

Our theoretical model predicts heterogeneity in the effect of the incentives' schemes on effort with respect to students' ability, at least for the competitive treatment. To control in a parsimonious way for individual ability, and for other individual characteristics, while assessing the effects of the treatments' scheme on average effort, we use linear regression models.

We run the analysis separately for males and females as previous results suggest that they react differently to incentives.

Table 4.8 presents the benchmark results of two baseline specifications for males and females: column (1) and (2) do not allow for heterogeneity in the treatment effects with respect to students' ability while in column (3) and (4) we include interactions between treatments and the ability indicator based on the average mark at previous exams. All regressions include controls for father education, risk aversion and trust. The top panel of Table 4.8 reports coefficients estimates while the bottom report p-values of both bilateral and unilateral tests: by specifying the direction in which the null hypothesis of no effect is violated (as predicted by theory), we increase the power of the t-test to detect significant deviations. As reference, we computed the power of a test to detect differences between any two treatments for males and females separately using the descriptive statistics (mean,

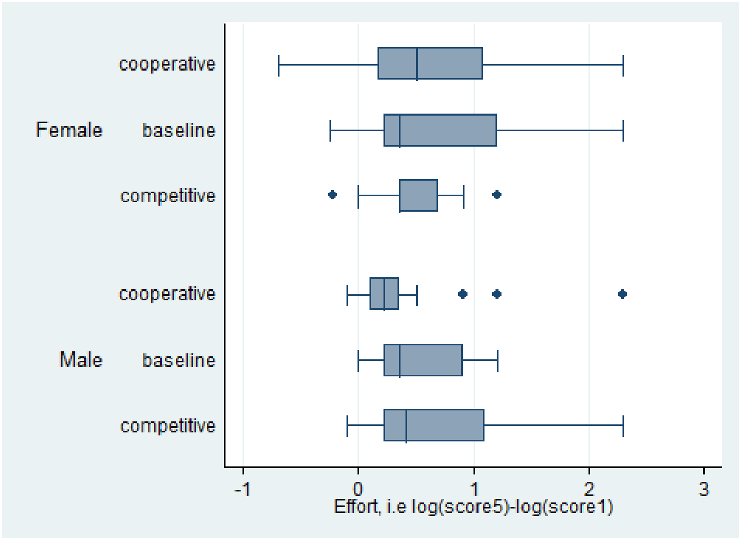


Figure 4.2: Box-plot showing the distribution of effort across treatments, by gender.

standard deviations and sample size) of our sample and significance level $\alpha = 0.5$.[21] Table 4.7 shows that we have little power to detect differences between the baseline and the cooperative treatment, while we have more power to detect differences between the baseline and the competitive treatment for both genders, even all values are quite low. In addition, the table reports the sample size required for a test to detect difference of the size we observe with power 0.8 (assuming constant sample sizes across groups): most of these sizes can be hardly met within a design structured as ours.

Table 4.7: Power of the two-sided and one-sided test of mean comparison across treatments and optimal size n^* for two sided tests with equal group sizes and power 0.8 . Males and Females.

Null Hypothesis	Males			Females		
	Power		n^*	Power		n^*
	2 sided	1 sided		2 sided	1 sided	
Baseline vs Cooperative	0.08	0.12	658	0.06	0.08	3078
Baseline vs Competitive	0.25	0.35	84	0.25	0.36	101
Cooperative vs Competitive	0.07	0.11	879	0.16	0.25	161

Results in Table 4.8[22] confirm previous results on the differential effects across treat-

[21]The power of similar tests for the pooled sample is lower: the gender heterogeneity makes point estimates of the average $log(e)$ less precise.

[22]With respect to control variables, **High parental education** is a binary indicator that takes the value 1 if

Table 4.8: Ordinary Least Squares Estimates of the Treatment Effects

	(1)	(2)	(3)	(4)
	No heterogeneity with ability		Heterogeneity with ability	
Variables	Males	Females	Males	Females
Constant	0.386**	0.652***	0.239	0.740**
	[0.174]	[0.178]	[0.262]	[0.294]
Cooperative	-0.125	-0.156	0.100	-0.135
	[0.178]	[0.219]	[0.263]	[0.382]
Competitive	0.331*	-0.235	0.492*	-0.363
	[0.189]	[0.211]	[0.272]	[0.359]
Coop · High Ability			-0.486	-0.117
			[0.401]	[0.475]
Comp · High ability			-0.266	0.080
			[0.398]	[0.449]
High ability			0.127	0.073
			[0.280]	[0.319]
High parental education	-0.249	-0.155	-0.317*	-0.219
	[0.151]	[0.185]	[0.173]	[0.210]
Frequent use of e-mail			0.139	-0.224
			[0.161]	[0.206]
Risk averse	0.290*	0.221	0.290*	0.280
	[0.147]	[0.173]	[0.150]	[0.198]
Truster	0.105	0.084	0.136	0.114
	[0.155]	[0.194]	[0.164]	[0.213]
Observations	50	55	50	55
R^2	0.237	0.066	0.311	0.157
P-values for the null of no effect against bilateral or unilateral H_1				
(R)$\equiv H_1$: $\beta > 0$; (L) $\equiv H_1$: $\beta < 0$				
Competitive				
1 sided (R)	0.039**	0.867	0.035**	0.844
2 sided	0.089*	0.266	0.071*	0.312
Cooperative				
1 sided (L)	0.240	0.238	0.649	0.362
2 sided	0.480	0.476	0.703	0.724
Competitive for high ability				
1 sided (R)			0.221	0.837
2 sided			0.441	0.326
Cooperative for high ability				
1 sided (L)			0.092*	0.186
2 sided			0.184	0.373

the highest qualification of at least one of the parents of the individual is above high school and 0 otherwise. **Risk averse** is a binary indicator that takes the value 1 if the answer of the individual on the risk aversion scale ments: there is evidence of a significant increase in effort under the competitive treatment with respect to the baseline for males but not for females.[23] The effect is statistically distinct from zero at 10% and not-negative at 5%. When we control for ability, we find that : (i) the positive incentive for males is higher for the low ability individuals (still significantly non-negative at 10%) and decreases substantially for high ability individuals; (ii) there is a negative and statistically significant (at 10%) detrimental effect of the cooperative treatment for high ability individuals only. However, the difference in effects of incentives between ability groups is not significant in our sample for the competitive case nor for the cooperative case. The magnitude of the effect ranges from 33% to 49% which is a strong increment of the exerted effort. Notice that this is in line with the findings of Angrist and Lavy (2009) who use monetary incentives based on the achievement of a specified score target.

For females, no statistically significant effect can be detected. The pattern of the effect of competitive incentives on effort for females is similar to the one detected for males but in the opposite direction: the point estimate of the effect is negative and, when we control for ability, point estimates of the effect of competitive incentives for females are negative for both low and high ability individuals but less so for high ability individuals.

We detect a significant increase in effort also in the baseline: we attribute this to the fact that students become more familiar with the instruments used for the test (*learning*). Students' ability does not play any role in determining the increase in effort in the baseline. Few regressors are relevant in determining changes in students effort: risk aversion and parental background attract significant coefficients in some specifications, suggesting that individuals who are risk averse tend on average to increase effort, while males with higher socio-economic background (here proxied by highly educated parents) tend to decrease effort, other things equal.

4.6 Conclusions

Our study investigates how two alternative incentive schemes affect students' effort, both from a theoretical and from an empirical point of view. To test the theoretical predictions, we run a field experiment in an undergraduate course at the University of Bologna (Italy). We randomly assign students to either a tournament, where coupled students compete to get the reward, a cooperative scheme where information sharing is allowed, or a control

---

is above 6 and 0 otherwise. **Truster (1)** is a binary indicator that takes the value 1 if the answer of the individual on the trust 1 scale is above 6 and 0 otherwise.

[23] Since we include control variates and 9 students do not answer the questionnaire, the sample size relevant for the regressions is 105 instead of 114.

treatment in which students can neither compete, nor cooperate. Differently from previous studies, none of our treatments involves pecuniary incentives but consists in extra points for their final grade.

The field-experiment data we collected confirm the theoretical predictions: we observe a weak ordering between the effort exerted by students under the different treatments with students in the competitive treatment exerting on average more effort with respect to students in the baseline and in the cooperative treatment.

We break down our results by gender and show that a significant difference emerges: only males react to incentives to compete while we cannot detect significant effect for females. Cooperation seems not to foster effort exertion and no gender effect emerges. In contrast with theoretical predictions we find that students' ability plays little role in determining the effectiveness of the incentives.

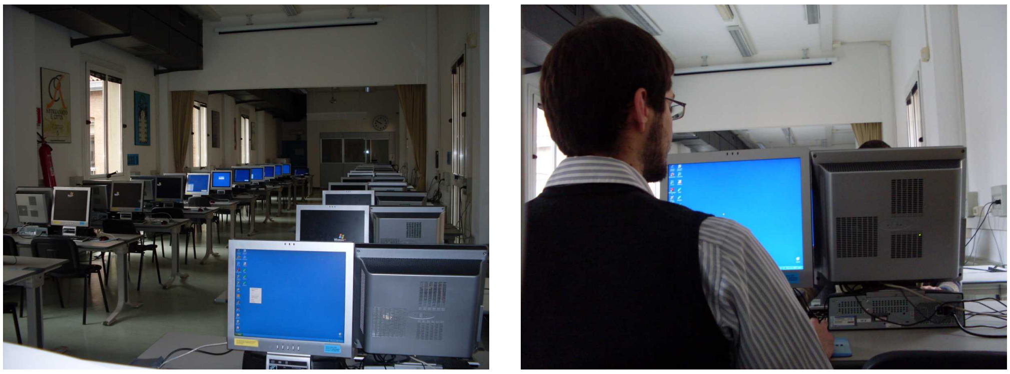
Our experimental results suggest that non-pecuniary incentives based on competition have the potential to increase students' effort as pecuniary incentives do (see Blimpo 2010) but at a much lower financial cost. In our case competition proves to work on males which is line with findings in several other contexts (see for example Gneezy and Rustichini 2004 and Niederle and Vesterlund 2010) where it has been shown that males are more prone to compete with respect to females. The estimated increase in effort induced for males in the competitive treatment ranges from 33% to 49%, meaning that, for example, if a student in the baseline spends 3 afternoons in preparing the test (roughly 10 hours), a student under the competitive scheme will spend one more afternoon. Moreover, highlighting the different effect of incentives to compete depending on gender, we complement the results in Angrist and Lavy (2009) who show that monetary incentives based on absolute performance are more effective for females.

Our study represents a first exploration of the effects of non-monetary incentives on students' performance and effort. It would be interesting to extend the inquiry to different samples, to verify whether our result holds for students with different majors (such as literature of philosophy), who are probably less trained to optimization, and for younger students at high school and middle-high school.

4.7 Appendix

4.7.1 Laboratory

Figure 4.3: The laboratory arrangement



4.7.2 Additional tables

Table 4.9: Descriptive statistics

			Descriptive Statistics- Stayers		
			Predetermined controls		
	assigned	stayers	score 1	exams' avg	score 5
Baseline (control)	47	37	1.80 *(0.81)*	24.76 *(1.8)*	2.91 *(0.24)*
Cooperative	42	41	1.92 *(0.84)*	24.88 *(2.3)*	2.80 *(0.50)*
Competitive	41	36	1.69 *(0.74)*	24.83 *(1.6)*	2.69 *(0.53)*
Full sample	130	114	1.81 *(0.80)*	24.83 *(1.9)*	2.80 *(0.45)*

Score 1: score at the first mock exam. Score 5: score at the last mock exam. Exams' avg: average score at previous exams. Stayers: students who participated to 5 experimental sessions.

In Table 4.10, we report the precise definition of questionnaire data used in the analysis.

Table 4.10: Description of questionnaire data.

Variable	**Corresponding question**	**Range**	**Coding**
gender	*gender*	0, 1	1 = male
age	*age*	0-100	age in years
freq. mail	*how frequently do you check your e-mail?*	1-5	1="more than once per day" 2="at least once per day"
freq. pc	*how frequently do you use the pc to study/work?*	1-5	3= "at least once per week" 4="less than once per week"
freq. chat	*how frequently do you exchange text messages via chat (msn, facebook, google talk, skype, etc.)?*	1-5	5="Never"
father edu.	*please, indicate the education level achieved by your father*	1-5	1="junior high school" 2="high school"
mother edu.	*please, indicate the education level achieved by your mother*	1-5	3="bachelor" 4="master" 5="Ph.D."
risk aversion	*I would describe myself as a risk-averse person.*	1-10	1="fully agree" 10="fully disagree"
trust 1	*Do you think that most people try to take advantage of you if they got a chance or would they try to be fair?*	1-10	1="people would try to take advantage" 10="people would try to be fair"
trust 2	*Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?*	1-10	1="you can never be too careful" 10="most people can be trusted"

Table 4.11: Descriptive statistics -mean, [median] and (standard deviation)- on lag between the use of chat and use of hints, by treatment and round. Questions 1-5

Lag & proportion of user of both chat and hint (seconds). Test 2.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	107.6 [21.7]	322.8 [252.2]	58.6 [6.3]	151.0 [114.1]	54.2 [6.4]
	(449.0)	(378.3)	(565.9)	(513.5)	(495.6)
Users (count)	14	12	11	13	16
Users (%)	35.0%	30.0%	27.5%	32.5%	40.0%
Competitive	80.6 [54.1]	-123.3 [-123.3]	130.4 [23.9]	13.3 [13.3]	25.3 [16.9]
	(67.9)	(131.1)	(1113.5)	(n.a.)	(21.3)
Users (count)	3	2	3	1	3
Users (%)	8.8%	5.9%	8.8%	2.3%	8.8%

Lag & proportion of user of both chat and hint (seconds). Test 3.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	492.5 [374.8]	496.2 [368]	-56.0 [-3.2]	167.6 [71.7]	76.4 [5.2]
	(693.1)	(577.4)	(423.0)	(427.3)	(303.5)
Users (count)	14	16	14	17	15
Users (%)	35.0%	40.0%	35.0%	42.5%	37.5%
Competitive	73.8 [7.6]	-164.2 [-164]	720.1 [720]	-194.0 [165.4]	97.0 [97.0]
	(131.3)	(951.9)	(223.6)	(380.5)	(123.7)
Users (count)	3	2	2	4	2
Users (%)	8.8%	5.9%	5.9%	11.7%	5.9%

Lag & proportion of user of both chat and hint (seconds). Test 4.

Treatment	Question 1	Question 2	Question 3	Question 4	Question 5
Cooperative	146.9 [55.9]	119.8 [20.8]	-15.8 [-3.3]	169.7 [40.8]	95.0 [4.1]
	(482.1)	(342.8)	(522.4)	(355.0)	(240.2)
Users (count)	14	17	17	17	22
Users (%)	35.0%	42.5%	42.5%	42.5%	55.0%
Competitive	194.3 [12.3]	8.5 [8.5]	180.1 [2.1]	458.8 [72.6]	322.3 [322]
	(365.8)	(n.a.)	(314.1)	(721.4)	(449.4)
Users (count)	4	1	3	3	2
Users (%)	11.7%	2.9%	8.8%	8.8%	5.9%

Table 4.12: Robustness checks, different measures of ability

	Measures of ability:					
	Symbolic digit test		Number of exams		Number of credits	
VARIABLES	Males	Females	Males	Females	Males	Females
Constant	0.346	0.939***	0.317	0.983***	0.369	1.211***
	[0.241]	[0.273]	[0.260]	[0.248]	[0.263]	[0.288]
Cooperative	-0.172	-0.212	-0.135	-0.393	-0.115	-0.616**
	[0.254]	[0.315]	[0.233]	[0.278]	[0.249]	[0.306]
Competitive	0.253	-0.248	0.401	-0.715**	0.248	-0.927**
	[0.272]	[0.327]	[0.279]	[0.336]	[0.310]	[0.359]
Coop · High Ability	0.117	-0.172	0.172	0.451	0.009	0.767*
	[0.420]	[0.440]	[0.414]	[0.461]	[0.380]	[0.444]
Comp · High ability	0.138	-0.198	-0.047	0.670	0.177	0.902**
	[0.409]	[0.453]	[0.471]	[0.431]	[0.435]	[0.437]
High ability	0.078	-0.144	-0.116	-0.328	-0.033	-0.551*
	[0.306]	[0.306]	[0.328]	[0.317]	[0.299]	[0.311]
High parental education	-0.284*	-0.299	-0.214	-0.182	-0.275	-0.188
	[0.164]	[0.209]	[0.171]	[0.214]	[0.171]	[0.196]
Frequent use of e-mail	0.033	-0.277	0.110	-0.303	0.056	-0.368*
	[0.176]	[0.204]	[0.179]	[0.199]	[0.173]	[0.196]
Risk averse	0.305*	0.280	0.286*	0.297	0.291*	0.259
	[0.154]	[0.180]	[0.157]	[0.182]	[0.154]	[0.180]
Truster	0.096	0.135	0.122	0.069	0.105	0.089
	[0.167]	[0.195]	[0.163]	[0.202]	[0.167]	[0.197]
Observations	50	54	50	54	50	54
R-squared	0.264	0.150	0.251	0.152	0.247	0.194
P-values for the null of no effect against bilateral or unilateral H_1						
(R)$\equiv H_1$: $\beta > 0$; (L) $\equiv H_1$: $\beta < 0$						
Competitive						
COMP 1 sided (R)	0.176	0.776	0.0754	0.983	0.212	0.995
COMP 2 sided	0.352	0.447	0.151	0.0333	0.424	0.00970
Cooperative						
COOP 1 sided (L)	0.249	0.250	0.281	0.0791	0.322	0.0219
COOP 2 sided	0.499	0.501	0.562	0.158	0.644	0.0438
Competitive for high ability						
COMP HIGH 1 sided (R)	0.0997	0.925	0.157	0.562	0.0702	0.539
COMP HIGH 2 sided	0.199	0.150	0.314	0.877	0.140	0.922
Cooperative for high ability						
COOP HIGH 1 sided (L)	0.432	0.116	0.543	0.561	0.360	0.676
COOP HIGH 2 sided	0.863	0.232	0.915	0.878	0.720	0.649

4.7.3 Additional figures

Figure 4.4: Empirical probability distribution of score 1 by treatment

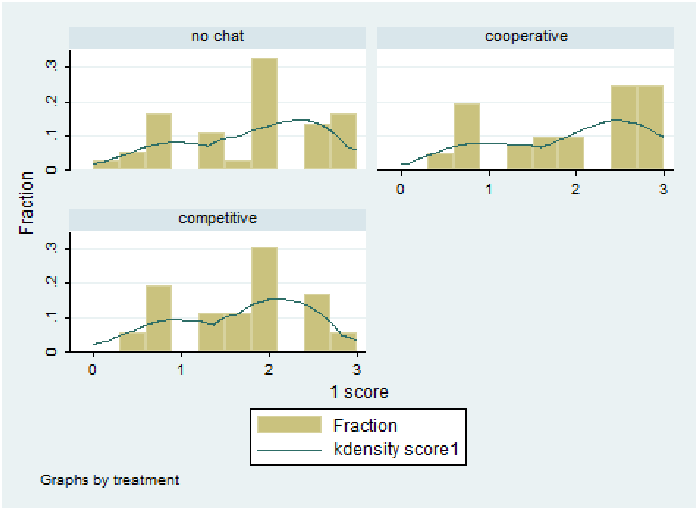


Figure 4.5: Empirical probability distribution of average score at previous exams by treatment

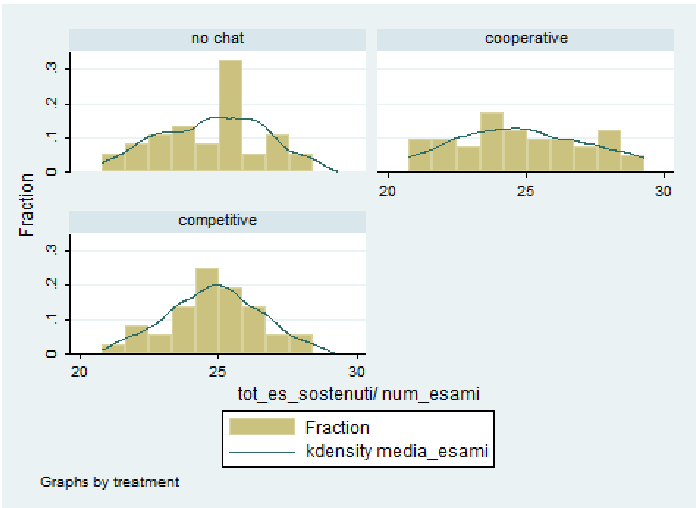
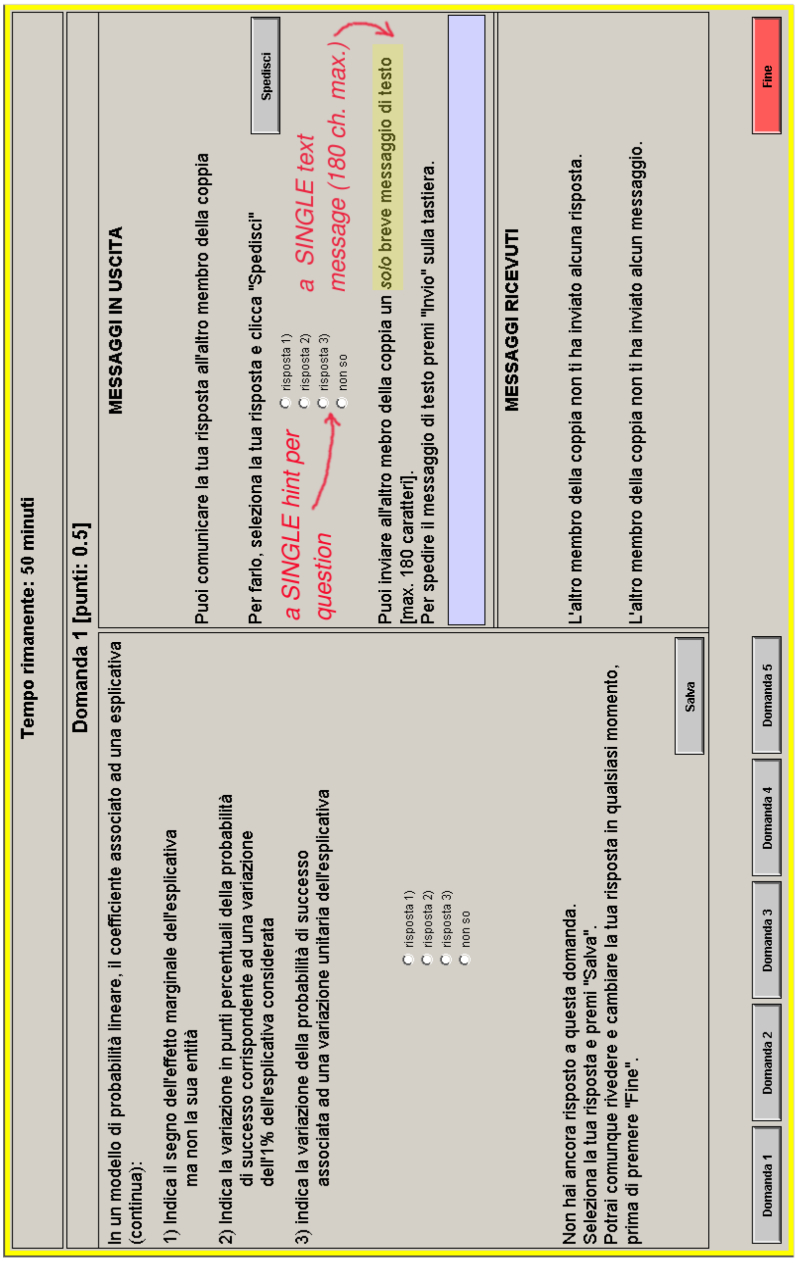


Figure 4.6: Screen-shot of the graphical interface for partial exams.



4.7.4 Examples of chat messages

Figure 4.7: Example of use of the chat under the cooperative scheme

A: Come on! Tell me which answers do you need. If you don't get 1.5 points, we will lose the bonus.
B: In my opinion the right one is the 2^{nd}
A: OK! I trust you

Figure 4.8: Example of use of the chat under the competitive scheme

A: In this case the 4^{th} is the best answer
B (replies): Why do you pass me this solution? Are you trying to screw me?
C: I know that you are going to pass me the wrong answers.
D: I'm not sure...probably the right answer is the 1^{st} *[she choses the 3^{rd}]*

Bibliography

ACLAND, D., AND M. LEVY (2011): "Habit Formation and Naivete in Gym Attendance: Evidence from a Field Experiment," *working paper Harvard University*.

AGRAVWAL, S., J. C. DRISCOLL, X. GABAIX, AND D. LAISBON (2008): "Learning in the credit market," *NBER WORKING PAPER SERIES - WP 13822*.

ALMALAUREA, . (2009): "XI Rapporto Almalaurea 2009," Report, Consorzio Interuniversitario AlmaLaurea - MIUR.

ANDERHUB, V., R. MÜLLER, AND C. SCHMIDT (2001): "Design and evaluation of an economic experiment via the Internet," *Journal of Economic Behavior & Organization*, 46(2), 227 – 247.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76(3), 583–618.

ANDERSON, J., S. BURKS, C. DEYOUNG, AND A. RUSTICHINI (2011): "Toward the Integration of Personality Theory and Decision Theory in the Explanation of Economic Behavior," *Working paper*.

ANGRIST, J., D. LANG, AND P. OREOPOULOS (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 1(1), 136–63.

ANGRIST, J., AND V. LAVY (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384–1414.

ARIELY, D., AND K. WERTENBROCH (2002): "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 13(3), pp. 219–224.

ARKES, H., R. (1996): "The psychology of waste," *Journal of Behaviour and Decision Making*, 9, 213–224.

AZMAT, G., AND N. IRIBERRI (2010): "The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness," Working Papers (Universitat Pompeu Fabra. Departamento de Economía y Empresa), No. 1216.

BALIGA, S., AND J. ELY (2011): "Mnemonomics: The Sunk Cost Fallacy as a Memory Kludge," *working paper Northwestern University*.

BENABOU, R., AND J. TIROLE (2004): "Willpower and Personal Rules," *The Journal of Political Economy*, 112(4), 848–886.

BENJAMIN, D., S. BROWN, AND J. SHAPIRO (2005): "Who Is Behavioral?," *Unpublished*.

BLIMPO, M. (2010): "Team Incentives for Education in Developing Countries. A Randomized Field Experiment in Benin," mineo, March 2010 - Stanford University.

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND T. W. BAS (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4), 972–1059.

BRATTI, M., D. CHECCHI, AND A. FILIPPIN (2008): "Should You Compete or Cooperate with Your Schoolmates?," IZA Discussion Papers 3599, Institute for the Study of Labor (IZA).

BURKS, S. V., J. P. CARPENTER, L. GOETTE, AND A. RUSTICHINI (2010): "Overconfidence is a Social Signaling Bias," *IZA Working Paper*, (4840).

CALZOLARI, G., AND M. NARDOTTO (2011): "Nudging consistent plans with information: a randomized field experiment on feedbacks and reminders," *Mimeo*.

CAMERER, C., AND D. LOVALLO (1999): "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89(1), 306–318.

CHARNESS, G., AND U. GNEEZY (2009): "Incentives to Exercise," *Econometrica*, 77(3), pp. 909–931.

COOPER, D., AND J. KAGEL (2005): "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games," *American Economic Review*, 95(3), 477–509.

DE PAOLA, M., AND V. SCOPPA (2008): "Peer Group Effects in the Academic Performance of Italian Students," *Applied Economics*, pp. 1466–4283.

——— (2010): "Monetary Incentives and Student Achievement in a Depressed Labour Market: Results from a Randomized Experiment," Working Paper n. 06 - 2010, UNICAL, Dipartimento di Economia e Statistica.

DELLA VIGNA, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2), 315–72.

DELLA VIGNA, S., AND U. MALMENDIER (2004): "Contract design and self-control," *Quarterly Journal of Economics*, CXIX, 353–402.

——— (2006): "Paying not to go to the gym," *American Economic Review*, 96, 694–719.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2008): "Representative Trusr and Reciprocity: Prevalence and Determinants," *Economic Inquiry*, 46(1), 84–90.

——— (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 100(3), 1238–60.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2005): "Individual Risk Attitudes: New Evidence From a Large, Representative, Experimentally-Validated Survey," Discussion Paper DP 1730, Institute for the Study of Labor, IZA.

DOMINITZ, J., AND C. F. MANSKI (2004): "How Should We Measure Consumer Confidence?," *Journal of Economic Perspectives*, 18(2), 51–66.

ELIAZ, K., AND R. SPIEGLER (2006): "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, 3, 689–714.

FALK, A., AND A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–58.

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

FREDERICK, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19(4), 25–42.

FRYER, R. J. (2010): "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *working paper Harvard University and NBER*.

FUDEMBERG, AND LEVINE (2011): "A Dual-Self Model of Impulse Control," *Working paper*.

GABAIX, X., AND D. LAISBON (2006): "Shrouded attributes, consumer myopia, and informative suppression in competitive markets," *Quarterly Journal of Economics*, CXXIV, 505–540.

GILBOA, I., AND D. SCHMEIDLER (2001): *A Theory of Case-Based Decisions*. Cambridge University Press.

GNEEZY, U., K. L. LEONARD, AND J. LIST (2009): "Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society," *Econometrica*, 77(5), 1637–1664.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance In Competitive Environments: Gender Differences," *The Quarterly Journal of Economics*, 118(3), 1049–1074.

GNEEZY, U., AND A. RUSTICHINI (2004): "Gender and Competition at a Young Age," *American Economic Review*, 94(2), 377–381.

GOURVILLE, J. T., AND D. SOMAN (1998): "Payment Depreciation: The effects of Temporally Separating Payments from Consumption," *Journal of Consumer Research*, 25, 160 – 174.

GRUBB, M. D. (2009): "Selling to Overconfident Consumers," *American Economic Review*, 99(5), 1770–1807.

GUL, F., AND W. PESENDORFER (2001): "Temptation and Self-Control," *Econometrica*, 69(6), 1403–1435.

HECKMAN, J., J. STIXRUD, AND S. URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

HECKMAN, J. J., AND Y. RUBINSTEIN (2001): "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *The American Economic Review*, 91(2), pp. 145–149.

HEIDHUES, P., AND B. KÖSZEGI (2010): "Exploiting Naïvete about Self-Control in the Credit Market," *American Economic Review*, 100(5), 2279–2303.

HEINECK, G., AND S. ANGER (2008): "The Returns to Cognitive Abilities and Personality Traits in Germany," *SOEP papers*, pp. 1–48.

HOLT, C. A., AND S. K. LAURY (2002): "Risk aversion and incentive effects," *American Economic Review*, 92, 1644–1655.

KARLAN, L., M. MCCONNELL, S. MULLAINATHAN, AND J. ZINMAN (2010): "Getting to the Top of Mind: How Reminders Increase Saving," *NBER working paper*.

KREMER, M., E. MIGUEL, AND R. THORNTON (2009): "Incentives to Learn," *Review of Economics and Statistics*, 91(3), 437–456.

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112(2), pp. 443–477.

LAIBSON, D., A. REPETTO, AND J. TOBACMAN (2007): “Estimating Discount Functions with Consumption Choices over the Lifecycle,” *National Bureau of Economic Research Working Paper 13314*.

LANG, F., D. WEISS, A. STOCKER, AND B. ROSENBLADT (2007): “Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP),” *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 127(1), 183–192.

LAZEAR, E., AND S. ROSEN (1981): “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89(5), 841–64.

LEUVEN, E., H. OOSTERBEEK, AND VAN DER KLAAUW B. (forth.): “The Effect of Financial Rewards on Studentsï¿½ Achievements: Evidence from a Randomized Experiment,” Discussion paper, Journal of the European Economic Association.

LEVITT, S. D., AND J. LIST (2007): “What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?,” *Journal of Economic Perspectives*, 21(2), 153–174.

LOWESTEIN, AND O’DONOGHUE (2005): “Animal spirits,” *Working paper*.

LYLE, D. (2009): “The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point,” *American Economic Journal: Applied Economics*, 1(4), 69–84.

MADRIAN, B. C., AND D. F. SHEA (2001): “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *The Quarterly Journal of Economics*, 116(4), pp. 1149–1187.

MALMENDIER, U., AND G. TATE (2008): “Who makes acquisitions? CEO overconfidence and the market’s reaction,” *Journal of Financial Economics*, 89(1), 20 – 43.

MANN, H., AND D. WHITNEY (1947): “On a test of whether one of 2 random variables is stochastically larger than the other,” *Annals of Mathematical Statistics*, 18, pp. 50–60.

MANSKI, CHARLES, F. (2004): “Measuring expectation,” *Econometrica*, 72, 1329–1376.

MAS, A., AND E. MORETTI (2009): “Peers at Work,” *American Economic Review*, 99(1), 112–145.

MIRAVETE, E. J. (2003): “Choosing the Wrong Calling Plan? Ignorance and Learning,” *The American Economic Review*, 93(1), 297–310.

NAGEEB, ALI, S. (2011): “Learning Self-Control,” *Quarterly Journal of Economics, forthcoming*.

NARDOTTO, M. (2011a): “On the relationship between risk and time preferences with cognitive abilities and personality traits,” *Mimeo*.

——— (2011b): “Stable Naive and Math Sophisticated, a Detailed Field Study on Overconfidence,” *Mimeo*.

NIEDERLE, M., AND L. VESTERLUND (2007): “Do Women Shy Away from Competition? Do Men Compete Too Much?,” *The Quarterly Journal of Economics*, 122(3), 1067–1101.

NIEDERLE, M., AND L. VESTERLUND (2010): “Explaining the Gender Gap in Math Test Scores: The Role of Competition,” *Journal of Economic Perspectives*, 24(2), 129–44.

O’DONOGHUE, T., AND M. RABIN (1999): “Doing It Now or Later,” *The American Economic Review*, 89(1), pp. 103–124.

——— (2001): “Choice and Procrastination,” *Quarterly Journal of Economics*, CXVI, 121–160.

PHELPS, E. S., AND R. A. POLLAK (1968): “On Second-Best National Saving and Game-Equilibrium Growth,” *The Review of Economic Studies*, 35(2), 185–199.

PRENDERGAST, C. (1999): “The Provision of Incentives in Firms,” *Journal of Economic Literature*, 37(1), pp. 7–63.

RABIN, M. (1998): “Psychology and Economics,” *Journal of Economic Literature*, 36(1), pp. 11–46.

RUBIN, D., C., AND E. WENZEL, A. (1996): “One Hundred Years of Forgetting: a Quantitative Description of Retention,” *Psychological Review*, 103, 734–760.

SACERDOTE, B. (2001): “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” *The Quarterly Journal of Economics*, 116(2), 681–704.

SEXTON, R., B. JOHNSON, N., AND A. KONAKAYAMA (1987): “Consumer Response to Continuous-Display Electricity-Use Monitors in a Time- of-Use Pricing Experiment,” *The Journal of Consumer Research*, 14, 55–62.

STANGO, V., AND J. ZINMAN (2010): “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Penalty Fees,” *working paper Graduate School of Management University of California, Davis*.

STINEBRICKNER, R., AND T. STINEBRICKNER (2008): “The Causal Effect of Studying on Academic Performance,” *The B.E. Journal of Economic Analysis & Policy*, 8(1).

THALER, R. H. (1999): “Mental accounting matters,” *Journal of Behavioral Decision Making*, 12(3), 183–206.

WILCOXON, F. (1945): “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, 1(6), pp. 80–83.

ZIMMERMAN, D. (2003): “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment,” *Review of Economics and Statistics*, 85(1), 9–23.