Journal of Behavioral and Experimental Economics 98 (2022) 101871

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Selection into experiments: New evidence on the role of preferences, cognition, and recruitment protocols☆

Petra Thiemann[a,e,*], Jonathan Schulz[b,f], Uwe Sunde[c,e,g,h], Christian Thöni[d,f]

[a] Lund University, Department of Economics, Sweden
[b] George Mason University, Department of Economics, United States
[c] University of Munich (LMU), Department of Economics, Germany
[d] University of Lausanne, Centre de droit public, Switzerland
[e] Institute for the Study of Labor (IZA)
[f] Centre for Decision Research and Experimental Economics (CeDEx)
[g] Centre for Economic Policy Research (CEPR)
[h] CESifo

ARTICLE INFO

JEL classification:
C93
D64
H41
L3

Keywords:
Classroom experiment
Selection
Recruitment
Preferences
Cognitive ability

ABSTRACT

We study selection into lab experiments based on data from two cohorts of first-year university students. We combine two experiments: a classroom experiment in which we elicit measures of time and risk preferences, overconfidence, trust, reciprocity, altruism, and cognitive reflection and a recruitment experiment with four treatment conditions that randomly vary the information provided in the e-mail sent to recruit lab participants. We find that students with higher cognitive skills are more likely to participate in experiments. By contrast, we find little evidence of selection along time and risk preferences, overconfidence, trust, and reciprocity, and our evidence of selection along altruism is inconclusive. In terms of recruitment conditions, mentioning financial incentives boosts the participation rate in lab experiments by 50 percent. Although recruitment conditions affect participation rates, they do not alter the composition of the participant sample in terms of the elicited characteristics. Finally, students who repeatedly participate in lab experiments are more patient than those who participate only once.

1. Introduction

Studying human behavior in laboratory experiments has become one of the predominant empirical methods in modern economics and the social sciences in general because lab experiments allow researchers to draw (causal) inferences based on controlled variation (Falk & Heckman, 2009). One major concern about lab experiments, however, is that subjects typically self-select into participation, meaning that the measures of interest (e.g., preference parameters) may be biased (Andersen, Harrison, Lau, & Rutström, 2010; von Gaudecker, van Soest, & Wengström, 2012) and the observed behaviors might not be representative of the population in question. A growing body of research is therefore studying the influence of preferences on selection into lab experiments, particularly risk preferences, trust, and reciprocity. Yet, many aspects of the selection of subjects into experiment participation remain poorly understood.

This study focuses on the roles of cognitive skills, overconfidence, time and risk preferences, social preferences (altruism, reciprocity), and trust for selection into lab experiments as well as on the role of the information in recruitment notifications. Whereas some of these characteristics have been studied extensively in the context of selection into lab experiments, others such as overconfidence and cognitive skills have received less attention. The effects of cognitive skills and overconfidence on selection into lab experiments are *a priori* unclear—one may expect positive, negative, or no effects. For instance, students with higher cognitive ability might be more likely to participate in lab experiments than their fellow students because their costs of completing cognitively demanding experimental tasks are relatively low or because they have

☆ A previous version of this paper was circulated under the title “Selection into Experiments: Evidence from a Population of Students” (Schulz, Sunde, Thiemann, and Thöni 2019).

* Corresponding author at: Lund University, Department of Economics, Scheelevägen 15B, 22363 Lund, Sweden.
E-mail addresses: petra.thiemann@nek.lu.se (P. Thiemann), jonathan.schulz77@gmail.com (J. Schulz), uwe.sunde@econ.lmu.de (U. Sunde), christian.thoeni@unil.ch (C. Thöni).

https://doi.org/10.1016/j.socec.2022.101871
Received 24 May 2021; Received in revised form 12 February 2022; Accepted 17 March 2022
Available online 24 March 2022

more free time than other students since they can master university coursework more easily. Conversely, they might spend more time on coursework and student jobs than those with lower cognitive ability and therefore might be less likely to participate in lab experiments. Overconfident individuals may be particularly likely to select into experiments based on their inflated belief about their own ability and earnings potential. However, they might be less likely to participate in experiments if they believe that they can easily earn money elsewhere. Because the relative magnitudes of such effects are difficult to predict, empirical evidence is needed to understand potential selection effects along both overconfidence and cognitive skills. Finally, little systematic evidence exists on the role of the information in recruitment notifications and little is known about how this information interacts with different characteristics; we discuss the related literature below.

Our analysis combines two controlled experiments that contribute to a better understanding of selection into experiments. We ran a classroom experiment to garner the characteristics of two cohorts of first-year students during students' second week at a public Swiss university that offers specializations in business, economics, finance, law, and international affairs. The characteristics we collected include overconfidence, time and risk preferences, trust, reciprocity, altruism, and cognitive reflection. In addition, we experimentally varied the content of recruitment e-mails, which we sent to the same population during their first week at university. In these e-mails, we asked the students to register in a database of experimental subjects. The goal of our research is to comprehensively study the factors that influence selection into lab experiments.

Our study yields three key results. First, we find sizable and robust evidence of selection into experiments along cognitive skills. By contrast, we find no selection effects along risk and time preferences, trust, reciprocity, or overconfidence. Our evidence on altruism is inconclusive: altruistic individuals are less likely to participate in experiments, but the effect is imprecisely measured. Second, although the recruitment treatments affect participation rates, we find no strong evidence that different recruitment conditions attract different types of students to participate in lab experiments. Third, we find relevant selection effects into repeat participation. Students who participate repeatedly are more patient and somewhat more overconfident than students who participate only once, which suggests that selection effects may drive behavioral differences between experienced and inexperienced participants.

Overall, our results provide good news for practitioners in the field of experimental economics and for readers of the experimental literature. We find no selection effects along most of the elicited measures. If anything, experimental economists can expect a sample of participants that have higher cognitive reflection scores than the overall (student) population. This implies that participants may reflect more on experimental tasks, understand those tasks better, and thus make fewer mistakes than the general population, which raises the validity and accuracy of the experimental results. In addition, the recruitment modalities do not matter much for selection: Mentioning financial incentives—as most experimenters already do—is crucial to recruiting a large number of students and does not lead to biased samples of participants. Including additional messages does not seem to impact the effect of recruitment e-mails. As a note of caution, re-inviting participants to lab experiments can impact the composition of student samples because patient students are more likely to return to the lab.

2. Contribution to the literature

This study contributes to the literature on selection into lab experiments in three ways.

Selection based on student characteristics. The approach we use to test for selection effects follows the seminal contributions by Cleave, Nikiforakis, & Slonim (2013) and Slonim, Wang, Garbarino, & Merret (2013), who investigate the role of preferences and demographic variables for participation in lab experiments among first-year students who attended an introductory tutorial class in microeconomics. These studies find differences between participants and non-participants along trust, altruism, and socio-demographic characteristics, but no differences along risk preferences and reciprocity. Other studies have investigated students' selection into lab experiments using different designs. For instance, Snowberg & Yaariv (2021) run an online elicitation of behavioral characteristics among undergraduates and find that the average lab participant is more risk averse, more willing to lie, and less generous than the undergraduate population, but they find no differences in other behaviors. Falk, Meier, & Zehnder (2014) exploit a donation decision that was mandatory for all university entrants and find no selection effects along generosity.[1]

Adding to this strand of the literature, we consider a large number of characteristics (risk, trust, reciprocity, altruism, patience, overconfidence, and cognitive ability). Two of these—overconfidence and cognitive ability—have received comparably little attention in this context, although they are frequently elicited in behavioral studies. The measure of cognitive ability we use (the cognitive reflection test, or CRT) is a key measure in the judgment and decision-making literature. It predicts patience, risk tolerance, and cognitive biases (e.g., Toplak, West, & Stanovich, 2011; Brañas Garza, Kujal, & Lenkei, 2019). Similarly, overconfidence has been used to explain important economic phenomena such as excess entry into markets, career choice, and value destroying mergers (e.g., Santos-Pinto & de la Rosa, 2020). The associated experimental literature is extensive.[2]

Recruitment experiments. Second, in terms of the design of our recruitment experiment, our study complements the studies by Harrison, Lau, & Rutström (2009), who vary the show-up fee; Krawczyk (2011), who emphasizes either pecuniary or non-pecuniary benefits in different recruitment e-mails; Lazear, Malmendier, & Weber (2012), who study sorting into experimental conditions in the context of sharing; and Abeler & Nosenzo (2015), who randomly allocate students to three recruitment e-mails that mention either monetary rewards, or appeal to the importance of helping research, or both. These studies suggest that emphasizing pecuniary benefits leads more people to register in a recruitment database but that different recruitment conditions do not attract students with different social preferences, risk preferences, or cognitive skills than the student population. An exception is the study by Harrison et al. (2009), who randomize the amount of the show-up fee across recruitment e-mails and find that increasing the show-up fee attracts relatively more risk-averse students.[3]

Our design allows us to comprehensively study the interaction effects between recruitment messages and individual characteristics. Concretely, our setup is similar to that of Abeler & Nosenzo (2015), but differs in some key aspects. Whereas the wording of three out of our four recruitment e-mails is close to their recruitment e-mails, we add a fourth type of e-mail, which allows us to test whether risk-averse students are motivated to participate if they can count on a safe reward. Moreover, whereas Abeler & Nosenzo (2015) focus on the effect of treatment variations on registrations in a recruitment database, our design allows us to investigate actual participation in experiments as an additional outcome. This is important since typically not all students who register in a database subsequently participate in experiments. Finally, we

[1] Broadly, our study also relates to work that compares the composition of students who select into experiment participation with the composition of broader (non-student) populations (see Anderson et al., 2013; Exadaktylos, Espin, & Brañas-Garza, 2013, Falk, Meier, and Zehnder, 2014; Arechar, Gächter, & Molleman, 2018; Snowberg & Yaariv, 2021).

[2] For instance, a joint search for the terms "overconfidence" and "experiment" in Google Scholar delivers more than 57,000 results.

[3] See also Harrison, Lau, & Yoo (2020) who show that selection (e.g., due to varying show-up fees or distance to the experiment location) leads to biased estimates of risk preferences.

explore the interaction between our recruitment variations and a large number of characteristics, including patience and overconfidence, in an explorative analysis.

Repeat participation. Third, our study relates to the small strand of the literature that investigates repeated participation in lab experiments. Matthey & Regner (2013) find a negative correlation between generosity in allocation decisions and the number of previous rounds of participation in other experimental sessions. In a similar vein, Benndorf, Moellers, & Normann (2017) find that experienced subjects show less trustworthiness and trust than inexperienced subjects.

These differences between experienced and inexperienced subjects raise the question of whether this is (i) due to sample selection (i.e., only certain subjects show up repeatedly) or (ii) a response to experiences in the lab. Using the data from our study, we can shed some light on the first channel.[4] Since we collected all measures before any lab experiments took place, the measures are not confounded by prior experiences in the lab. This is a new design feature compared with previous studies of repeat participation and allows us to investigate whether there are *ex-ante* differences between subjects who participate in experimental sessions only once and subjects who participate repeatedly. There are no clear hypotheses on selection into repeat participation in the literature and our analysis is therefore exploratory.

3. Experimental design and procedures

3.1. Setup

The data we use for this study provides information on two cohorts (2011 and 2012) of first-year students at a public Swiss university, the University of St. Gallen. In addition to background characteristics, the data includes information on seven characteristics (time preferences, risk preferences, overconfidence, trust, reciprocity, altruism, and cognitive reflection) that we elicited in a classroom experiment.

We combine these measures with the results of a randomized recruitment experiment. In this experiment, we invited first-year students to register in a recruitment database for lab experiments. We randomized four types of invitation e-mails that 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.

3.2. Design of the classroom experiment

Our pen-and-paper classroom experiment was carried out in the second week of the semester during the last 20 min of the students' first Introductory Economics tutorial. Tutorials are groups that meet every other week and review the course material together with an instructor. The Introductory Economics course is compulsory for all first-year students at the University of St. Gallen.[5] Nearly all the students of a cohort thus participated in the first tutorial sessions (here, a cohort is defined as all the students who enter the first year of their undergraduate degree in the fall of 2011 and 2012, respectively). The classroom experiment was not pre-announced, that is, the students only found out about it at the end of the tutorial session.

In each cohort, the experiment was conducted in 38 tutorial groups (76 groups in total). All the tutorials took place on the same day with the exception of two tutorials, which took place three days later. The average tutorial group size was 23 students and varied across tutorials (standard deviation 8.4).[6] Trained research assistants instructed the students verbally and supervised the experiment. Moreover, we collected identifiers for each tutorial to account for the possible dependence of answers within tutorials in our empirical analysis.

The behavioral tasks of the classroom experiment were linked to financial incentives. The participants were informed that, once they had handed in their sheets, we would draw one participant per tutorial to receive the experimental earnings of one randomly selected task. The randomly selected subjects received CHF 94 (CHF 1 equaled $1 at the time of our study) in cash on average. Thus, the stakes were high for those selected for payout. The probability of being paid was on average 4.3 percent per group and the average expected payoff of a participant amounted to CHF 4. Since tutorial group sizes, and thus expected payoffs, varied, we control for tutorial group size throughout the analysis.[7]

During the classroom experiment, we collected six measures using incentivized tasks: risk preferences, trust and reciprocity (using a trust game), time preferences, overconfidence, and altruism. In addition, we collected a measure of cognitive ability, using a version of Frederick (2005)'s CRT. We elicited five of the measures in both cohorts (risk preferences, time preferences, overconfidence, altruism, CRT) and two of the measures only in the second cohort (trust and reciprocity). Trust and reciprocity were added into the questionnaire of the second cohort to investigate the students' social preferences more comprehensively. The measures we used to elicit these seven characteristics correspond to standard measures used in lab and field experiments at the time of implementation (see Table 1 for an overview and Section C.1 in the appendix for more details). The pen-and-paper form of the experiment implied that we could not prevent students from skipping individual questions or tasks, which means that we do not have observations for the full sample for all measurements.[8]

We varied the order in which the students completed the

Table 1
Overview of the elicited measures.

Measures	Elicitation
Risk preferences	Investment game following Gneezy & Potters (1997).
Trust and reciprocity	Trust game following Berg, Dickhaut, & McCabe (1995); added in the second cohort.
Overconfidence	Relative ability judgements based on year-guessing tasks, similar to Ewers & Zimmermann (2015) and Schulz & Thöni (2016).
Altruism	Donation decision, see Schulz, Thiemann, & Thöni (2018).
Time preferences (patience)	Choice table based on Dohmen, Falk, Huffman, & Sunde (2010).
Cognitive reflection test	Following Frederick (2005); two questions in the first and four questions in the second cohort.

Note: The table presents an overview of the measures that were elicited during the classroom experiment. See Section C.1 in the appendix for more details about the procedures.

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

[5] The university offers majors in Business Administration, Economics, International Affairs, Legal Studies, and Law & Economics.

[6] The distribution of tutorial group sizes is depicted in Figure B.1 in the appendix.

[7] Experimental evidence suggests that randomly paying only a subset of individuals leads to similar decisions in the dictator game compared with paying all subjects (Clot, Grolleau, & Ibanez, 2018). Furthermore, Charness, Gneezy, & Halladay (2016) conclude—based on a literature review—that paying only a subset of tasks or individuals is at least as effective as the "pay all" approach and could even be more effective.

[8] The fraction of missings in our measures is small: out of the seven tasks, three were completed by more than 95 percent of the students (risk, trust, altruism), another three were completed by between 90 and 95 percent of the students (reciprocity, patience, CRT), and the overconfidence task was completed by 81 percent of the students. See Figures A.2 and A.3 in the appendix for details.

incentivized tasks that measure risk preferences, time preferences, trust, reciprocity, and overconfidence across and within the tutorials (see Section C.5 in the appendix for more details). This variation served two purposes. First, to mitigate spillovers between the participants' decisions we used two task orders in each tutorial, such that students who sat next to each other had different task orders. Second, we varied the task order across tutorials in a given time slot, which permits us to control for order effects. Our results are not sensitive to the task order.

3.3. Design of the recruitment experiment

The recruitment experiment was conducted by inviting students to participate in lab experiments by e-mail. The invitation process followed two steps. First, we invited students to register in a *recruitment database* for future lab experiments. In both cohorts, this invitation was sent out to all first-year students during the first two weeks of the teaching period (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, students added their e-mail address to the recruitment database (we used the ORSEE database system; see Greiner, 2015). Second, we invited the students in the database to participate in specific lab experiments throughout the academic year.

Treatment variation. In the recruitment experiment, we randomly varied the e-mail invitation to the recruitment database.[9] We randomized the students into four equally-sized groups, one baseline (B) and three treatment groups (T1–T3). In each group, the recruitment e-mail differed by just one sentence. In the baseline, this sentence emphasized the monetary rewards (*Money*). In the treatment groups, this sentence highlighted the scientific value of the experiments (*GreaterGood*), mentioned a guaranteed minimum monetary reward of CHF 10 (*Money10*), or combined the scientific value and monetary reward argument (*GreaterGoodMoney*). The wording was as follows (translated from German):

B	*Money*	"By participating, you earn money."
T1	*GreaterGood*	"These studies provide us with valuable scientific insights."
T2	*Money10*	"By participating you earn money (at least CHF 10)."
T3	*GreaterGoodMoney*	"These studies provide us with valuable scientific insights. By participating, you earn money."

The students who registered in the database received e-mail invitations for up to five lab experiments throughout the academic year. These e-mails did not have any treatment variations. As our main outcome, we investigate participation in at least one lab experiment, but we also study whether the recruitment treatments change the students' willingness to register in the recruitment database as an intermediate outcome.

To uncover which pieces of information most affect student registrations and subsequent participation in lab experiments, our analysis concentrates on the three *ceteris paribus* comparisons across the treatments out of the six possible pairwise treatment comparisons. By *ceteris paribus* comparisons, we mean the effect of adding or subtracting information, rather than substituting one piece of information for another. First, we test whether mentioning financial rewards in addition to the students' contribution to research boosts database registration rates and lab experiment participation rates (*GreaterGood* vs. *GreaterGoodMoney*). Second, we investigate whether mentioning the contribution to research in addition to the financial rewards increases registration and participation rates (*Money* vs. *GreaterGoodMoney*). Finally, we test the consequences of mentioning a guaranteed minimum payoff in addition to the mere possibility of earning money (*Money* vs. *Money10*).

One potential threat to the validity of the treatment comparisons are spillover effects across students. For instance, the students may talk about the recruitment e-mail, their registration in the database, or their experiences in the lab. Such interactions may bias the treatment effects toward zero as the information spreads across the cohort. To mitigate these concerns, we conduct robustness checks in which we only consider early registrations (i.e., registrations within one week of the receipt of the recruitment e-mail), which are arguably less affected by spillovers than later registrations. Specifically, this allows us to exclude spillovers due to actual participation. We find that our results are robust when only considering early registrations.

4. Samples and descriptive statistics

4.1. Samples

In total, 2363 students received an invitation to register in the database and 2241 of them were enrolled on the first-year Introductory Economics course.[10] Out of these students, 1740 students (78 percent) participated in the classroom experiment.

We did not announce the classroom experiment in advance to avoid that show-up at the tutorial would depend on the classroom experiment. Because participation in the classroom experiment was voluntary, not all the students in a tutorial session participated. Using attendance counts from the tutorial sessions, we calculate that 90 percent of those individuals who showed up at the tutorials also participated in the classroom experiment (93 percent in the first cohort and 86 percent in the second cohort).

Using enrollment records, we also verify that the background characteristics of participants in the classroom experiment are representative for the population of first-year students in both cohorts. The age, gender, nationality, mother tongue, and region of origin of participants in the classroom experiment are not statistically different from those of all first-year students (see Table A.1 in the appendix for more details). In the following sections, the analysis focuses on the sample of participants in the classroom experiment.

4.2. Descriptive statistics

4.2.1. Registration in the database and participation in experiments

Figure 1 illustrates the students' decisions to register in the recruitment database and subsequently participate in lab experiments in the sample of participants in the classroom experiment ($N = 1{,}740$). In this sample, 23.4 percent of the students registered in the recruitment database and 11.6 percent participated in at least one lab experiment; 6.7 percent of the students participated in lab experiments once and 4.9 percent participated twice or more. Although participation in up to five experiments was possible, a negligible number of students (seven) participated in more than two experiments.[11]

4.2.2. Elicited measures

Our experimental design yields substantial variation in the behavioral measures (see Figure A.2 in the appendix). Responses in the domains of risk preferences, trust, reciprocity, and time preferences vary along the full support of possible choices. For the first- and second-mover responses in the trust game (trust and reciprocity), the responses are moderately left-skewed.

[9] Section C.3 in the appendix shows the full text of the recruitment e-mails. In the years prior to the experiment, the e-mail used to recruit students for experiments was similar to the *GreaterGoodMoney* e-mail.

[10] The reason why 122 students were not enrolled on the Introductory Economics course is unknown. The most likely reason is dropout before the end of the course enrollment period (i.e., before the beginning of the second week).

[11] In the full sample of the students who received a recruitment e-mail—including those who did not participate in the classroom experiment—these fractions were similar (see Appendix Figure A.1).

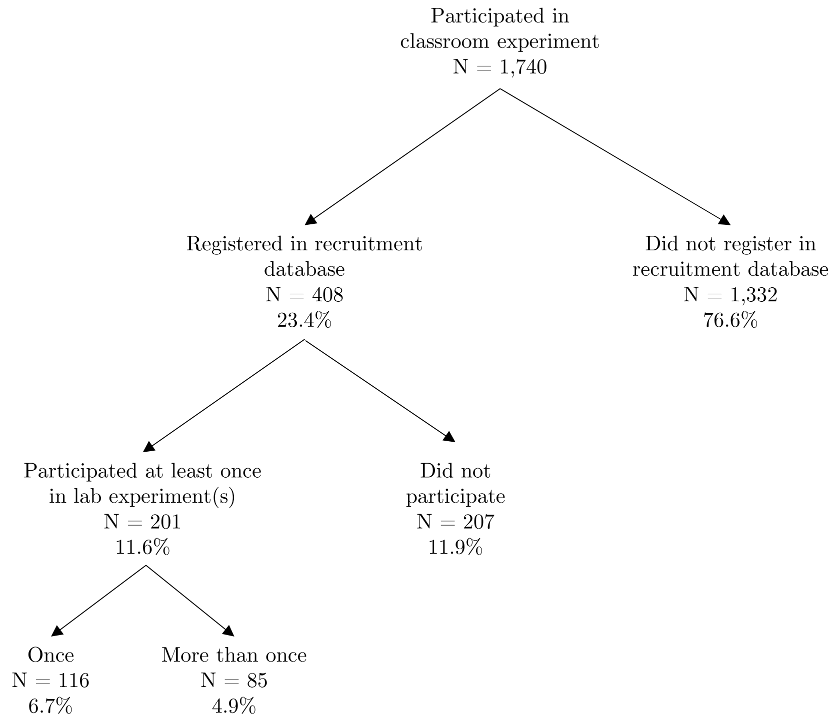

Fig. 1. Student decisions: registration in the recruitment database and participation in lab experiments. *Note:* The sample consists of all participants in the classroom experiment (N = 1,740). The percentages are rounded to one decimal place.

The overconfidence measure is surprisingly symmetrically distributed; the mean is slightly positive but not significantly different from zero (p-value of t-test = 0.111; all the p-values reported in this article are based on two-sided tests). There is also some variation in the level of altruism: about one-quarter of students are willing to donate part of their experimental earnings to charity, and the contributions range from 0 to 100 percent.[12] Finally, we find substantial variation in the number of correct answers in the CRT (see Figure A.3 in the appendix). Since cognitive ability may confound selection effects, we control for the CRT throughout the analysis.

5. Results

5.1. Selection into experiments along the elicited measures

We evaluate selection into lab experiments along the elicited measures. We study two outcomes: registration in the recruitment database (our intermediate outcome) and participation in at least one lab experiment (our main outcome; both coded as binary variables).

Mean comparisons. We start by comparing the averages of the students' characteristics among those who registered in the database and those who did not (see Figure A.4 in the appendix). We find a highly statistically significant and large difference in cognitive skills: registered students have on average a 0.24 standard deviation higher CRT score than unregistered students ($p < 0.001$, Wilcoxon rank sum test). They are also more risk tolerant, more patient, and less altruistic than unregistered students; all these differences are statistically significant at the 5 percent level (see the p-values in Figure A.4 in the appendix). Registered students are also more trusting, but the difference is only weakly significant (p = 0.096, Wilcoxon rank sum test). We find no statistically significant differences for reciprocity and overconfidence.

Next, we study whether the selection effects persist when we focus on participation in lab experiments instead of registration in the database. Figure 2 displays the raw differences in the elicited measures between the participants in at least one lab experiment and the non-participants. The latter group consists both of students who registered in the experimental database, but did not participate in any lab experiment and of unregistered students (see Fig. 1).[13]

Across the samples of participants and non-participants, we find highly significant and large differences in cognitive reflection: the level of CRT is 0.24 standard deviations higher among participants than non-participants (p = 0.003, Wilcoxon rank sum test). By contrast, we find no significant raw differences between participants and non-participants in any of the other elicited measures (see the p-values reported in Fig. 2).

Regression results and robustness. To explore the robustness of these unconditional results, we regress the outcomes on the elicited measures controlling for cohort, the order of the preference elicitation, the size of the tutorial groups, the recruitment conditions, and gender (Table 2). Further, we interact the CRT with a cohort dummy to account for the difference in the number of CRT questions across cohorts (two questions in the first cohort and four questions in the second cohort).[14] To interpret the regression coefficients, we standardize all the elicited measures to have a mean of zero and a standard deviation of one,[15] with the exception of altruism, which we express as the fraction of experimental

[12] Donating to charity serves as measure of altruism according to Falk et al. (2018). Our setting is comparable to a standard dictator game (see Engel, 2011, for a survey). The donating rates in our tasks are lower than those in the literature. This may be related to the size of the perceived stake or ambiguity in earnings because the students did not know their actual earnings before they decided whether to donate to charity and at which proportion (e.g., Brañas-Garza, Kovarik, & Lopez, 2020).

[13] The group also includes students who signed up to participate in lab experiments but did not show up, but this number was small (12 students).

[14] All the results are based on linear probability models estimated using OLS. Our results are similar when we use logit models instead (results not shown).

[15] To standardize the measures, we use the z-score formula $z_i = \frac{x_i - \mu_X}{\sigma_X}$, where z_i is the standardized value (z-score) for person i, x_i denotes the original value, μ_X denotes the within-cohort mean of the original value, and σ_X is the within-cohort standard deviation of the original value. A cohort consists of all undergraduate students who enter their first year in the fall of that academic year.

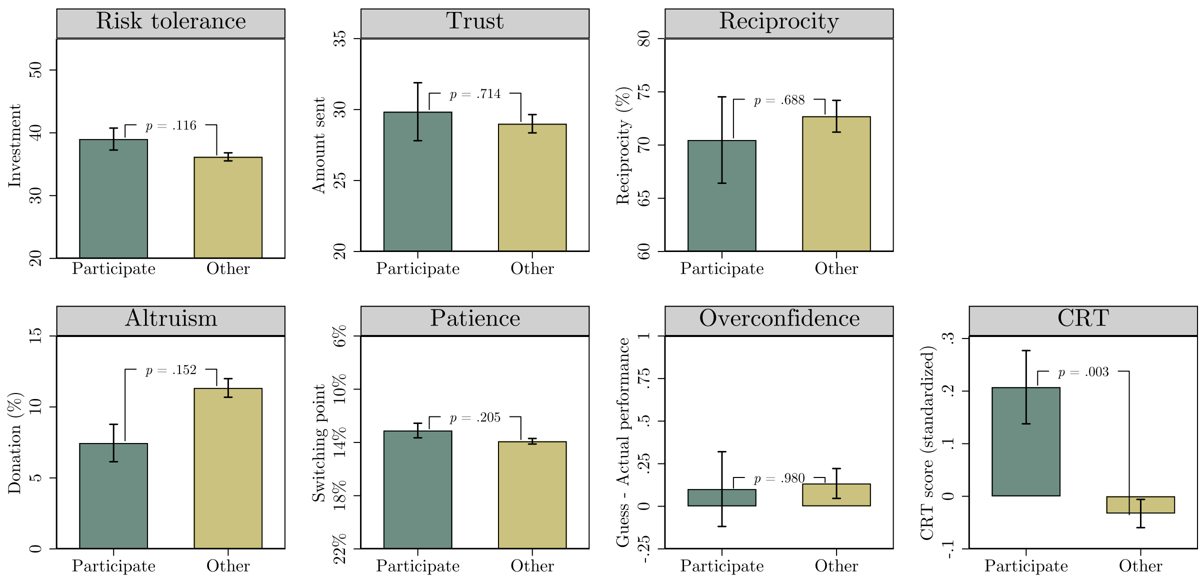

Fig. 2. Selection into lab experiments. *Note:* The figure compares the characteristics of participants in at least one lab experiment (left bar in each panel) with the characteristics of students who never participated (right bar in each panel). The whiskers indicate standard errors and the *p*-values are from Wilcoxon rank sum tests. Reciprocity is reported as the percentage of the payoff-equalizing transfer sent back by the second mover in the trust game. Donations are reported as the percentage of experimental earnings donated to charity. The sample consists of the participants in the classroom experiment (*N* = 1,740).

earnings donated to charity. We run separate regressions for each measure to maximize the sample size.

The CRT has a statistically significantly positive and sizable effect both on registrations in the database and on participation rates in lab experiments (Table 2). A one-standard deviation increase in the CRT is associated with an increase in the registration rate by 2.9 percentage points in the first cohort and by 6.7 percentage points in the second cohort; these numbers correspond to increases of 12 and 27 percent compared with the average registration rate (Panel A, column 7). The CRT also affects participation in lab experiments. A one-standard deviation increase in the CRT is associated with an increase in the participation rate of about 2.7 percentage points in both cohorts, which corresponds to an increase of 23 percent over the average participation rate (Panel B, column 7). This effect is robust across the specifications in which CRT is used as a control variable (Table 2, columns 1–6).

We also observe a significantly negative effect of altruism (fraction donated to charity) on registration rates, but only a weakly significant negative effect on participation rates in lab experiments. An increase in the fraction of earnings donated to charity of 50 percentage points (e.g., from no donation to donating half of experimental earnings) corresponds to a decrease in the registration rate of 6.1 percentage points, a decrease of 25 percent compared with the average registration rate. This effect partly translates into lower participation; however, the effect on participation is only weakly significant. An increase in the fraction of earnings donated to charity of 50 percentage points corresponds to a decrease in the participation rate of 2.4 percentage points, a decrease of 20 percent compared with the average participation rate.

None of the other measures show any (strongly) statistically significant effects for either registration or participation rates (Table 2, columns 1–3 and 5–6). First, we find no evidence of an influence of risk preferences, trust, or reciprocity. These results largely corroborate earlier findings (Falk, Meier, & Zehnder, 2014; Cleave, Nikoforakis, and Slonim, 2013). Moreover, we find no statistically significant effect of overconfidence on registration or participation rates, only a weakly significant effect of patience on registration rates, and no statistically significant effect of patience on participation rates. With respect to participation rates, the effects of risk preferences, trust, reciprocity, patience, and overconfidence are not only insignificant but small (absolute values are ≤ 1 percentage point; Panel B, columns 1–3 and 5–6).

We conducted several further robustness checks (see Tables A.2 and A.3 in the appendix). First, we vary the control variables included in the model. We start with a regression that does not include controls for gender and the CRT (column 1) and sequentially add the gender dummy (column 2) and the CRT (column 3; this column corresponds to our preferred specification, as displayed in Table 2). Next, we split the sample by cohort (columns 4 and 5). Moreover, we assess whether our results might be blurred due to spillover effects across students (column 6). We only consider students as “registered” and “participants” if they registered early in the recruitment database (i.e., within one week of receiving the recruitment e-mail). These students are arguably less affected by spillover effects that arise as students share information and lab experiences.[16] Finally, we run specifications in which we jointly include all the elicited measures (columns 7 and 8).

The results of these robustness checks are in line with those from our preferred specification. The effect of the CRT is robust across the specifications. It is, however, smaller in the first cohort (column 4) and when we consider students as “registered” and “participants” if they registered early (column 6); the latter change is mechanical—at least to some extent. The effect of altruism (fraction donated) on registrations in the database is strong and statistically significant in all but the most restrictive specification (column 8), whereas its effect on participation is only significant in some specifications and imprecisely measured when we add control variables or drop observations from the sample. Similarly, we find (weakly) significant effects of patience on registration rates in the database in four out of eight specifications, but a weakly significant effect on participation in only one. Finally, we find no selection effects for any of the other variables in any of the specifications.

Summing up, our results suggest that the willingness to take risks, trust, reciprocity, patience, and overconfidence of lab experiment participants are very similar to those of non-participants. The evidence on altruism is inconclusive: participants are less altruistic than non-

[16] See Alexeev & Freer (2019) who study spillover effects in selection into lab experiments in a network model.

Table 2
Registration in the recruitment database and participation in at least one lab experiment.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A.	Dependent variable: registered in recruitment database (binary)						
Risk	0.012						
	(0.012)						
Trust		0.014					
		(0.017)					
Reciprocity			0.014				
			(0.016)				
Altruism				-0.121***			
				(0.033)			
Patience					0.018*(0.010)		
Overconfidence						0.015	
						(0.013)	
CRT	0.064***	0.072***	0.073***	0.068***	0.070***	0.067***	0.067***
	(0.016)	(0.016)	(0.018)	(0.016)	(0.016)	(0.019)	(0.016)
CRT × first cohort	-0.036*			-0.044**	-0.039*	-0.045*	-0.038*
	(0.020)			(0.021)	(0.021)	(0.023)	(0.020)
Female	0.025	0.085**	0.078**	0.014	0.019	0.018	0.018
	(0.024)	(0.032)	(0.035)	(0.023)	(0.024)	(0.025)	(0.023)
Frac. registered	0.25	0.24	0.25	0.24	0.25	0.25	0.24
R-squared	0.031	0.063	0.062	0.036	0.030	0.029	0.030
Panel B.	Dependent variable: participated in at least one lab experiment (binary)						
Risk	0.005						
	(0.008)						
Trust		-0.002					
		(0.011)					
Reciprocity			-0.010				
			(0.010)				
Altruism				-0.048*			
				(0.026)			
Patience					0.008		
					(0.007)		
Overconfidence						0.004	
						(0.011)	
CRT	0.026**	0.024*	0.024*	0.026**	0.028**	0.032**	0.027**
	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.014)	(0.012)
CRT × first cohort	-0.005			-0.007	-0.003	-0.013	-0.005
	(0.016)			(0.016)	(0.016)	(0.018)	(0.016)
Female	-0.004	-0.017	-0.021	-0.010	-0.006	0.007	-0.006
	(0.017)	(0.022)	(0.024)	(0.016)	(0.017)	(0.019)	(0.016)
Frac. participated	0.12	0.09	0.10	0.12	0.12	0.13	0.12
R-squared	0.022	0.027	0.028	0.023	0.021	0.024	0.022
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	1,578	646	617	1,570	1,507	1,316	1,590

Note: The table presents OLS estimates of regressions of registration in the recruitment database (Panel A) and participation in at least one lab experiment (Panel B, both are binary outcomes) on individual characteristics. The characteristics are standardized to have a mean of zero and a standard deviation of one, with the exception of altruism, which is measured as the fraction of experimental earnings donated to charity. The controls are: recruitment treatments, university cohort, the order of the tasks in the classroom experiment, and tutorial group size. Robust standard errors, clustered at the tutorial level, in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

participants, but the effect is imprecisely measured. As a key difference, we find that participants have higher CRT scores than non-participants. This finding suggests that participants take more time to reflect on a question or task than the average student. Thus, one may expect these individuals to be more diligent and careful when they perform tasks in lab experiments. This likely raises the accuracy of the results of lab experiments and implies that lab results on cognitive biases provide a lower bound of the cognitive biases in the population from which participants are drawn.

5.2. Recruitment experiment

Mean comparisons. Fig. 3 shows the registration rates in the recruitment database and participation rates in at least one lab experiment among all participants in the classroom experiment. We report the results of the three *ceteris paribus* comparisons outlined in Section 3.3 (*Money vs. GreaterGoodMoney; Money vs. Money10; GreaterGood vs. GreaterGoodMoney*).

Focusing on registration rates we find that out of the three pairwise comparisons, only the difference between the *GreaterGoodMoney* and the *GreaterGood* treatment is significant. The registration rates are 8 percentage points (42 percent) higher in the *GreaterGoodMoney* treatment, compared to the *GreaterGood* treatment ($p = 0.004$, Wilcoxon rank sum test). We do not find any statistically significant differences between the *Money* and *Money10* conditions or the *Money* and *GreaterGoodMoney* conditions.

Participation rates are 5 percentage points (56 percent) higher in the *GreaterGoodMoney* treatment, compared to the *GreaterGood* treatment ($p = 0.037$, Wilcoxon rank sum test), and we do not find any statistically significant differences across the other two treatment pairs (*Money* vs. *Money10, Money* vs. *GreaterGoodMoney*). In sum, these findings suggest that mentioning a safe amount or benefits to society in addition to financial rewards neither changes registration nor participation rates in a statistically significant way. Omitting information on financial rewards, however, has a large and significantly negative effect on both registration and participation rates.

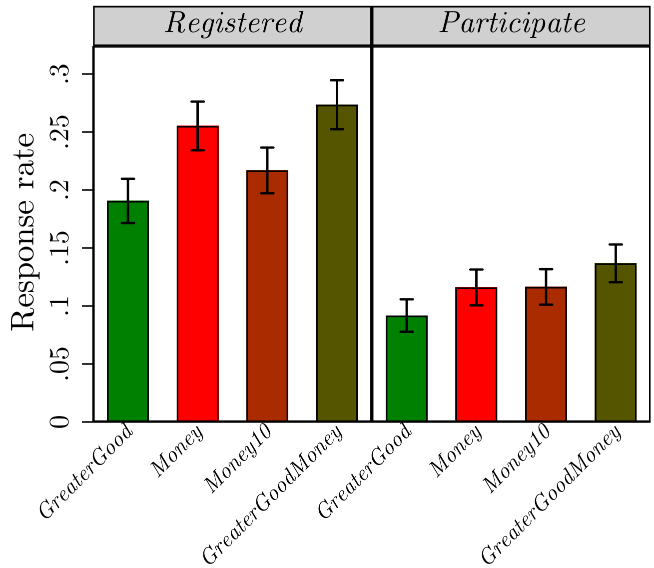

Fig. 3. Result of the recruitment experiment. *Notes:* The left panel shows the fraction of students who registered in the recruitment database in each treatment condition. The right panel shows the fraction of participants in at least one lab experiment in each treatment condition. The sample consists of all participants in the classroom experiment ($N = 1{,}740$). The whiskers indicate standard errors.

Regression results and robustness. We conduct regression analyses to confirm the robustness of our results (Table A.4 in the appendix). We start by regressing the outcomes (registration, participation) on the recruitment treatments while controlling for cohort. We then present two robustness checks. First, we address concerns about treatment spillovers across students (even columns of Table A.4): we classify students as "registered" only if they registered in the database within a week of receiving the recruitment e-mail; similarly, we recode participation as zero if a student registered in the database more than one week after receiving the recruitment e-mail. This robustness check aims at mitigating concerns about spillovers that arise between those students who have already participated in a lab experiment and those who did not register in the database early. Second, we conduct the analyses in the larger sample of all students who received a recruitment e-mail (columns 5–8).

The regression results confirm our findings from the raw mean comparisons (Table A.4, columns 1 and 3). Mentioning financial incentives in addition to benefits to society raises the registration rate by 8.3 percentage points (column 1; *GreaterGoodMoney* vs. *GreaterGood*), an increase of 44 percent relative to the registration rate in the *GreaterGood* treatment. This difference is statistically significant ($p = 0.004$). Similarly, mentioning financial incentives in addition to benefits to society increases the participation rate by 4.6 percent (column 3; *GreaterGoodMoney* vs. *GreaterGood*), an increase of 51 percent relative to the participation rate in the *GreaterGood* treatment. By contrast, we find no statistically significant effect of mentioning a safe amount of CHF 10 compared to only mentioning the possibility of earning money. If anything, the registration and participation rates are lower in the *Money10* treatment than in the *Money* treatment, but the effect is not statistically significant at conventional levels in our main specifications (columns 1 and 3; *Money10* vs. *Money*). Finally, we find no statistically significant effect of mentioning benefits to society in addition to monetary rewards (columns 1 and 3; *GreaterGoodMoney* vs. *Money*).[17]

The effects are largely robust to choosing alternative specifications. First, the effects are similar but more precisely measured in the larger sample of all students who received a recruitment e-mail (columns 5–8). Second, we find a high level of consistency between the robustness check based on early registrations and our main specifications. As a final check of the validity of our selection estimates, we confirm in a regression analysis that participation rates in the classroom experiment are balanced across treatment conditions (Table A.5).

In sum, our findings corroborate the prior evidence by Abeler & Nosenzo (2015), who study the effect of mentioning monetary incentives on registration rates in an experimental database. We extend their findings with our results on participation rates in lab experiments. We find that adding information on monetary rewards into a recruitment e-mail increases participation. We conclude that financial incentives are important to motivate students to participate in lab experiments.

5.3. Treatment-specific selection effects

In this section, we combine the results of the two experiments and investigate whether the different recruitment treatments attract different types of students to participate in lab experiments. Our main outcome variable is a binary variable that indicates whether a student has participated in at least one lab experiment. The analysis is carried out for the sample of participants in the classroom experiment (recall that elicited measures are missing for non-participants).

The number of treatment-specific selection effects that can be tested in our data is potentially large: seven characteristics and four treatments yield up to 42 treatment comparisons. To structure the analysis, we therefore start by testing directed hypotheses suggested in the literature (e.g., Abeler & Nosenzo, 2015; Krawczyk, 2011). We then present an explorative analysis of all possible treatment-specific selection effects.

Tests of directed hypotheses. Our setup allows us to test two sets of hypotheses that are based on the experimental literature: (1) Does mentioning a certain show-up fee in addition to financial incentives attract relatively more risk-averse individuals to participate in lab experiments? (2) Does mentioning financial incentives in addition to benefits to society increase the participation rate of selfish individuals? Relatedly, does mentioning financial incentives, but not mentioning benefits to society, discourage altruistic individuals from participating?

To answer the first question, we investigate whether the *Money10* treatment attracts more risk averse students than the *Money* treatment or all the other treatments. Table 3 shows that selection into participation along the willingness to take risk is unaffected by mentioning a certain show-up fee. This also holds when investigating registrations in the database as an outcome (see Table A.6 in the appendix).

To answer the second question, 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. While the level of altruism among the participants differs slightly across the recruitment treatments, none of the treatment comparisons is statistically significant at conventional levels (Table 4). We also find no statistically significant interaction effects when we investigate registrations in the recruitment database (see Table A.7 in the appendix).

These results are in line with those of Abeler & Nosenzo (2015), who do not detect any differences in the effect of pro-sociality on registrations between recruitment treatments that emphasize the social value of participation and those that do not. We confirm that these results hold

[17] From a practical standpoint, one may also compare the *GreaterGood* and *Money* treatments. We find significant differences in registration rates across the treatments, but no significant differences in participation rates in lab experiments.

Table 3
Interaction effects: mentioning a safe amount and selection based on willingness to take risk. Dependent variable: participated in at least one lab experiment (binary).

	(1)	(2)	(3)	(4)
	participated in at least one lab experiment			
Comparison:	Money 10 vs. Money		Money 10 vs. all other treatments	
Risk	-0.013	-0.003	0.010	0.015**
	(0.012)	(0.011)	(0.007)	(0.007)
Risk × Money10	0.023	0.010	0.001	-0.008
	(0.020)	(0.017)	(0.018)	(0.015)
Controls	✓	✓	✓	✓
Early registrations		✓		✓
Fraction participated	0.12	0.09	0.12	0.09
Observations	857	857	1723	1723
R-squared	0.028	0.028	0.013	0.016

Note: The table shows the results of OLS regressions with participation in at least one lab experiment (binary) as the dependent variable. The sample in columns 1 and 2 consist of only individuals in the *Money*10 and *Money* treatments, and the sample in columns 3 and 4 consists of individuals in all the treatments. The controls are: a dummy for the *Money*10 treatment, a cohort dummy, the order of the experimental tasks, and tutorial group size. The even columns code participation as 1 if the student registered in the recruitment database within one week of receiving the recruitment e-mail and subsequently participated in at least one lab experiment. Robust standard errors, clustered at the tutorial level, in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

not only for registrations in the recruitment database but also for participation in lab experiments. Our findings, however, contrast with the evidence provided by Krawczyk (2011), who finds that individuals recruited by emphasizing monetary (as opposed to non-monetary) benefits are less likely to display pro-social behavior (i.e., participate in a non-paid survey) and less altruistic in general.

In sum, our evidence suggests that variations in the recruitment treatments do not systematically alter the participant sample: mentioning financial incentives, benefits to society, or a certain show-up fee does not change the average willingness to take risk and altruism among the participants.

Explorative analysis of treatment-specific selection effects. The directed hypotheses above might miss some important treatment-specific selection effects that have not previously been investigated. This section thus explores a larger set of treatment-specific selection effects.

We start with a non-parametric analysis based on the Kruskal–Wallis test. This test explains whether a measure (e.g., the CRT) is balanced across the treatment conditions in a sample. For instance, a low p-value from the Kruskal–Wallis test in the sample of lab participants would indicate that the CRT is unbalanced across treatments among lab participants, thus suggesting a treatment-specific selection effect. With just two treatments, one may perform balancing tests using the Wilcoxon rank sum test. However, since our experiment includes four treatment variations, we use the Kruskal–Wallis test, which extends the rank sum test to a multi-treatment setting.

Table A.8 in the appendix displays the results of the Kruskal–Wallis test for all the measures in three samples: the full sample of participants in the classroom experiment (column 1), the sample of students who registered in the database (column 2), and the sample of participants in at least one lab experiment (column 3). Because we randomized the students into treatments, the characteristics are balanced across treatments in the full sample of participants in the classroom experiment. Moreover, we find no treatment-specific selection effects with respect to database registrations: all the p-values of the Kruskal–Wallis test are far from conventional levels of significance. Finally, we find no treatment-specific selection effects in the sample of lab participants, with the exception of overconfidence (p of the Kruskal–Wallis test = 0.033).

We further explore the treatment-specific selection effects in regression analyses in which we interact treatment dummies with individual characteristics. We regress the outcomes (registration, participation) on treatment dummies, individual characteristics, and all interaction terms between treatment dummies and individual characteristics. Our results are in line with the results of the Kruskal–Wallis test. We find that the interaction terms are not jointly significant in these regressions (Table A.9 in the appendix). In line with our earlier result, however, the overconfidence among the participants of lab experiments differs across treatment conditions. Overconfident participants are more likely to select into participation in the *GreaterGood* and *GreaterGoodMoney* treatments than in the *Money* and *Money10* treatments (Table A.9, column 3). This effect disappears when we restrict the sample to the second cohort (Table A.9, column 4).

In sum, this explorative analysis does not reveal strong evidence of treatment-specific selection effects. If anything, we find some evidence of treatment-specific selection along overconfidence. Further research is therefore needed to understand the selection of overconfident individuals into lab experiments.

Table 4
Interaction effects: mentioning financial incentives and selection based on altruism (fraction donated). Dependent variable: participation in at least one lab experiment (binary).

	(1)	(2)	(3)	(4)	(5)	(6)
	participated in at least one lab experiment					
Comparison:	GreaterGoodMoney vs. GreaterGood		GreaterGoodMoney vs. Money		GreaterGood vs. Money	
Altruism	-0.089**	-0.051	-0.079**	-0.071**	-0.083*	-0.049
	(0.042)	(0.038)	(0.034)	(0.029)	(0.044)	(0.040)
Altruism × GreaterGoodMoney	0.024	0.015	0.010	0.033		
	(0.072)	(0.067)	(0.063)	(0.055)		
Altruism × GreaterGood					0.001	-0.021
					(0.059)	(0.053)
Controls	✓	✓	✓	✓	✓	✓
Early registrations		✓		✓		✓
Fraction participated	0.13	0.10	0.12	0.09	0.11	0.09
Observations	859	859	850	850	835	835
R-squared	0.021	0.021	0.019	0.012	0.020	0.016

Note: The table shows results of OLS regressions with participation in at least one lab experiment (binary) as the dependent variable. The sample in columns 1 and 2 consists of the individuals in the *GreaterGoodMoney* and *GreaterGood* treatments, the sample in columns 3 and 4 consists of the individuals in the *GreaterGoodMoney* and *Money* treatments, and the sample in columns 5 and 6 consists of the individuals in the *GreaterGood* and *Money* treatments. The controls are: recruitment treatments, a cohort dummy, the order of the experimental tasks, and tutorial group size. The even columns code participation as 1 if the student registered in the recruitment database within one week of receiving the recruitment e-mail and subsequently participated in at least one lab experiment. Robust standard errors, clustered at the tutorial level, in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

5.4. One-timers and repeat participants

For those students who registered in the database, we can not only track whether they participated in lab experiments, but also check whether they participated repeatedly. In this section, we analyze whether repeat participants differ systematically from one-timers. Recall that the behavioral measures were collected before any lab experiments took place, which allows us to rule out that the results are influenced by laboratory experiences.

Mean comparisons. In our analysis, we restrict the sample to the 201 participants in lab experiments. Among these, 58 percent participated only once, almost all of the remaining subjects participated twice, and a very small number participated three to five times. We find significant differences between one-timers and repeat participants for two measures: repeated participation is associated with significantly higher levels of patience and overconfidence (see Figure A.5 in the appendix). For the trust game, we find substantial but insignificant differences, which suggest that repeat participants are more trusting but less reciprocating than one-timers.

Regression results and robustness. We test the robustness of our results in different regression specifications. Notably, controlling for gender might be important, given prior evidence that male subjects are more likely to participate repeatedly (Guillen & Veszteg, 2012).

Table 5 displays the results of the OLS regressions of repeat participation on individual characteristics and the standard set of controls, including gender and the CRT. The analysis corroborates Guillen & Veszteg (2012)'s finding that women are less likely to participate repeatedly, but the gender effect is imprecisely measured. Moreover, the positive relationship between patience and repeat participation is robust to including controls and the effect is sizable: a one-standard deviation increase in patience increases the probability of participating repeatedly by 9.9 percentage points, which corresponds to an increase of 23 percent compared to the average probability to participate repeatedly. The result on overconfidence is also sizable but only weakly significant in the regression analysis: a one-standard deviation increase in overconfidence maps to an increase in repeat participation of 7.7 percentage points, an increase of 18 percent over the average probability to participate repeatedly. None of the other coefficients of the elicited measures are significant.

Furthermore, we conduct a number of robustness checks in which we vary the control variables, investigate the results for each cohort separately, and restrict the sample to those students who registered in the database within the first week of receiving the recruitment e-mail (see columns 1–6 of Table A.10 in the appendix). We also test specifications in which we include all the elicited measures jointly (see columns 7 and 8 in Table A.10 in the appendix). The results for both patience and overconfidence are largely robust across specifications, but we find larger and more precise results for both measures in the first cohort (column 4) and in the sample of students who registered in the database within the first week (column 6). The results for all the other measures are insignificant across all the robustness checks, with the exception of trust, which is weakly significant when we include all measures together (column 8).

To sum up, we find sizable and significant selection effects into repeat participation along patience. We view these results as indicative given the low number of repeat participants and the small sample size, which leads to noisily estimated coefficients of some of the characteristics. Further research is needed to address these concerns and analyze repeat participation more in depth.

6. Discussion

This study provides novel evidence of selection into lab experiments based on an extensive battery of behavioral measures, the influence of the information provided in the recruitment process on participation rates in lab experiments, and the interplay between these two aspects. The analysis is based on classroom experiments conducted among two cohorts of first-year university students to elicit their preferences across various domains, cognitive ability, and overconfidence. We combined the elicited measures with a recruitment experiment that varied the information in e-mail invitations to participate in lab experiments.

Our study provides good news for experimental economists.

Table 5
One-timers and repeat participants.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: participated more than once (binary)						
Risk	-0.045						
	(0.032)						
Trust		0.047					
		(0.079)					
Reciprocity			-0.075				
			(0.069)				
Altruism				-0.043			
				(0.179)			
Patience					0.099**		
					(0.041)		
Overconfidence						0.077*	
						(0.046)	
CRT	0.038	0.041	0.045	0.038	-0.023	0.085	0.037
	(0.061)	(0.059)	(0.058)	(0.061)	(0.057)	(0.096)	(0.060)
CRT × first cohort	-0.046			-0.047	-0.003	-0.112	-0.050
	(0.080)			(0.079)	(0.078)	(0.106)	(0.078)
Female	-0.148*	-0.134	-0.143	-0.126	-0.132*	-0.128	-0.137
	(0.083)	(0.122)	(0.121)	(0.082)	(0.078)	(0.088)	(0.083)
Controls	✓	✓	✓	✓	✓	✓	✓
Frac. participated repeatedly	0.44	0.51	0.50	0.44	0.44	0.44	0.44
Observations	194	61	62	192	186	172	194
R-squared	0.099	0.180	0.183	0.088	0.140	0.142	0.092

Note: The table presents OLS estimates of regressions of participation in more than one lab experiment (binary) on individual characteristics. The regressions are conducted in the sample of participants of at least one lab experiment. The characteristics are standardized to have a mean of zero and a standard deviation of one, with the exception of altruism, which is measured as the fraction of experimental earnings donated to charity. The controls are: recruitment treatments, university cohort, the order of the tasks in the classroom experiment, and tutorial group size. Robust standard errors, clustered at the tutorial level, in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Informing potential participants about financial rewards is critical to recruiting a large number of subjects but does not change the composition of the participant pool. The only significant selection effect we find is along cognitive skills: participants in lab experiments exhibit higher cognitive reflection scores, which may contribute to the accuracy of experimental results. In addition, patient individuals are more likely to participate repeatedly. Overall, these findings suggest that selection into lab experiments along student characteristics and the role of recruitment procedures for this selection might be less of a concern than suggested previously, at least as far as student populations are concerned.

Several questions remain for further research. First, we study a population that is often recruited for laboratory experiments—students from business- and economics-related fields—but the results may vary across other fields of study. Further research could aim to explore such differences systematically. Second, we cannot quantify the importance of having a representative sample for ensuring the external validity of experimental studies. Many experimenters are not primarily interested in studying a representative population *per se*, but instead care about correctly measured treatment comparisons. We leave it to further research to quantify the impact of selection on the external validity of treatment comparisons. Finally, in our setup, the number of times that students could participate was low due to the limited number of lab experiments that were ran at the time. Further research could explore the selection into repeat participation in student samples that are invited to a larger number of experiments.

Declaration of Competing Interest

None.

Acknowledgements

We thank the editor and two anonymous referees for their instructive comments and suggestions. We thank seminar participants at the University of Zurich and Lund University, Dorothea Kübler, and Roel van Veldhuizen 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.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.socec.2022.101871.

References

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

Alexeev, A., & Freer, M. (2019). Selection in the lab: A network approach. *Working Paper*.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2010). Preference heterogeneity in experiments: Comparing the field and laboratory. *Journal of Economic Behavior and Organization, 73*(2), 209–224.

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

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

Benndorf, V., Moellers, C., & Normann, H.-T. (2017). Experienced vs. inexperienced participants in the lab: Do they behave differently? *Journal of the Economic Science Association, 3*(1), 12–25. https://doi.org/10.1007/s40881-017-0036-z

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*(1), 122–142. https://doi.org/10.1006/game.1995.1027http://www.sciencedirect.com/science/article/pii/S0899825685710275

Brañas-Garza, P., Kovarik, J., & Lopez, M. C. (2020). No moral wiggles in € 5 and € 1000 dictator games under ambiguity. *Manuscript*.

Brosig, J., Riechmann, T., & Weimann, J. (2007). Selfish in the end? An investigation of consistency and stability of individual behavior. *MPRA Paper No. 2035*.

Charness, G., Gneezy, U., & Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization, 131*, 141–150.

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

Clot, S., Grolleau, G., & Ibanez, L. (2018). Shall we pay all? an experimental test of random incentivized systems. *Journal of Behavioral and Experimental Economics, 73*, 93–98. https://doi.org/10.1016/j.socec.2018.01.004http://www.sciencedirect.com/science/article/pii/S2214804318300363

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610.

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

Exadaktylos, F., Espin, A. M., & Brañas-Garza, P. (2013). Experimental subjects are not different. *Nature: Scientific Report, 1213*. https://doi.org/10.1038/srep01213

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics, 133*(4), 1645–1692.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535–538.

Falk, A., Meier, S., & Zehnder, C. (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.

Brañas Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics, 82*(August). https://doi.org/10.1016/j.socec.2019.101455

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

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *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., & Veszteg, R. F. (2012). On "lab rats". *Journal of Socio-Economics, 41*(5), 714–720. https://doi.org/10.1016/j.socec.2012.07.002

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

Harrison, G. W., Lau, M. I., & Yoo, H. I. (2020). Risk attitudes, sample selection, and attrition in a longitudinal field experiment. *Review of Economics and Statistics, 102*(3), 552–568. https://doi.org/10.1162/rest_a_00845

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

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics, 4*(1), 136–163.

Matthey, A., & Regner, T. (2013). On the independence of history: Experience spill-overs between experiments. *Theory and Decision, 75*(3), 403–419. https://doi.org/10.1007/s11238-012-9346-z

Santos-Pinto, L., & de la Rosa, L. E. (2020). Overconfidence in labor market. In K. Zimmermann (Ed.), *Handbook of labor, human resources and population economics*. Springer.

Schulz, J., Sunde, U., Thiemann, P., & Thöni, C. (2019). Selection into experiments: Evidence from a population of students. *Technical Report*. CeDEx Discussion Paper No. 2019-09.

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

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

Slonim, R., Wang, C., Garbarino, E., & Merret, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior & Organization, 90*, 43–70.

Snowberg, E., & Yaariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review, 111*(2), 687–719. https://doi.org/10.1257/AER.20181065

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition, 39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization, 81*(2), 664–676. https://doi.org/10.1016/j.jebo.2011.10.006