I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12807

Selection into Experiments: Evidence from a Population of Students

Jonathan Schulz
Uwe Sunde
Petra Thiemann
Christian Thöni

NOVEMBER 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12807

Selection into Experiments: Evidence from a Population of Students

Jonathan Schulz
George Mason University and CeDEx

Uwe Sunde
University of Munich, IZA and CEPR

Petra Thiemann
Lund University and IZA

Christian Thöni
University of Lausanne

NOVEMBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Selection into Experiments: Evidence from a Population of Students

This study investigates the selection into lab experiments among university students based on data from two cohorts of a university's first-year students. The analysis combines two experiments: a classroom experiment in which we elicited measures for risk, time, social preferences, confidence, and cognitive skills using standard measures from the experimental literature; and a recruitment experiment that varied information provided in a typical e-mail recruitment procedure for lab participants. In the recruitment experiment, students were randomly assigned to four conditions that highlighted altruistic motives or financial incentives. We find significant treatment effects: mentioning financial incentives boosts the participation rate in lab experiments by 50 percent. In terms of selection, we find that more selfish individuals and individuals with higher cognitive reflection scores are more likely to participate in experiments, but we find little evidence for selection along risk preferences, time preferences, and overconfidence. Although the recruitment conditions affect participation rates, they do not alter the composition of the participant sample in terms of behavioral measures and cognitive skills.

JEL Classification: C93, D64, H41, L3

Keywords: classroom experiment, selection, recruitment, preferences, cognitive abilities

Corresponding author:
Petra Thiemann
Lund University
Department of Economics
Tycho Brahes väg 1
SE-223 63 Lund
Sweden

E-mail: petra.thiemann@nek.lu.se

1 Introduction

Studying human behavior in laboratory experiments has become one of the predominant empirical methods in modern economics and in the social sciences in general. Laboratory experiments provide useful insights into decision making because they allow researchers to draw causal inferences based on controlled variation—including the possibility to randomly allocate participants to treatments, to control and hold all relevant aspects of the decision environment fixed, and to provide well-defined monetary incentives that are linked to the participants' decisions. Thus, lab experiments entail a degree of internal validity that cannot be obtained as easily or at all with data from other sources. As a consequence, lab experiments are widely used in economics (Falk and Heckman, 2009).

One of the major concerns raised in this context has been that most lab experiments are conducted with selected subject pools since participation is voluntary and subjects deliberately decide to participate, i.e., self-select into participation. In the vast majority of lab experiments in economics, the subject pool consists of undergraduate economics students. Thus, the subjects may not even be representative of the student population. As a consequence, causal inference may not be generalizable, and estimates of parameters of interest, e.g., of preference parameters, may be biased (Andersen et al., 2010, von Gaudecker et al., 2012). This threatens the external validity of experimental findings since the observed behaviors—even very basic ones—might not be representative for the population in question. Whereas a growing body of research compares the behaviors of student subject pools and subject pools from other strata of the population (discussed below), the question of self-selection into experiment participation has received little attention.

This paper presents the results from large controlled experiments that we conducted to investigate selection of student samples into lab experiments. The study provides two major methodological contributions. First, the analysis is based on a large and comprehensive population of two cohorts of first-year students at a Swiss university. During the students' second week at university, we ran a classroom experiment to elicit an extensive list of students' traits including social preferences, risk preferences, time preferences, confidence, and cognitive reflection. This list includes a large portion of the traits that have been studied in the experimental economics literature. For several reasons, voluntary selection may lead to biased measures of these traits. Our paper presents a comprehensive investigation of self-selection into lab experiments along these traits based on a representative student population that comprises different academic disciplines and cohorts.

Second, to better understand the factors that influence selection, we experimentally varied the content of recruitment e-mails with which we invited the students to sign into a database of experimental subjects. We implemented four treatments, which either emphasized the social aspect (supporting research), the financial aspect (earning a substantial amount of money), the risk aspect (earning a certain amount for sure), or the combination of the social and financial aspect.

Our design allows us to investigate several hypotheses and concerns about the selection of subject pools into typical lab experiments in economics, based on the population of freshmen of two subsequent years, before they are exposed to any subject-related education. For instance, it is not entirely clear whether student subjects are particularly self-interested individuals who react to financial incentives for participation, or whether student subjects are rather altruistic individuals who like to engage in the creation of scientific knowledge and to potentially help the researchers at their university. Furthermore, usually uncertain earnings may discourage risk-averse students from participating in experiments, while overconfident students may select into the experiments based on inflated beliefs about their own ability and about potential experimental earnings. Finally, opportunity costs of forgoing leisure or study time may play a role. These opportunity costs may be lower for students with high cognitive ability, compared to students with low cognitive ability, since high-ability students have to exert less effort to understand an experiment; at the same time, they do not have to study as much to pass exams. This may lead to an over-representation of students with higher ability in experimental subject pools.

The results reveal new insights regarding the selection of students into a recruitment database for potential experimental subjects and into actual participation in lab experiments. In our recruiting experiment, we show that emphasizing financial incentives in addition to benefits to society increases both the probability to subscribe to the database and the probability to participate in experiments by about 50 percent. By contrast, mentioning benefits to society in addition to financial incentives neither affects the probability to subscribe to the recruitment database nor the probability to participate in experiments. The same is true for the explicit reference to a show-up fee in addition to financial incentives.

In terms of social preferences, we find that participants of experiments are less altruistic compared to individuals who do not participate in experiments, but we do not find significant differences in levels of trust or reciprocity between participants and non-participants. Moreover, we find no evidence that participants of experiments are more willing to take risk, more patient, or more overconfident than non-participants. We do find that students who participate in experiments display higher cognitive skills, as measured by a cognitive reflection test, than non-participants, however.

Despite the strong effects of recruitment treatments on participation, we do not find that the different recruiting conditions change the composition of the participant sample in terms of behavioral measures and cognitive skills. First, the prospect of a certain show-up fee in addition to financial incentives does not dis-proportionally attract students with low confidence or willingness to take risk into the lab. Second, mentioning financial incentives in addition to benefits to society does not dis-proportionally appeal to selfish students. Likewise, mentioning benefits to society in addition to, or instead of, financial incentives does not dis-proportionally draw altruistic students in. Finally, we also do not find any significant interaction effects between the recruitment treatments and selection along time preferences or cognitive skills. Additional evidence sheds light on the selectivity related to repeat participation in lab experiments.

Contribution to the literature. The results of this paper contribute to the recent debate on the selection of participants into experiments. Broadly speaking, this debate has two distinct dimensions. One dimension relates to the comparability of experimental subject pools with the population at large, while the other dimension is concerned with the determinants of selection into participation in lab experiments, including preferences, traits, and the influence of information provided in the recruitment process. Both strands of the literature address different aspects. On the one hand, studies that solely investigate the role of particular factors for self-selection into experiments—such as the information given to potential participants during the recruitment process—provide valuable insights into which factors attract people to the lab. These studies also assess whether different invitation treatments attract different types of people. However, they do not investigate selection bias since information about the population is lacking. On the other hand, studies that compare the selected sample to the population at large provide insights into the dimensions of selection. However, they do not assess what information in the recruitment process may mitigate or exaggerate biases, which would allow researchers to construct recruitment protocols that minimize potential selection issues. This study combines these two dimensions by providing a comprehensive investigation of the determinants of experiment participation.

Overall, the existing empirical evidence is mixed regarding the question whether experimental subject pools are selected. For example, Exadaktylos, Espin, and Branas-Garza (2013) as well as Snowberg and Yariv (2018) find little evidence for selection, while Slonim, Wang, Garbarino, and Merret (2013) show significant selection along several dimensions. Likewise, lab subjects differ from particular groups of the population. Falk, Meier, and Zehnder (2014) focus on the comparability of experimental subject pools in terms of social preferences and find no differences between a student sample which voluntarily selected into lab experiments and the student population at the University in Zurich. When they compare the self-selected student participants to voluntary participants from the general Zurich population, however, they find that the general population is more generous on average in the repayment stage of a trust game than the student population. Anderson et al. (2013) measure other-regarding preferences in three samples drawn from college students, non-student adults from the community surrounding the college, and a sample of trainee truckers. They do not find differences between truckers and the non-student adults; but—like Falk, Meier, and Zehnder (2014)—they find that the students are less pro-social than the non-student adults. Harrison, Lau, and Rutström (2009) compare a self-selected sample to a representative sample of a population (in their case, the Danish population) in the context of risk preferences. Similar to our study, they also gather information on the selection process by using randomized variation in a second experiment that varied the recruitment e-mails to prospective student participants. They find that the experimental sample is more risk averse than the population, and that increasing the (certain) show-up fee attracts relatively more risk-averse students. Two studies compare student samples with samples from Amazons Mechanical Turk (MTurk). Arechar, Gächter, and Molleman (2018) show that the basic behavioral patterns in a public goods game found in a typical student sample are replicable with a sample from MTurk. Snowberg and Yariv (2018) compare a student sample with a representative sample from the US population and with a sample from MTurk. In contrast to Arechar, Gächter, and Molleman (2018), the authors find substantial differences between the samples. Student participants are less altruistic and more willing to take risk than participants from the representative and the MTurk samples; student participants also display higher strategic sophistication and cognitive skills than participants from the other two samples.

Our study complements these contributions by comparing the composition of individuals that self-select into experiment participation to the composition of the population from which these individuals are drawn; that is, to the universe of two cohorts of first-year students at a university. While we cannot compare the student sample to the population at large, our analysis provides a step forward by rigorously testing how actual experimental subjects are selected from the whole student population which typically forms the basis for lab experiments. In particular, a classroom experiment reaches individuals that would normally not self-select into lab experiments, but that could do so; this is in contrast to comparing student participants to individuals that probably would never have the opportunity to self-select into participation, such as the entire adult population of a country, as studied by Harrison, Lau, and Rutström (2009), the population of a city as in Falk, Meier, and Zehnder (2014), or a mix of samples of adult students and non-students as in Anderson et al. (2013). Even though our results are based on a student population, they indicate what type of people in other populations may be more or less likely to self-select into lab experiments.

In this sense, our setting is similar to that of Slonim, Wang, Garbarino, and Merret (2013) and Cleave, Nikiforakis, and Slonim (2013), who conducted large scale classroom experiments on different aspects including social (trust game) and risk preferences among students, e.g., in the context of an introductory tutorial class for microeconomics. They find that students that subsequently participate in lab experiments do not significantly differ from non-participants. Investigating subgroups, however, they find that female laboratory participants are on average less risk averse (than female non-participants), while male participants are on average more risk averse (than male non-participants). Our sample differs in that there is no pre-selection of students into courses (e.g., microeconomics) or fields (e.g., economics), since we have access to the entire cohorts of first-year students. Moreover, we elicit a comprehensive list of attitudes and traits. Our setting is thus similar to that of Snowberg and Yariv (2018), who ran an online elicitation of behavioral characteristics among the undergraduates at Caltech and combined it with data on participation in lab experiments. The authors find that the average lab participant is more risk averse, more willing to lie, and less generous than the overall undergraduate population. These differences, however, are small in magnitude. In contrast to Snowberg and Yariv (2018), we conduct a classroom experiment instead of an online survey. Our classroom experiment took place at the beginning of students' undergraduate studies and before the students had to elect majors and courses. Thus, we measure selection based on underlying preferences that are unaffected by course content or educational choices of students.

In terms of our recruitment experiment, our study complements the studies by Harrison, Lau, and Rutström (2009), who vary the show-up fee; Krawczyk (2011), who emphasizes either pecuniary or non-pecuniary benefits in different recruitment e-mails; Lazear et al. (2012), who study sorting into experimental conditions in the context of sharing; and Abeler and Nosenzo (2015), who randomly allocate students to three different recruitment e-mails that either mention monetary rewards, or appeal to the importance of helping research, or both. These studies suggest that emphasizing pecuniary benefits leads more people to sign into a database of subjects for experiments; however, subjects who were recruited via the pecuniary benefits treatment were less likely to participate in a non-paid survey compared to the subjects in the other recruitment conditions. Sign-up rates drop considerably when monetary rewards are not mentioned, but the studies do not find differences across the treatments in terms of selection along social preferences, risk preferences, or cognitive skills. While this literature focuses mainly on risk and social preferences, our study is broader in that we also investigate the effect of self-selection on time preferences, confidence, and cognitive reflection. Adding these preferences or traits is important since a large and expanding literature focuses on cognition, time preference, and overconfidence. Moreover, our randomized experiment includes the different variants of recruitment e-mails studied in the literature in a single setting, and our sample is unusually large (two cohorts of university freshmen). In addition, we study not only subscription to a recruiting data base, but also actual participation in experiments. Thus, our study provides a comprehensive view on selection into participation in lab experiments.

2 Experimental design and procedures

2.1 Setup

The dataset used for this study contains information about behavioral characteristics that were elicited in a classroom experiment (time preferences, risk preferences, overconfidence, and social preferences), cognitive ability (measured by a cognitive reflection test), and background characteristics for two cohorts (2011 and 2012) of first-year students at a public Swiss university, the University of St. Gallen. We combine these measures with results of a randomized recruiting experiment. In this experiment, we invited incoming first-year students to sign up for lab experiments. We randomized four different types of invitation e-mails that each emphasized different motives to participate in experiments. The combined data allows us to study selection into experiments based on both individual characteristics and recruitment conditions.

2.2 Classroom experiment

The pen-and-paper classroom experiment was carried out in the second week of the semester during the last 20 minutes of the students' first Introductory Economics tutorial. Tutorials are groups that meet every other weeks and review the course material together with an instructor. The course Introductory Economics is compulsory for all first-year students,[1] so nearly all students of a cohort participate in the first tutorial sessions.[2] In each cohort, the study was undertaken in 38 tutorials; all of them took place on the same day with the exception of two tutorials (three days later). To ensure that students did not copy answers from their neighbors, we distributed two different sets of instructions, alternating between neighbors. The instructions differed in the order of questions and experimental tasks, so that neighbors would not complete the tasks simultaneously. Student experimenters, who underwent careful training, instructed the students verbally and supervised the experiment. Moreover, we collected classroom identifiers for each tutorial to account for possible dependence of answers within classrooms in our empirical analysis.

Out of all students who attended the tutorials, 90 percent participated in the classroom experiment (93 percent in the first and 86 percent in the second cohort). To verify that our samples are indeed representative for the sample of first-year students, we merged information on participation in the classroom experiment to enrollment records. Out of the two cohorts of first-year students—2,294 students in total—75 percent participated in the classroom experiment (1,722 students). The turnout of the classroom experiment was

[1] Notice that the University of St. Gallen offers undergraduate majors only in the subjects Business Administration, Economics, International Affairs, Legal Studies, and Law & Economics.

[2] In addition, the tutorial is open for students who are not first-year students, but who are required to repeat the course.

different across cohorts, with 86 percent in the first cohort and 65 percent in the second cohort. We thus conduct our main analysis for the pooled two cohorts as well as for the first cohort only.

In terms of student background, the participants in the classroom experiment are representative for the population of first-year students in both cohorts (Table A.1). No significant differences between the two groups exist in terms of age, gender, nationality, and region of origin, i.e., whether the students come from the canton (state) of St. Gallen. An exception is the students' mother tongue: Because the experiment was carried out in German (the languge of instruction in the course), students with a non-German mother tongue were less likely to participate. This difference emerged although we offered an English translation of the instructions.

We financially incentivized the behavioral tasks of our classroom experiment. One subject in each tutorial was randomly determined for pay-out. The amount of money paid out depended on the participants' choices in the incentivized tasks. The randomly selected subjects received on average CHF 94 (≈ $94) in cash. Thus, for those selected for payoff the stakes were high.

2.3 Behavioral measures and cognitive ability

During the classroom experiment, we collected a set of six different behavioral measures in incentivized tasks: measures of time preferences, risk preferences, and overconfidence (one measure each) as well as social preferences (three measures). As a control variable, we collected information on a students' cognitive ability, using a version of Frederick's (2005) Cognitive Reflection Test (CRT). The CRT was not incentivized.

2.3.1 Risk preferences, time preferences, and overconfidence

Risk preferences Our risk elicitation follows Gneezy and Potters (1997) and mimics an investment decision. The students received an endowment of 80 Francs and had to decide which portion of the endowment to invest in a risky asset, and which portion to keep. The portion that was invested in the risky asset yielded a return of 2.5 with a probability of 50 percent and was lost otherwise. For example, if a student invested the whole amount, he/she would receive CHF 200 with a probability of 50 percent, and nothing otherwise. The portion that a student decided to keep was paid out for sure. Thus, a risk-neutral (or risk-seeking) individual would choose to invest the whole amount, while a risk-averse person may invest less. We chose this method for its relative simplicity and promptness: subjects only have to make one choice. This basic setting is easy to understand, whereas more complex methods can produce inconsistent results (Charness, Gneezy, and Imas 2013).

Time preferences To elicit time preferences we adapted the design of Dohmen, Falk, Huffman, and Sunde (2010) and provided the students with a choice table comprising six rows. The students had to make one choice in each row: either they could decide to earn CHF 100 (≈ $100) today or a larger payment in 6 months. Future payments started at CHF 102 in the first row and increased in increments of four up to CHF 122 in the sixth row. One out of the six rows was randomly determined for payout according to the participant's choice. We used the students' switching points from the early payment to a larger future payment as a measure of patience. To minimize the concern that uncertainty about future payouts influenced students' choices, we made all—contemporaneous and future—payments via bank transfer directly after the experiment and in the presence of the student. This procedure was emphasized both at the top of the questionnaire and when we explained the task to elicit time preferences. The instructions read: "[...] we are going to make a bank transaction in your presence in the payment room [after the experiment]." Thus, the participants had a written and ex ante credible account of the payment conditions.

Overconfidence Our measure for overconfidence focuses on relative ability judgments. First, the students had to complete a year-guessing task. They were financially incentivized to state the year of three historical events correctly. If the year was stated correctly, they obtained CHF 20 Francs for each question. For every year that they deviated from the correct answer, we reduced the payout by CHF 2. Second, we incentivized the students to place themselves into deciles according to their relative ability in the year-guessing task. That is, we asked students whether they thought they were among the top 10 percent (top 20 percent, . . . , bottom 10 percent) among all students who completed the guessing task. The students received CHF 10 Francs when they placed themselves into the correct decile.[3] To compute overconfidence, we subtracted a student's guess of his/her own placement in the performance distribution from the actual one. Thus, positive numbers refer to the extent of overconfidence, while negative numbers to underconfidence. A similar procedure is used by Ewers and Zimmermann (2015) and Schulz and Thöni (2016).

2.3.2 Social preferences

Altruism We measure altruism based on a donation decision. At the end of the classroom experiment we asked the students whether they wanted to donate part or all of their potential earnings to a charitable organization. The choice set was not restricted: The

[3]We chose this small incentive compared to the year-guessing task to mitigate potential hedging. That is, a subject answering all three year-guessing questions wrongly forgoes a potential maximal earning of 60 Francs compared to a potential earning of 10 Francs. As a second measure against hedging we made sure that the year guessing-task and the placing task were always on different pages to make the possibility of hedging less salient.

students were provided with a blank field and could choose their preferred organization. Half of the participants in the second cohort (the 2012 cohort) received a list of 5 charitable organizations in addition to the blank field. This treatment had a profound impact on donation behaviour, which we document in a different study (see Schulz, Thiemann, and Thöni, 2018). In our subsequent analysis we control for this treatment intervention to account for differences in donation behavior across cohorts which arises from the list treatment in the second cohort.

Trust and Reciprocity We elicited trust and reciprocity among the students of the second cohort. Our design, a trust game, follows Berg, Dickhaut, and McCabe (1995). Two students were paired with each other. One student (the first mover) received an endowment of CHF 50 and decided which portion of the endowment to transfer to the other student (the second mover). The amount that the first mover transferred was then tripled. Subsequently, the second mover had to decide which amount to send back to the first mover.

We made two main changes to Berg et al.'s (1995) original design. First, in our setup each student took both roles, i.e., the role of the first mover as well as the role of the second mover. This role reversal has two advantages. On the one hand, the students were forced to understand the complete game before they took a decision. On the other hand, we obtained data on the student's behavior in both roles. Second, we employed the strategy method for second mover's decision. To this end, we first constrained the first mover's decision to six options: they could transfer between CHF 0 and CHF 50 in increments of CHF 10. Second, the second movers decided how much to send back conditional on each potential transfer amount. Thus, the second movers had to make six choices. The strategy method allowed us to identify individual degrees of reciprocity by obtaining a full conditional schedule of back-transfers. The strategy method also had a procedural advantage: given the time constraint, it would have been infeasible to match and inform a second mover about a first mover's decision. To pay out a participant, we randomly paired the participants within each tutorial group after the experiment was finished.

In this experiment, selfish second movers who want to maximize their payoff would not send anything back. Moreover, first movers who anticipate a selfish second mover would likely choose a low transfer. Thus, this experiment allows us to assess trust (based on first movers' decision) as well as reciprocity (based on second movers' decisions).

2.3.3 Cognitive Reflection Test

To control for cognitive ability throughout the study, we also measured cognitive abilities using a version of Frederick's (2005) Cognitive Reflection Test (CRT). The CRT measures the ability to suppress a spontaneous and wrong answer in favor of a deliberate and correct answer. We minimized the possibility that the students could copy answers from their neighbors or that they heard the questions from students who took part in the classroom experiment before them. To this end, we created several new questions in the spirit of the original CRT. This ensured that within a tutorial, neighbors would never have the same questions, and later tutorials would not get the same questions that earlier tutorials had received before. We made changes to the original CRT questions by framing the questions differently and by scaling numbers up and down. We also added two additional types of questions, where, in the spirit of the CRT, the intuitive answer was wrong (see Appendix Section A.2 for all questions). In the first cohort we administered two CRT questions; in the second cohort we administered four CRT questions to increase the variation. In our analysis we standardize the CRT measure to have a mean of 0 and a standard deviation of 1 in each cohort and control for cohort dummmies in order to account for the change in the number of questions across cohorts. The CRT was not incentivized.

2.4 Recruiting Experiment

The recruiting experiment was conducted in the context of inviting students to participate in lab experiments by e-mail. The invitation process followed two steps: first, we invited the students to subscribe to a *recruitment database* for future lab experiments; second, we invited the students in the database to participate in specific lab experiments. To be precise, we invited all students of the two cohorts to subscribe to the recruitment database during the first two weeks of the teaching period in their first semester at university (four days before the classroom experiment in the first cohort, and nine days before the classroom experiment in the second cohort). By clicking on a link in the invitation e-mail, the students added their e-mail address to the recruitment database (we use the database system "ORSEE", see Greiner (2015)). We then invited the students to participate in lab experiments throughout the academic year.

In the recruitment experiment, we experimentally varied the e-mail invitation to the recruitment database (and not the invitation to the individual experiments). The invitation to the database was sent out to the two full cohorts of students. Appendix A.3 shows the text of the recruitment e-mails.

We randomized the students into four different treatment conditions. In each of the conditions, the students received similar recruitment e-mails, which differed by just one sentence. This sentence would either highlight the scientific value of the experiments (*GreaterGood*), emphasize the monetary reward (*Money*), mention a guaranteed minimum monetary reward of CHF 10 (*Money10*), or combined the scientific value and the monetary reward argument (*GreaterGoodMoney*):

1. *GreaterGood*: *"These studies provide us with valuable scientific insights."*
2. *Money*: *"By participating you earn money."*
3. *Money10*: *"By participating you earn money (at least 10 CHF)."*
4. *GreaterGoodMoney*: *"These studies provide us with valuable scientific insights. By participating you earn money."*

The students who subscribed to the database received up to five invitations to lab experiments throughout the academic year. These e-mails did not have any treatment variations. As our main outcome variable, we investigate the participation in at least one lab experiment, but we also study whether the recruitment treatment changes the willingness to subscribe to the recruitment database.

In total, 2,363 students received an invitation to subscribe to the database.[4] Out of all students who received a recruitment e-mail, 1,740 students (74 percent) participated in the classroom experiment.[5] The sample of 1,740 students is the main sample that we use for the analysis of selection into experiments. In addition, we use the full sample of 2,363 students to study the effect of the recruitment treatments.

Table A.2 summarizes the treatment variation, the subscription and participation rates for two samples: for all students who received a recruitment e-mail, and for the sub-sample of students who participated in the classroom experiment. As intended in the randomization, each treatment reaches about a quarter of the students (Table A.2). Restricting the sample to students that participated in lab experiments does not change the distribution of treatments—indicating that the recruitment treatments did not affect the participation in the tutorials and the classroom experiment.

Overall, about 20 percent of the students who received a recruitment e-mail enrolled into the recruitment database, and 11 percent of the students participated in at least one experiment (Table A.2). Conditional on participating, the students participated on average in 1.5 experiments; although the students could participate in up to five experiments, a negligible number of students participated in more than two experiments (7 students). Students in the classroom experiment subscribed to the database at a higher

[4]Notice that this number is three percent higher than the number of first-year students that we obtained from enrollment records (2,294 students). The numbers differ slightly because they come from two different sources—an e-mail database that we received from the university administration to send out recruitment e-mails, and an administrative database that contained student enrollment information, merged to information on participation in the classroom experiment. Since the two databases do not contain the same identifier, we cannot assess which of the students who received a recruitment e-mail were not first-year students.

[5]A small number of students (18 students) are likely not first-year students, but students who repeat the course—this is why this number differs from the number of first-year students who participated in the classroom experiment, reported in Section 2.2 (there, we speak of 1,722 students).

rate, compared to the overall population (23 percent versus 20 percent), but the rates of participation in at least one lab experiment are almost identical across the two samples (11 percent versus 11.6 percent).

3 Results

3.1 Behavioral measures

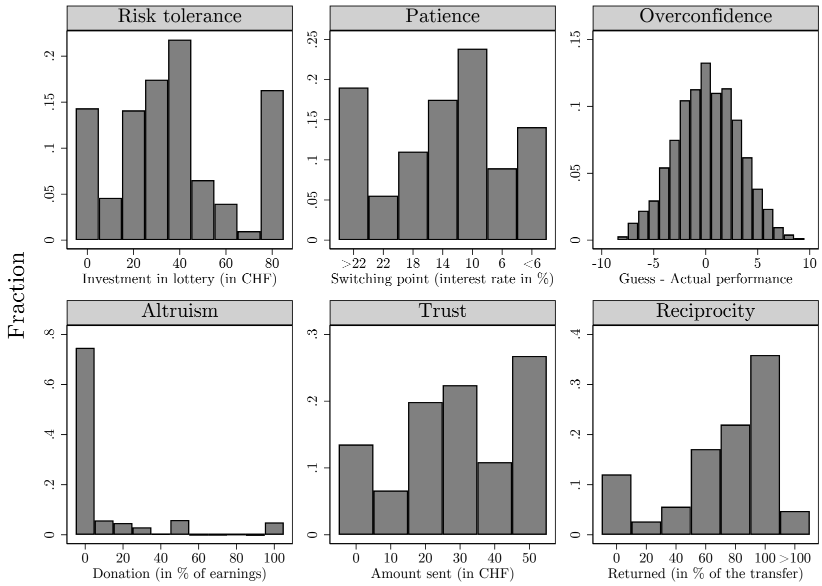
In our classroom experiment almost all students completed the behavioral tasks. More than 94 percent of the participants completed the tasks for time preferences, risk preferences, and altruism, which were elicited for both cohorts of students, as well as the the trust game that was added for the cohort of 2012. Moreover, 91 percent of the participants completed the CRT. The response rate was lower for the overconfidence task (81 percent). Table A.3 displays the sample sizes and summary statistics of all behavioral measures.

Our experimental design yields substantial variation in the behavioral measures (see Figure 1). In the domains risk and patience, the students chose all possible answers; we observe, however, bunching at the extremes. The overconfidence measure is surprisingly symmetrically distributed; the mean is positive but close to zero, indicating that the students on average display only slight overconfidence in the knowledge task. We also obtain variation in the level of altruism: about one-quarter of students is willing to donate part of their experimental earnings to charity, and the contributions range from 0 percent to 100 percent of experimental earnings. Finally, in the trust game, both the receiver and the sender task yield variation in the choices along the full support of possible choices.

To facilitate the interpretation of behavioral measures, we standardize all measures to have a mean of 0 and a standard deviation of 1, with the exception of altruism, which we express as the fraction of experimental earnings donated to charity.

The six behavioral measures that we include in the analysis capture distinct aspects of human behavior. We provide two pieces of evidence for this claim. First, we present a correlation matrix of the six behavioral measures (see Appendix table A.4). The correlations among the measures are small in magnitude, yet we find some significant correlations. Overconfidence and willingness to take risk are significantly and positively correlated with each other. Moreover, we find significant correlations among the measures of social preferences, i.e., between altruism and trust, and between trust and reciprocity. Trust is also positively correlated with patience and willingness to take risk. However, with the exception of the correlation between trust and reciprocity, which are measured in the same trust game, all of the correlations are small (0.16 or less). Second, we conduct a principal component analysis (PCA, see Appendix Table A.5). We find no clear evidence that the

Figure 1: Distribution of behavioral measures



Note: Histograms of behavioral measures. The data is pooled from both cohorts, except for the measures of trust and reciprocity, which are only collected in the cohort of 2012.

measures can be reduced to less than six components; all eigenvalues of the PCA are close to 1. Based on the low correlations and high eigenvalues in the PCA, we argue that it is important to investigate selection along each of the behavioral measures separately.

Througout the analysis, we present results with and without controlling for cognitive ability, since cognitive ability is significantly correlated with willingness to take risk and patience. This result is both shown in prior research (Dohmen, Falk, Huffman, and Sunde 2010) and evident in our data (see Table A.4).

3.2 Recruiting experiment

In evaluating the recruitment experiment, our focus is on the comparison between the *Money* and the *GreaterGoodMoney* treatments, between the *GreaterGood* and the *GreaterGoodMoney* treatments, and between the *Money* and the *Money10* treatments. We find the following results:

> **Result 1**—Mentioning financial incentives in addition to benefits to society increases both the probability to subscribe to the recruitment database and the probability to participate in experiments.

> **Result 2**—Mentioning benefits to society in addition to financial incentives neither affects the probability to subscribe to the recruitment database nor the probability to participate in experiments.

Result 3—Mentioning a certain show-up fee in addition to financial incentives for participation neither affects the probability to subscribe to the recruitment database nor the probability to participate in experiments.

These findings are illustrated in Figure 2, which shows the subscription rates to the recruitment database and the participation rates in at least one lab experiment among all students who received a recruitment e-mail. Mentioning benefits to society only (*GreaterGood*) leads to a subscription rate of 15.0 percent and to a participation rate of 8.4 percent. Including information on financial incentives in addition to societal benefits (*GreaterGood-Money*) increases the subscription rate by 8.3 percentage points—an increase of 55 percent over the baseline—and the participation rate by 4.3 percentage points—an increase of 51 percent over the baseline. These large increases are significant (the p-values of Wilcoxon ranksum tests are $p < 0.000$ for subscription and $p < 0.017$ for participation).[6] Our findings are line with prior evidence by Abeler and Nosenzo (2015), who study the effect of mentioning monetary incentives on subscriptions to an experimental data base. We show in addition that mentioning financial incentives not only increases subscription rates, but also affects actual participation. Overall, we conclude that financial motives are critical for students' participation in experiments.

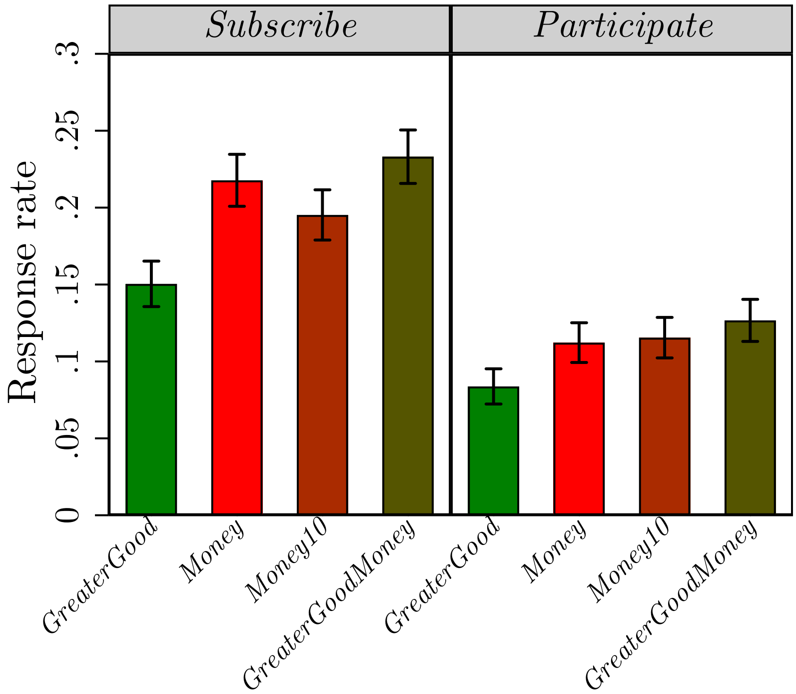
By contrast, mentioning benefits to society in addition to the financial incentives does not significantly alter the decision to subscribe to the recruitment database, nor does it significantly affect the decision to participate in experiments. In the *Money* treatment, the subscription rate is only slightly lower than in the *GreaterGoodMoney* treatment (difference of 1.5 percentage points, $p < 0.527$). This finding on subscriptions also confirms prior evidence by Abeler and Nosenzo (2015). The difference in the participation rates in at least one experiment across the two treatments is equally insignificant (difference of 1.4 percentage points, $p < 0.442$).

Finally, mentioning a fixed show-up fee in addition to financial incentives (*Money10*) does not attract more individuals to the database or to the lab than only mentioning financial incentives (*Money*). In fact, the subscription rate is slighly lower when mentioning the fixed show-up fee in addition to financial incentives (difference of -2.3 percentage points, $p < 0.339$), potentially because some individuals adjust their earnings expectations downwards when the amount of CHF 10 is mentioned. We do not find any differences in participation rates across the two treatments (difference of 0.3 percentage points, p-value < 0.861).

Our results also hold up in a regression analysis (see Table 1). We use linear probability models in all our regressions and estimate the models using OLS.[7] Our outcome variables are dummy variables, which indicate whether an individual subscribed to the

[6] In the following, all reported p-values of mean comparisons are based on Wilcoxon ranksum tests.
[7] The results are robust to using logit and probit specifications. Results are available upon request.

Figure 2: Result of the recruiting experiment



Note: The left panel shows the fraction of students who subscribed to the recruiting database in each treatment condition. The right panel shows the fraction of students who participated in at least one experiment in each treatment condition. The sample consists of all students who received a recruitment e-mail ($N = 2,363$).

recruitment database or participated in at least one experiment, respectively. In the regression analysis, we include cohort dummies because participation and subscription rates differ across cohorts.[8] Moreover, we report results for the whole sample of students who received a recruitment e-mail, in addition to the sample of students who participated in the classroom experiment. We find that the results across the different samples are virtually identical; moreover, controlling for cohort does not change our findings.

3.3 Selection into experiments and behavioral measures

In this section, we evaluate the selection into economic experiments along the different experimentally elicited behavioral measures. Before investigating selection into lab ex-

[8]The subscription rates are nearly identical in the two cohorts (24 percent in the first and 23 percent in the second cohorts). The participation rates differ across cohorts because more experiments were offered to students in the first cohort, compared to students in the second cohort. The participation rates were 14 percent in the first cohort and 9 percent in the second cohort.

Table 1: Results of the recruiting experiment

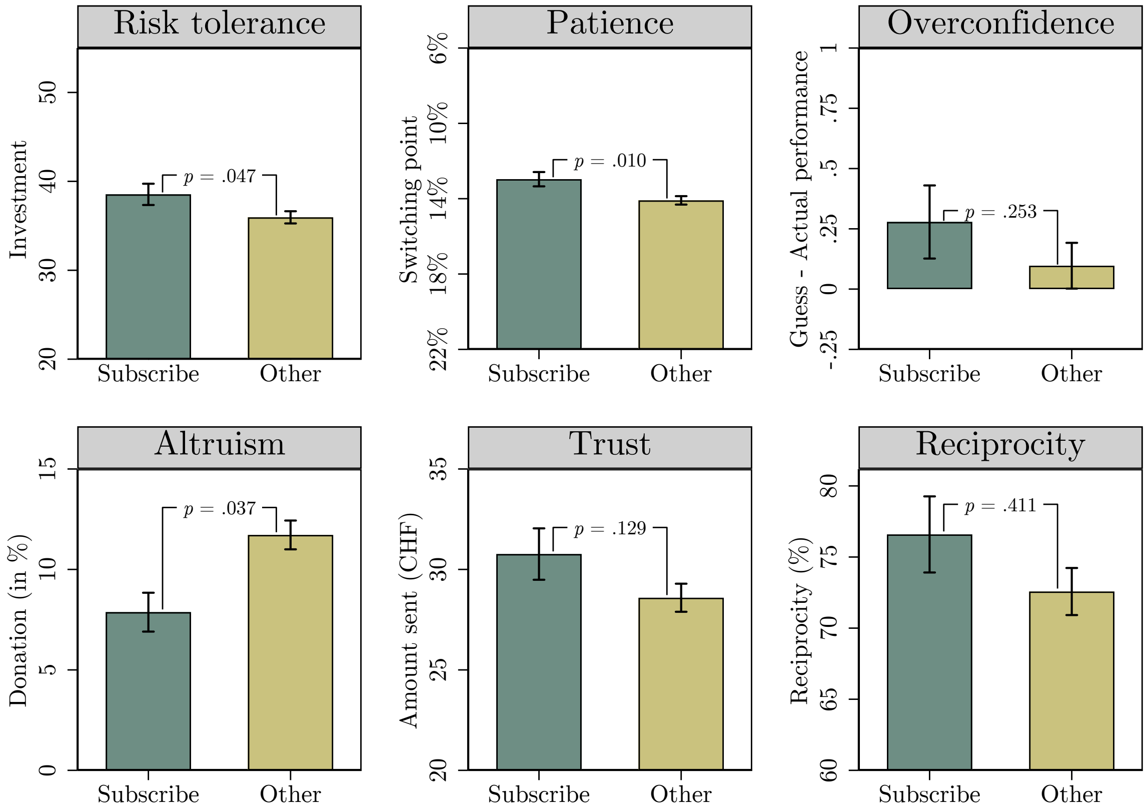
	(1)	(2)	(3)	(4)
	Sample: recruitment e-mail		Sample: classroom experiment	
	subscribe	participate	subscribe	participate
Money	0.067***	0.028	0.065**	0.024
	(0.023)	(0.018)	(0.029)	(0.022)
Money10	0.045*	0.031*	0.026	0.024
	(0.023)	(0.018)	(0.029)	(0.022)
GreaterGoodMoney	0.082***	0.044**	0.083***	0.046**
	(0.023)	(0.018)	(0.029)	(0.022)
Constant	0.146***	0.103***	0.198***	0.112***
	(0.018)	(0.014)	(0.022)	(0.017)
Cohort	✓	✓	✓	✓
p-values of t-tests:				
- GreaterGood vs. GreaterGood-Money	0.000	0.016	0.004	0.034
- Money vs. GreaterGoodMoney	0.513	0.395	0.512	0.310
- Money vs. Money10	0.335	0.893	0.180	0.996
Fraction subscribed/participated	20%	11%	23%	12%
Observations	2,363	2,363	1,740	1,740
R-squared	0.006	0.007	0.006	0.006

Note: Dependent variables are dummy variables for subscription to the recruitment database and participation in at least one experiment. The baseline condition (constant term) is the *GreaterGood* treatment. The results are presented for two samples: all students who received a recruitment e-mail (columns 1 and 2), and the sub-group of students who participated in the classroom experiment (columns 3 and 4). All regressions include a cohort dummy. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

periments, we briefly describe selection into the recruitment database, which indicates a general willingness to participate in lab experiments. Figure 3 shows the averages of the six behavioral measures in the group of students who subscribed, compared to the students who did not subscribe. The results suggest that the sample of students in the database are somewhat more risk tolerant and patient than the non-subscribers, but hardly differ in overconfidence. On the three measures regarding social preferences the data suggests that subscribers are less altruistic when it comes to donations, but tend to be more pro-social in the trust game, albeit the latter differences are not significant. Multivariate regression analysis reveals that subscriptions are higher among students with higher cognitive skills and among students with lower levels of altruism.[9]

[9]See also Appendix (Tables A.6 and A.7).

Figure 3: Selection into the recruitment database



Note: Average behavioral measures for subjects who expressed their willingness to participate in the experiment by subscribing to the database in comparison to the subjects who did not subscribe. Spikes indicate standard errors, p-values from Wilcoxon rank-sum tests.

Compared to selection into the recruitment database, selection into actual participation in lab experiments is more relevant to the question whether laboratory samples adequatley represent the (student) population. For the remainder of the analysis we therefore focus on participation (i.e., participation in at least one experiment). For the first set of measures—willingness to take risk, patience, and overconfidence—we find the following result:

> **Result 4**—We find no evidence that participants of lab experiments are more willing to take risk, more patient, or more overconfident than individuals who do not participate in experiments.

The raw differences in willingness to take risk, patience, and overconfidence between the participants and the non-participants are all small and insignificant. The willingness to take risk is 0.11 standard deviations higher among the participants, compared to the non-participants, but this difference is insignificant (p-value of the Wilcoxon ranksum test is $p < 0.117$). Similarly, patience is 0.10 standard deviations larger among the participants, compared to the non-participants, and this difference is not statistically different either ($p < 0.206$). Moreover, the level of overconfidence does not differ appreciably among the two samples (difference of -0.01 standard deviations, $p < 0.980$).

A regression analysis confirms these findings. Table 2 shows results of OLS regressions of a participation dummy on each of the behavioral measures. In all regressions, we control for the student cohort. In addition, we present regressions with and without controlling for CRT. In particular, cognitive skills may confound the results because risk aversion, patience, and cognitive ability are related (Dohmen, Falk, Huffman, and Sunde 2010). We do not find any significant relationships between the probability to participate in lab experiments and the three behavioral measures; controlling for cognitive ability and gender does not affect this result. Moreover, the results are robust to including the first cohort only (see Table A.8).

Table 2: Participation in at least one experiment based on characteristics: risk, patience, and overconfidence

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: participate					
Risk (std)	0.011	0.007				
	(0.007)	(0.008)				
Patience (std)			0.011	0.008		
			(0.007)	(0.007)		
Overconfidence (std)					0.001	0.005
					(0.010)	(0.011)
CRT score (std)		0.023***		0.026***		0.024**
		(0.009)		(0.008)		(0.010)
Female		-0.008		-0.010		0.005
		(0.017)		(0.017)		(0.019)
Cohort	✓	✓	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓	✓	✓
Observations	1,723	1,578	1,635	1,507	1,415	1,316
Fraction participated	11.7%	12.3%	11.7%	12.3%	12.6%	13.1%
R-squared	0.009	0.014	0.007	0.013	0.008	0.013

Note: All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We consistently find that higher cognitive ability increases the probability of participating in experiments, however. A one-standard deviation increase in cognitive reflection improves the probability to participate by between 2 and 3 percentage points in all regressions (Table 2). This is a large effect and corresponds to about 20 percent over the baseline (the average participation probability is 12 percent). This effect can neither be explained by variation in risk aversion nor variation in patience among individuals with different levels of cognitive reflection (Table 2, columns 2 and 4). Furthermore, the result also holds when inspecting the raw differences between participants and non-participants. The level of CRT is 0.24 standard deviations higher in the sample of participants compared to the sample of non-participants ($p < 0.003$). We thus obtain the following result:

> **Result 5**—Individuals who participate in lab experiments display higher cognitive skills, as measured by a cognitive reflection test, than individuals who do not participate in experiments.

Result 5 is an important finding in favor of the credibility of lab experiments. The individuals who participate in experiments take more time to reflect on a question or task than the average student—even without being incentivized for it. Thus, one may expect that these individuals are also more diligent and careful when performing tasks in the laboratory than the non-participants. This likely benefits the accuracy of results from lab experiments and implies that lab results regarding cognitive bias might even provide a lower bound.

Our second set of measures captures social preferences: altruism, as well as trust and reciprocity. Our results are as follows:

> **Result 6**—We find evidence that participants of lab experiments are less altruistic, compared to individuals who do not participate in experiments. By contrast, we find no evidence that levels of trust and reciprocity differ among participants and non-participants.

First, the level of altruism differs significantly across the samples of participants and non-participants. Participants are willing to donate a lower fraction of their earnings, compared to non-participants.[10] Thus, self-interest—rather than altruism—may attract people to the lab; this observation is in line with our previous finding that financial incentives are more effective in recruiting students to the lab than emphasizing the benefits to society. The difference in the fraction of earnings donated to charity across the samples of non-participants and participants is insignificant when we perform a raw sample comparison (difference of -4 percentage points in the fraction of earnings that a student is willing to donate, $p < 0.152$). However, in our regression analysis we find that a higher level

[10] Recall that these are earnings in the classroom experiments, so both participants and non-participants in the lab experiments answered this question.

of altruism is significantly predictive of a lower participation probability (Table 3): An increase in the fraction of earnings that a student is willing to donate—from no donation to a donation of his/her complete experimental earnings—maps into a 6 percentage points lower participation probability. This corresponds to a decrease in the participation rate by 50 percent compared to the baseline (the participation rate in the full sample is 12 percent). Controlling for CRT, the coefficient decreases slightly in size and is significant at the 10% level. Moreover, the coefficient becomes smaller when we include the first cohort only (see Table A.8). Yet, our evidence suggests that lab participants are more self-interested than individuals who do not participate in lab experiments. This finding is in line with the recent study by Snowberg and Yariv (2018), which documents that the average lab participant is less generous in a dictator game, compared to the average non-participant.

Table 3: Participation in at least one experiment based on characteristics: social preferences

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: participate					
Altruism (frac. donated)	-0.059***	-0.049*				
	(0.022)	(0.026)				
Trust (std)			0.004	0.002		
			(0.010)	(0.011)		
Reciprocity (std)					-0.005	-0.010
					(0.009)	(0.010)
CRT score (std)		0.021**		0.023*		0.023*
		(0.008)		(0.012)		(0.013)
Female		-0.014		-0.021		-0.025
		(0.016)		(0.022)		(0.024)
Cohort	✓	✓	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓	✓	✓
Observations	1,700	1,570	750	646	709	617
Fraction participated	11.6%	12.2%	8.8%	9.4%	9.4%	10.0%
R-squared	0.009	0.013	0.005	0.016	0.005	0.016

Note: All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Second, using the results from the trust game that we added in the second cohort, we do not find any significant differences in the levels of trust and reciprocity between the participants and non-participants. Participants are neither significantly less trusting than non-participants (difference of 0.05 standard deviations, $p < 0.714$), nor are they significantly less reciprocal than non-participants (difference of -0.06 standard deviations,

$p < 0.689$). The regression analysis supports the findings that the levels of trust and reciprocity do not differ across the samples (Table 3, columns 3 and 4). It may seem surprising at first sight that participants are less altruistic than non-participants, but not less reciprocal or trusting. We have two potential explanations for this finding: On the one hand, we measure trust and reciprocity only in the second cohort and thus in a smaller sample. On the other hand, the role of social preferences may differ depending on the task used to elicit social preferences—i.e., whether the task includes a strategic interaction or not. The latter explanation parallels a finding by Snowberg and Yariv (2018): The authors show that the average participant differs from the average non-participant when social preferences are elicited in a dictator game, a game without strategic interaction, but not when social preferences are elicited in a prisoner's dilemma, a game with strategic interaction. We find a similar pattern: Based on our trust game, which involves strategic interaction, we find no differences in social preferences between participants and non-participants. By contrast, based on the donation decision, which is not strategic, we find different levels of altruism between the participants and non-participants.

In sum, we do not find any selection effects along risk preferences, time preferences, overconfidence, as well as trust and reciprocity. We do, however, find sizable and significant selection effects along a measure of altruism and a measure of cognitive reflection. The positive selection in terms of cognitive reflection may benefit the internal validity of experimental results.

3.4 Treatment-specific selection effects

In our recruitment experiment, we varied the e-mail invitation to our recruitment database. In this section, we investigate whether these different recruitment treatments attract different types of students into the lab. Out of the large number of potential interactions we limit our attention to the cases in which we have clear hypotheses about these interactions. We focus on three questions: (1) Does mentioning a certain show-up fee in addition to financial incentives attract more risk-averse individuals to the lab? (2) Does mentioning a certain show-up fee in addition to financial incentives attract students with lower confidence into the lab? (3) Does mentioning financial incentives in addition to benefits for society increase the participation rate of selfish individuals? Similarly, does mentioning financial incentives, but not mentioning the benefits for society discourage altruistic individuals from participating?

To address the first question, we investigate whether the $Money10$ treatment attracts relatively more risk-averse individuals, compared ot the $Money$ treatment. The result is as follows:

Result 7—The selection into participation along willingness to take risk is unaffected by mentioning a certain show-up fee in addition to financial incentives.

Table 4 shows the results of OLS regressions of a participation dummy on the $Money10$ dummy, and interactions of this dummy with willingness to take risk. We do not find that the $Money10$ treatment selects individuals with lower willingness to take risk, compared to the $Money$ treatment. This result also holds when we compare the $Money10$ treatment with all other treatments, none of which mentions a certain show-up fee.

Table 4: Interaction effects: Mentioning a safe amount and selection based on willingness to take risk and overconfidence

	(1)	(2)	(3)	(4)
	Dependent variable: participate			
Comparison	Money10 vs. Money		Money10 vs. all other treatments	
Money10	-0.001	-0.004	0.006	0.007
	(0.024)	(0.027)	(0.019)	(0.022)
× Risk	0.025		0.000	
	(0.022)		(0.021)	
× Overconfidence		0.024		-0.014
		(0.025)		(0.019)
Cohort	✓	✓	✓	✓
CRT score (std)	✓	✓	✓	✓
Risk (std)	✓		✓	
Overconfidence (std)		✓		✓
Observations	781	662	1,581	1,317
Fraction participated	12.9%	12.3%	13.9%	13.1%
R-squared	0.007	0.007	0.011	0.010

Note: The sample in columns 1 and 2 consists of only individuals in the $Money10$ and $Money$ treatments. The sample in columns 3 and 4 consist of the full sample. All regressions include a cohort dummy and control for CRT and the main effects. Columns 1 and 3 control for the main effect of risk preferences, colums 2 and 4 control for the main effect of overconfidence. Robust standard errors, clustered at tutorial level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

This finding is different from the results by Harrison, Lau, and Rutström (2009), who show that increasing the amount of a certain show-up fee attracts more risk-averse students. However, different from their setting, the certain payment is unchanged in our setting; we only manipulate the information that students receive before signing up to the recruitment database. In this setting, simply mentioning the show-up fee does not alter the composition of the participant sample.

Second, we test whether mentioning a safe amount reduces the level of overconfidence in the participant sample. Again, we compare the $Money10$ treatment with the $Money$

treatment and with all other treatments. We do not find any significant interaction effects between the *Money*10 treatment and overconfidence (Table 4), thus leading to our result:

> **Result 8**—The selection into participation along overconfidence is unaffected by mentioning a certain show-up fee in addition to financial incentives.

Third, we test how the composition of the participant sample changes if we switch the mentioning of benefits to society on and off, and if we switch the mentioning of financial incentives on and off. We find the following two results:

> **Result 9**—The selection into participation along altruism is unaffected by mentioning financial incentives in addition to benefits to society.

> **Result 10**—The selection into participation along altruism is unaffected by mentioning benefits to society in addition to, or instead of, financial incentives.

Table 5 compares the selection based on altruism between the *GreaterGoodMoney* and the *GreaterGood* conditions, as well as between the *GreaterGoodMoney* and the *Money* conditions. We do not find that mentioning financial incentives in addition to societal benefits attracts less altruistic individuals on average (column 1); similarly, we do not find that mentioning the benefits to society in addition to—or instead of—financial incentives attracts more altruistic individuals (columns 2 and 3). Our results are in line with Abeler and Nosenzo (2015) who do not detect any differences in pro-sociality between recruitment treatments that emphasize the social value of participation, and the ones that do not emphasize this social value. Our findings, however, contrast with earlier evidence by Krawczyk (2011) who show that individuals who were recruited by emphasizing monetary (as opposed to non-monetary) benefits were less likely to display pro-social behavior (i.e., participating in a non-paid survey) and were also less altruistic in general.

In sum, our evidence suggests that variations in the recruitment treatments do not alter the participant sample: Mentioning financial incentives, benefits to society, and a certain show-up fee do not change the average level of willingness to take risk, overconfidence, and altruism among the participants.[11] This implies that choosing the recruitment e-mail that maximizes participation—in this case, the *GreaterGoodMoney* e-mail—does not result in a more selected sample compared to the other recruitment treatments.

[11] For completeness, we also investigate the selection under the different recruitment conditions along all behavioral characteristics that we collected (seven characteristics, including CRT). Appendix Table A.11 presents the results, where we focus only on *ceteris paribus* treatment variations, i.e., comparisons between the *GreaterGoodMoney* and the *GreaterGood* treatments, between the *GreaterGoodMoney* and the *Money* treatments, and between the *Money* and the *Money*10 treatments. Out of the 21 possible comparisons, only one difference is significant. Thus, we conclude that the recruitment method does not have any overall effect on selection along any of the seven characteristics we study.

Table 5: Interaction effects: mentioning financial incentives and selection based on pro-sociality

	(1)	(2)	(3)
	Dependent variable: participate		
Comparison	GreaterGoodMoney vs. GreaterGood	GreaterGoodMoney vs. Money	GreaterGood vs. Money
Altruism	-0.079	-0.079*	-0.078
	(0.052)	(0.041)	(0.052)
× GreaterGoodMoney	0.013	0.013	
	(0.080)	(0.064)	
× GreaterGood			-0.002
			(0.068)
Cohort	✓	✓	✓
CRT score (std)	✓	✓	✓
Recruitment treatments	✓	✓	✓
Observations	800	792	770
Fraction participated	13.4%	11.7%	11.2%
R-squared	0.018	0.023	0.021

Note: The sample in column 1 consists of the individuals in the *GreaterGoodMoney* and *GreaterGood* treatments, the sample in column 2 consist of the individuals in the *GreaterGoodMoney* and *Money* treatments, and the sample in column 3 consist of the individuals in the *GreaterGood* and *Money* treatments. All regressions include a cohort dummy and control for CRT and the main effects of the recruitment treatments. Robust standard errors, clustered at tutorial level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

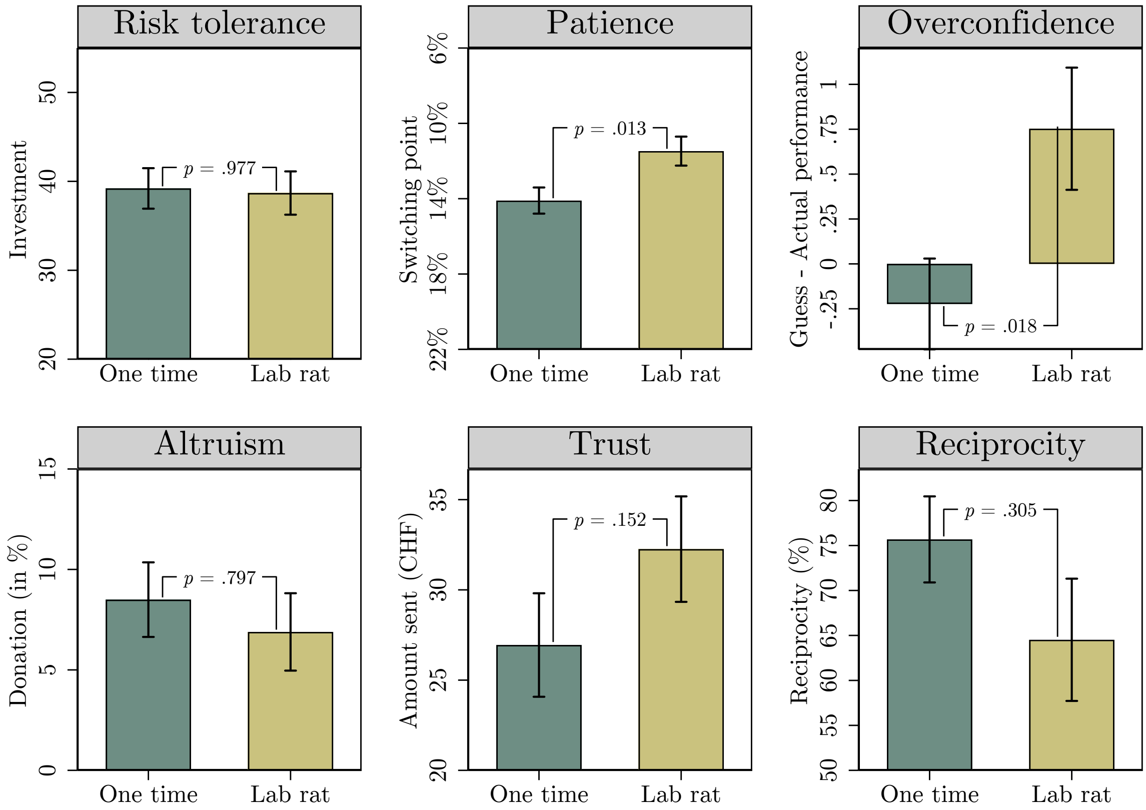
3.5 One timers vs. lab rats

For those students who subscribed to the database we can not only track whether they participated in experiments, but we can also check whether they participated repeatedly. This allows us to address the concern that subjects with prior lab experience may systematically differ from inexperienced subjects. Matthey and Regner (2013) find a negative correlation between generous behavior in allocation decisions and the number of previous participations in other experimental sessions. In a similar vein, Benndorf, Moellers, and Normann (2017) find that experienced subjects show less trustworthiness and trust than inexperienced subjects. Finding differences between experienced and inexperienced subjects begs the question of whether this is (i) due to sample selection (only some subjects show up repeatedly), or (ii) a response to experiences made in the lab. With the data from our study we can shed some light on the first channel.[12] In particular, our experimental setup allows us to investigate whether there are ex-ante differences between subjects who participate only once and subjects who participate repeatedly in experimental sessions. There is no prior evidence in the literature regarding what sort of effects to expect. The following analysis is therefore rather exploratory.

Figure 4 depicts the behavioral measures of one-time participants relative to subjects who repeatedly participated in experimental sessions ('Lab rat' in the figure). Recall that the behavioral measures have been collected prior to any participation, i.e., it can be ruled out that the results are influenced by experiences made in the laboratory. The standard errors in Figure 4 are larger than in Figure 3 because the sample is restricted to the 259 subjects who participated in lab experiments. Among these, 58 percent participated only once, almost all of the remaining subjects participated twice in experiments, and a very small number participated three to five times. Significant differences emerge for two measures. Repeated participation is significantly associated with higher levels of patience, as well as higher levels of overconfidence. For the trust game we find substantial but insignificant differences, suggesting less trust but more trustworthiness (reciprocity) among one-time participants. Part of the differences between one-time and repeated participation might be due to gender differences. Guillen and Veszteg (2012) show that male subjects are more likely to participate repeatedly. In regression analysis we confirm that the differences documented in Figure 4 remain unchanged when controlling for gender (see Appendix Table A.10).

[12] There is a (small) literature on the second channel, which studies the same set of subjects repeatedly over time. Brosig, Riechmann, and Weimann (2007) find that subjects become more selfish in subsequent sessions, while Volk, Thöni, and Ruigrok (2012) find no systematic changes in preferences for cooperation.

Figure 4: Lab rats and onetimers



Note: Average behavioral measures for subjects who participated in one experiment (One-time) or more than one experimental session (Labrat). Spikes indicate standard errors, *p*-values from Wilcoxon rank-sum tests.

4 Conclusion

This paper provides novel evidence on the influence of information provided in the recruitment process on participation rates in lab experiments, on selection into lab experiments based on an extensive battery of behavioral measures, and on the interplay between these two aspects. The analysis is based on a unique combination of classroom experiments conducted among two cohorts of first-year university students to elicit their preferences across various domains as well as cognitive abilities and self-confidence, with a recruiting experiment which varied the information contained in invitations to participate in lab experiments.

Taken together, the empirical results from the recruitment experiment indicate that providing information about financial rewards increases subscriptions to the subject pool and ultimately participation rates in lab experiments. We find little evidence, though, that emphasizing monetary rewards during the recruitment process affects selection of lab participants with particular characteristics. Overall, we find that participants in lab experiments exhibit higher cognitive skills and tend to be less altruistic than non-participants. For the other behavioral measures (risk and time preferences, trust, reciprocity, and self-confidence), we do not find evidence for selection. Additional findings suggest that repeated participation is selected for by subjects with higher patience and self-confidence.

These findings have relevant implications for the interpretation of experimental findings from student populations. In particular, the results suggest that the selection of subject pools in lab experiments along behavioral measures and the role of recruiting procedures for this selection might be less of a concern than suggested previously, at least as far as student subject pools are concerned. An exception is selection along cognitive skills: We document higher cognitive skills among participants of lab experiments, compared to non-participants. Lab results regarding cognitive biases might thus provide a lower bound for cognitive biases in a population.

Acknowledgements. We thank seminar participants at the University of Zurich and Lund University for helpful comments. We thank Martin Kolmar for giving us access to the tutorials. We are indebted to Sharon Pfister and Darjusch Tafreschi for their invaluable help in the design of the questionnaires as well as the preparation and implementation of the data collection. We thank numerous administrators, tutors, and helpers at the University of St.Gallen for their cooperation in the data collection process. Financial support from the SNSF (grant: 100018_135166) and the European Research Council (AdG 295707) is gratefully acknowledged.

References

Abeler, J., and D. Nosenzo (2015): "Self-selection into laboratory experiments: Pro-social motives vs. monetary incentives," *Experimental Economics*, 18(2), 195–214.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2010): "Preference Heterogeneity in Experiments: Comparing the Field and Laboratory," *Journal of Economic Behavior and Organization*, 73(2), 209–224.

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

Arechar, A. A., S. Gächter, and L. Molleman (2018): "Conducting interactive experiments online," *Experimental Economics*, 21(1), 99–131.

Benndorf, V., C. Moellers, and H.-T. Normann (2017): "Experienced vs. inexperienced participants in the lab: Do they behave differently?," *Journal of the Economic Science Association*, 3(1), 12–25.

Berg, J., J. Dickhaut, and K. McCabe (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10(1), 122 – 142.

Brosig, J., T. Riechmann, and J. Weimann (2007): "Selfish in the End? An Investigation of Consistency and Stability of individual Behavior," MPRA Paper No. 2035.

Charness, G., U. Gneezy, and A. Imas (2013): "Experimental methods: Eliciting risk preferences," *Journal of Economic Behavior & Organization*, 87, 43–51.

Cleave, B., N. Nikiforakis, and R. Slonim (2013): "Is there selection bias in laboratory experiments? The case of social and risk preferences," *Experimental Economics*, 16(3), 372–382.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010): "Are risk aversion and impatience related to cognitive ability?," *American Economic Review*, 100(3), 1238–60.

Ewers, M., and F. Zimmermann (2015): "Image and misreporting," *Journal of the European Economic Association*, 13(2), 363–380.

Exadaktylos, F., A. M. Espin, and P. Branas-Garza (2013): "Experimental subjects are not different," *Nature: Scientific Report*, 1213.

Falk, A., and J. J. Heckman (2009): “Lab Experiments are a Major Source of Knowledge in the Social Sciences,” *Science*, 326(5952), 535–538.

Falk, A., S. Meier, and C. Zehnder (2014): “Do Lab Experiments Misrepresent Social Preferences? The Case of Self-Selected Student Samples,” *Journal of the European Economic Association*, 11(4), 839–852.

Frederick, S. (2005): “Cognitive reflection and decision making,” *Journal of Economic Perspectives*, 19(4), 25–42.

Gneezy, U., and J. Potters (1997): “An experiment on risk taking and evaluation periods,” *The Quarterly Journal of Economics*, 112(2), 631–645.

Greiner, B. (2015): “Subject pool recruitment procedures: organizing experiments with ORSEE,” *Journal of the Economic Science Association*, 1(1), 114–125.

Guillen, P., and R. F. Veszteg (2012): “On “lab rats”,” *Journal of Socio-Economics*, 41(5), 714–720.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2009): “Risk attitudes, randomization to treatment, and self-selection into experiments,” *Journal of Economic Behavior and Organization*, 70(3), 498–507.

Krawczyk, M. (2011): “What brings your subjects to the lab? A field experiment,” *Experimental Economics*, 14(4), 482–489.

Lazear, E. P., U. Malmendier, and R. A. Weber (2012): “Sorting in Experiments with Application to Social Preferences,” *American Economic Journal: Applied Economics*, 4(1), 136–163.

Matthey, A., and T. Regner (2013): “On the independence of history: Experience spill-overs between experiments,” *Theory and Decision*, 75(3), 403–419.

Schulz, J. F., P. Thiemann, and C. Thöni (2018): “Nudging generosity: Choice architecture and cognitive factors in charitable giving,” *Journal of Behavioral and Experimental Economics*, 74, 139–145.

Schulz, J. F., and C. Thöni (2016): “Overconfidence and career choice,” *PloS one*, 11(1), e0145126.

Slonim, R., C. Wang, E. Garbarino, and D. Merret (2013): “Opting-in: Participation Bias in Economic Experiments,” *Journal of Economic Behavior and Organization*, 90, 43–70.

Snowberg, E., and L. Yariv (2018): "Testing the Waters: Behavior across Participant Pools," *NBER Working Paper*, 24781.

Volk, S., C. Thöni, and W. Ruigrok (2012): "Temporal stability and psychological foundations of cooperation preferences," *Journal of Economic Behavior & Organization*, 81(2), 664–676.

von Gaudecker, H.-M., A. van Soest, and E. Wengström (2012): "Experts in Experiments," *Journal of Risk and Uncertainty*, 45(2), 159–190.

Appendix

A.1 Tables

Table A.1: Representativeness of the participants in the classroom experiment for the full population of first-year students

	(1) Classroom experiment	(2) Full sample	(3) Difference	(4) p-value
Panel A. Both cohorts				
Age	19.94	20.02	-0.08	0.168
Male	0.65	0.66	-0.01	0.565
Nationality: Swiss	0.76	0.77	-0.01	0.581
Nationality: German/Austrian	0.19	0.18	0.01	0.364
Nationality: Other	0.05	0.06	0.00	0.614
Non-German mother tongue	0.09	0.11	-0.03	0.009
Entrance test	0.18	0.16	0.01	0.252
High school St. Gallen	0.13	0.13	0.00	0.655
# Observations	1,722	2,294		
% of full sample	75%			
Panel B. Cohort of 2011				
Age	19.97	20.01	-0.04	0.589
Male	0.66	0.67	-0.01	0.753
Nationality: Swiss	0.76	0.77	-0.01	0.717
Nationality: German/Austrian	0.19	0.18	0.01	0.553
Nationality: Other	0.05	0.06	0.00	0.743
Non-German mothertongue	0.08	0.10	-0.02	0.050
Entrance test	0.18	0.17	0.01	0.472
High school St. Gallen	0.13	0.13	0.00	0.774
# Observations	955	1,111		
% of full sample	86%			
Panel C. Cohort of 2012				
Age	19.91	20.02	-0.12	0.146
Male	0.64	0.65	-0.01	0.524
Nationality: Swiss	0.76	0.77	-0.01	0.672
Nationality: German/Austrian	0.19	0.18	0.01	0.489
Nationality: Other	0.05	0.06	0.00	0.706
Non-German mother tongue	0.10	0.12	-0.02	0.123
Entrance test	0.17	0.15	0.01	0.442
High school St. Gallen	0.14	0.13	0.01	0.710
# Observations	767	1,183		
% of full sample	65%			

Note: The table compares student background characteristics among the overall population of first-year students and the participants in the classroom experiment for all students who could be identified as first-year students in administrative enrollment records.

Table A.2: Recruitment experiment: Treatments and participation probabilities

	(1)	(2)	(3)	(4)
	Sample: recruitment e-mail		Sample: classroom experiment	
	Obs.	Mean	Obs.	Mean
Panel A. Recruitment treatments				
GreaterGood	2,363	24.8%	1,740	24.4%
Money	2,363	25.3%	1,740	24.8%
Money10	2,363	24.9%	1,740	25.2%
GreaterGoodMoney	2,363	25.1%	1,740	25.6%
Panel B. Recruitment results				
Subscribed to database	2,363	19.9%	1,740	23.4%
Participated				
At least one experiment	2,363	11.0%	1,740	11.6%
Number participated	259	1.47	201	1.44
One experiment	259	57.5%	201	57.7%
More than one experiment	259	42.5%	201	42.3%

Note: Panel A shows the treatment probabilities in the sample of all students who received the recruitment e-mail, as well as among the experiment participants. Panel B shows the probabilities of subscribing to the database and of participating in the experiments. The maximum possible number of experiments a student could participate in was 5.

Table A.3: Summary statistics of the behavioral measures and CRT collected in the classroom experiment

	(1)	(2)	(3)	(4)	(5)
A. Behavioral measures and CRT collected in both cohorts (N = 1,740)					
	N	Mean	SD	Min	Max
Risk	1,724	36.50	25.02	0	80
Patience	1,636	3.05	1.98	0	6
Overconfidence	1,416	0.13	3.07	-8	9
Altruism (fraction donated)	1,700	0.11	0.25	0	1
CRT score	1,593	1.75	1.20	0	4
B. Behavioral measures collected only in the second cohort (N = 754)					
	N	Mean	SD	Min	Max
Trust	751	29.08	16.85	0	50
Reciprocity	710	217.49	112.21	0	760

Note: The table shows summary statistics for the behavioral measures and CRT collected in the classroom experiment. Trust and reciprocity were only collected in the second cohort of the experiment.

Table A.4: Correlation among behavioral measures

	(1) Risk	(2) Patience	(3) Over-confidence	(4) Altruism	(5) Trust	(6) Reciprocity
Patience (std)	-0.020					
	(0.412)					
Overconfidence (std)	0.059	-0.026				
	(0.027)	(0.338)				
Altruism (fraction donated)	0.033	0.062	-0.069			
	(0.176)	(0.013)	(0.010)			
Trust (std)	0.137	0.152	-0.014	0.063		
	(0.000)	(0.000)	(0.763)	(0.088)		
Reciprocity (std)	0.003	0.087	-0.028	0.085	0.327	
	(0.941)	(0.025)	(0.545)	(0.025)	(0.000)	
CRT score (std)	0.140	0.090	-0.002	-0.035	0.051	0.064
	(0.000)	(0.001)	(0.932)	(0.168)	(0.196)	(0.113)

Note: The table displays pairwise correlations between the behavioral measures and CRT used in the main analysis, based on the individuals who took part in the classroom experiment and completed the respective task. All measures are standardized to have a mean of 0 and a standard deviation of 1, except for altruism, which is the amount donated as a fraction of experimental earnings. P-values are in parentheses.

Table A.5: Principal component analysis of behavioral measures

	(1)	(2)	(3)	(4)
	A. Both cohorts (N = 1,317)		B. Only 2^{nd} cohort (N = 427)	
Component	Eigenvalue	Proportion explained	Eigenvalue	Proportion explained
1	1.116	0.279	1.430	0.238
2	1.045	0.261	1.145	0.191
3	0.973	0.243	1.007	0.168
4	0.866	0.217	0.963	0.161
5	-	-	0.795	0.133
6	-	-	0.659	0.110
TOTAL		1		1

Note: The table displays the results of a principal component analysis of the behavioral measures. Columns 1 and 2 show the results for the four behavioral measures that were collected in both cohorts: willingness to take risk, patience, overconfidence, and altruism. Columns 3 and 4 show the results including the measures collected only in the second cohort: trust and reciprocity. Columns 1 and 3 display the eigenvalue of each component, and columns 2 and 4 show the proportion of the total variation explained by each of the components.

Table A.6: Subscription to the recruitment database based on characteristics: risk, patience, and overconfidence

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: subscribe					
Risk (std)	0.018*	0.013				
	(0.011)	(0.012)				
Patience (std)			0.022**	0.019*		
			(0.010)	(0.011)		
Overconfidence (std)					0.010	0.015
					(0.012)	(0.013)
CRT score (std)		0.042***		0.046***		0.037***
		(0.011)		(0.011)		(0.012)
Female		0.018		0.013		0.014
		(0.024)		(0.023)		(0.025)
Cohort	✓	✓	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓	✓	✓
Observations	1,723	1,578	1,635	1,507	1,415	1,316
Subscribed in sample	23.6%	24.6%	24.0%	24.9%	24.5%	25.3%
R-squared	0.008	0.017	0.006	0.018	0.010	0.017

Note: All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.7: Subscription to the recruitment database based on characteristics: social preferences

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: subscribe					
Altruism (frac. donated)	-0.123***	-0.123***				
	(0.032)	(0.034)				
Trust (std)			0.025	0.021		
			(0.016)	(0.016)		
Reciprocity (std)					0.025	0.015
					(0.016)	(0.016)
CRT score (std)		0.042***		0.069***		0.068***
		(0.011)		(0.016)		(0.018)
Female		0.007		0.073**		0.067*
		(0.023)		(0.031)		(0.035)
Cohort	✓	✓	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓	✓	✓
Observations	1,700	1,570	750	646	709	617
Subscribed in sample	23.6%	24.5%	22.5%	23.8%	23.8%	25.0%
R-squared	0.011	0.020	0.014	0.041	0.013	0.038

Note: All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Participation in at least one experiment based on behavioral characteristics—first cohort only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: participate							
Risk (std)	0.009	0.004						
	(0.010)	(0.011)						
Patience (std)			0.004	0.003				
			(0.010)	(0.010)				
Overconfidence (std)					0.003	0.007		
					(0.015)	(0.015)		
Altruism (frac. donated)							-0.044	-0.040
							(0.035)	(0.037)
CRT score (std)		0.023*		0.025**		0.021*		0.020*
		(0.012)		(0.012)		(0.012)		(0.011)
Female		-0.002		-0.004		0.006		-0.009
		(0.023)		(0.023)		(0.024)		(0.022)
Recruitment treatments	✓	✓	✓	✓	✓	✓	✓	✓
Observations	983	938	934	898	915	881	974	932
Fraction participated	13.6%	14.1%	13.6%	13.9%	14.3%	14.6%	13.6%	13.9%
R-squared	0.003	0.007	0.002	0.007	0.004	0.007	0.003	0.006

Note: The sample is restricted to the cohort of 2011. All regressions include dummies for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Participation in at least one experiment: joint test of behavioral measures

	(1)	(2)	(3)	(4)
	Dependent variable: participate			
Risk (std)	0.012	0.008	0.013	0.006
	(0.009)	(0.010)	(0.016)	(0.018)
Patience (std)	0.005	0.002	0.009	0.005
	(0.008)	(0.008)	(0.014)	(0.015)
Overconfidence (std)	-0.001	0.003	-0.006	-0.003
	(0.010)	(0.011)	(0.011)	(0.013)
Altruism (fraction donated)	-0.037	-0.026	-0.038	-0.021
	(0.028)	(0.031)	(0.050)	(0.062)
Trust (std)			-0.007	-0.006
			(0.016)	(0.016)
Reciprocity (std)			-0.003	-0.008
			(0.012)	(0.012)
CRT score (std)		0.024**		0.032*
		(0.010)		(0.017)
Female		0.005		0.012
		(0.020)		(0.036)
Cohort	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓
p-value of F-test: joint significance of behavioral measures	0.39	0.80	0.73	0.98
Observations	1,317	1,239	427	381
Fraction participated	12.7%	13.1%	10.5%	11.0%
R-squared	0.007	0.012	0.007	0.022

Note: All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. The table reports and F-test for the joint significance of risk, patience, overconfidence, and altruism in columns 1 and 2, and for the joint significance of risk, patience, overconfidence, the altruism, trust, and reciprocity in columns 3 and 4. Robust standard errors, clustered at tutorial level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A.10: One-time participants vs. lab rats

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: participate more than once							
Risk (std)	-0.033						0.011	-0.035
	(0.033)						(0.037)	(0.102)
Patience (std)		0.106***					0.108**	0.058
		(0.036)					(0.043)	(0.076)
Overconfidence (std)			0.101**				0.106**	0.033
			(0.044)				(0.045)	(0.130)
Altruism (frac. donated)				-0.074			-0.193	-0.336
				(0.186)			(0.145)	(0.397)
Trust (std)					0.064			0.140
					(0.064)			(0.104)
Reciprocity (std)						-0.080		-0.078
						(0.066)		(0.073)
CRT score (std)	-0.012	-0.031	-0.018	-0.010	0.005	0.009	-0.032	-0.057
	(0.038)	(0.039)	(0.041)	(0.037)	(0.056)	(0.053)	(0.043)	(0.108)
Female	-0.158*	-0.141*	-0.138	-0.137*	-0.161	-0.185	-0.118	-0.196
	(0.082)	(0.075)	(0.086)	(0.082)	(0.140)	(0.132)	(0.083)	(0.260)
Cohort	✓	✓	✓	✓	✓	✓	✓	✓
Recruitment treatments	✓	✓	✓	✓	✓	✓	✓	✓
p-value of F-test: joint significance of behavioral measures							0.006	0.007
Observations	194	186	172	192	61	62	162	42
Fraction part. more than once	43.8%	43.5%	43.6%	44.3%	50.8%	50.0%	43.8%	54.8%
R-squared	0.041	0.087	0.080	0.033	0.091	0.099	0.137	0.249

Note: The sample consists of all students who participated in at least one experiment. The dependent variable is coded as one if the student participated in more than one experiment and as zero otherwise. All regressions include dummies for the cohort and for the treatment group in the recruitment experiment. Robust standard errors, clustered at tutorial level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Treatment-specific selection effects: All results

	(1)	(2)	(3)
	Dependent variable: participate		
Comparison	GreaterGoodMoney vs. GreaterGood	GreaterGoodMoney vs. Money	Money vs. Money10
Risk (std)			
difference	-0.037	-0.010	0.006
p-value	(0.671)	(0.139)	(0.793)
Patience (std)			
difference	0.010	0.041	-0.025
p-value	(0.072)	(0.139)	(0.257)
Overconfidence (std)			
difference	0.013	0.067	-0.025
p-value	(0.004)	(0.139)	(0.329)
CRT score (std)			
difference	0.018	0.018	0.029
p-value	(0.506)	(0.139)	(0.327)
Altruism (frac. donated)			
difference	0.078	0.027	-0.090
p-value	(0.713)	(0.139)	(0.259)
Trust (std)			
difference	0.008	-0.006	-0.017
p-value	(0.852)	(0.139)	(0.626)
Reciprocity (std)			
difference	0.024	-0.004	-0.007
p-value	(0.914)	(0.139)	(0.823)

Note: The table presents results from regressions of participation in at least one experiment on student characteristics and interaction terms of the student characteristics with recruitment treatment dummies. For example, row 1 is based on a regression that contains willingness to take risk as well as interaction terms between willingness to take risk and the four recruitment treatment indicators. The reported values are the differences in selection effects across recruitment treatments. For example, row 1, column 1, contains the interaction effect between the *GreaterGoodMoney* treatment and willingess to take risk, minus the interaction effect between the *GreaterGood* treatment and willingness to take risk. The cell also contains the corresponding p-value of a test for difference between the two interaction effects. Each row is based on a single regression. All regressions control for the cohort of the experiment and for CRT.

A.2 Cognitive Reflection Test

The cognitive reflection test (Frederick 2005) is a widely used measure to assess an individual's ability to suppress an intuitive and spontaneous wrong answer in favor of a reflective and deliberative right answer. It consists of three questions. To minimize the probability that subjects can copy their answer from their neighbor (or obtain the correct answer from participants of earlier sessions), we created six question sets consisting of questions that are very similar or identical to the three questions by Frederick (2005).

1. Bat-and-ball-type questions

    - 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? (Set 1)
    - A stock and a stock-option cost $110 in total. The stock costs $100 more than the stock-option. How much does the stock-option cost? (Set 2)
    - A motorist and his car weigh 1100kg in total. The motorist weighs 1000kg less than the car. How much does the motorist weigh? (Set 3)
    - A cyclist and his cycle weigh together 120kg. The cycle weighs 100kg less than the cyclist. How much does the cycle weigh? (Set 4)
    - A broom and a dustpan weigh 1.1 kg in total. The broom weighs 1 kg more than the dustpan. How much does the dustpan weigh? (Set 5)
    - A bottle of wine and a corkscrew cost together 60 CHF. The bottle of wine costs 50 CHF more than the corkscrew. How much does the bottle of wine cost? (Set 6)

2. Machine-type questions

    - If it takes 10 concrete mixers 10 minutes to mix 10 tons of concrete, how long would it take 100 concrete mixers to mix 100 tons of concrete? (Set 1)
    - If it takes 5 bulldozers 5 minutes to level 5 m^2, how long would it take 10 bulldozers to level 10 m^2? (Set 2)
    - If it takes 10 workers 10 minutes to make 10 widgets, how long would it take 50 workers to make 50 widgets? (Set 3)
    - If it takes 5 printers 5 minutes to print 5 posters, how long would it take 100 printers to make 100 posters? (Set 4)
    - If it takes 10 people 10 minutes to make 10 widgets, how long would it take 100 people to make 100 widgets? (Set 5)

- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Set 6)

3. Lily-pad-type questions

- On a corn field vermin are spreading. Every day the affected area doubles in size. If it takes 32 days until the whole field is affected, how long would it take until half of the field is affected? (Set 1, see Frederick, 2005)
- In a lake, there are algae. Every day, the affected area doubles in size. If it takes 100 days until the whole lake is affected by algae, how long would it take until half the lake is affected? (Set 2)
- On a wheat field vermin are spreading. Every day the affected area doubles in size. If it takes 60 days until the whole wheat field is affected, how long would it take until half of the field is affected? (Set 3)
- On a lake an oil film is spreading. Every day, the area doubles in size. If it takes 24 days for the oil film to cover the entire lake, how long would it take for the oil film to cover half of the lake? (Set 4)
- 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? (Set 5)
- On a meadow there are primrose. Every year the area where primrose are growing doubles in size. If it takes 10 years for the primrose to cover the entire meadow, how long would it take for the primrose to cover half of the meadow? (Set 6)

A.3 Recruitment e-mail

The recruitment e-mail contained the following text (translation from German to English by the authors):

> Dear Students
>
> At the University of St.Gallen we regularly conduct economic experiments. You have now the opportunity to register yourself for participation without commitment.
> *By participating you earn money.*
>
> On top of that you gain exciting insights on an innovative research area. To date more than 2,500 students participated and 98 percent indicated that they would like to take part in further experiments in the future. There is no special knowledge necessary. By clicking on the link below you register without commitment. Once registered you will receive invitations for individual experiments.
>
> Thank you for your interest!
>
> With kind regards,
> Prof. Dr. Christian Thöni
> Behavioral and Experimental Economics

The text in italics had four different versions, which were randomized across students:

1. *GreaterGood*: *"These studies provide us with valuable scientific insights."*
2. *Money*: *"By participating you earn money."*
3. *Money&GreaterGood*: *"These studies provide us with valuable scientific insights. By participating you earn money."*
4. *Money10*: *"By participating you earn money (at least 10 CHF)."*