DO LAB EXPERIMENTS MISREPRESENT SOCIAL PREFERENCES? THE CASE OF SELF-SELECTED STUDENT SAMPLES

Armin Falk
University of Bonn

Stephan Meier
Columbia University

Christian Zehnder
University of Lausanne

Abstract
Social preference research has received considerable attention among economists in recent years. However, the empirical foundation of social preferences is largely based on laboratory experiments with self-selected students as participants. This is potentially problematic as students participating in experiments may behave systematically different than nonparticipating students or nonstudents. In this paper we empirically investigate whether laboratory experiments with student samples misrepresent the importance of social preferences. Our first study shows that students who exhibit stronger prosocial inclinations in an unrelated field donation are not more likely to participate in experiments. This suggests that self-selection of more prosocial students into experiments is not a major issue. Our second study compares the behavior of students and participants recruited from the general population in a trust experiment. In general, we find very similar behavioral patterns for the two groups, but nonstudents make significantly more generous repayments suggesting that results from student samples might be seen as a lower bound for the importance of prosocial behavior. (JEL: C90, D03)

1. Introduction

Social preferences such as concerns for distributional fairness and reciprocity have received considerable attention in recent economic research (see, e.g., Cooper and Kagel, forthcoming). The empirical foundation of social preferences is largely based on laboratory experiments using self-selected students as samples. This is a potential problem, as students participating in experiments might behave systematically different than nonparticipating students or nonstudents. If participating students behave more or less prosocially than the population of interest, our laboratory results provide a biased estimation of the potential of social preferences for the analysis of economic

The editor in charge of this paper was Stefano DellaVigna.

Acknowledgments: We thank the editor, three anonymous referees, Simon Gächter, Lorenz Götte, Bruce Kogut, Rafael Lalive, and John List for helpful comments.

E-mail: armin.falk@uni-bonn.de (Falk); sm3087@columbia.edu (Meier); christian.zehnder@unil.ch (Zehnder)

Journal of the European Economic Association August 2013 11(4):839–852
 DOI: 10.1111/jeea.12019

outcomes. Were this to be the case we would need to be more careful in plugging behavioral assumptions derived from observations in the lab into models used to derive implications for the general population.

In this paper we provide empirical evidence on whether laboratory experiments with student samples systematically misrepresent social preferences. In particular, we address two potential problems.

First, experiments rely on volunteers, creating a problem of self-selection. This may bias outcomes in experiments if participants exhibit a stronger or weaker prosocial inclination than people who do not participate. A priori, the direction of a potential selection effect is unclear. If people's participation decision is mainly money driven, one might expect an overrepresentation of self-interested payoff-maximizers in the participant pool. However, it is also possible that social motives determine people's decision to participate (e.g., people may want to help the researcher or foster the advancement of science), which would speak for an overrepresentation of prosocially inclined participants.[1] While a drastic overestimation of prosocial motives would be especially troubling for the literature on social preferences, it is, of course, also important to know whether there is a bias in the other direction.

Second, most laboratory experiments are conducted with university undergraduates. While using students as subjects is very convenient, they are not representative of the general population in many dimensions. The important question for our context is whether they also differ with respect to social preferences, so that using them as participants distorts the measurements of social preferences in experiments.

Our first study analyzes whether participating students are more prosocial than nonparticipating students. The ideal data set to test for potential differences between participants and nonparticipants would provide information on prosocial preferences of *all* students while observing who participates in experiments and who does not. This type of data is usually not available simply because we have proxies for preferences typically only for participants in experiments. Moreover, if we know preferences from nonexperimental data—for example, survey studies—we do not observe decisions to participate in an experiment. In our first study we present results using a novel data set that combines preference measures for both participants and nonparticipants. In particular, we use a naturally occurring donation decision as a measure of participants' and nonparticipants' prosocial inclination. Our results show that students with stronger prosocial inclinations are neither more likely to participate in experiments (extensive margin), nor do they participate more often (intensive margin). These findings resonate with a complementary study by Cleave, Nikiforakis, and Slonim (forthcoming) who also don't find a selection bias regarding social preferences. However, while Cleave et al. make use of tutorials of introductory microeconomics to obtain a laboratory measure of social preferences for about 600 students, we identify prosocial inclinations using a naturally occurring field donation that gives us access to data for more than

1. Levitt and List (2007) and List (2009) focus on the latter possibility when they argue that behavior in the lab might not be a good indicator of behavior in the field.

16,000 students. While both approaches have their advantages, the fact that both studies ultimately emphasize a nonresult makes a large number of observations relevant, because it increases the precision of the estimation and reduces the possibility to find a null effect by chance. In fact, we show that our sample allows us to estimate the null result with a small confidence interval.

Our second study uses a version of the trust game (Berg, Dickhaut, and McCabe, 1995) to investigate whether measurements of social preferences change if the usual student subjects are replaced with participants from the general population. In contrast to many existing studies, we use the same recruitment procedure, the same instructions, the same decision process, and the same financial incentives for both our subject pools. Our results reveal no significant difference in first mover trusting behavior between students and nonstudents. However, the repayment level is significantly lower for students than for nonstudents. Our results are in line with earlier studies that also show that prosocial behavior is even more frequently observed with nonstudent participants (see, e.g., Fehr and List, 2004; Bellemare and Kröger, 2007; Dohmen et al., 2008; Burks, Carpenter, and Götte, 2009; Belot, Duch, and Miller, 2010).

Our paper contributes to a recent methodological debate about the role of experimental economics in the social sciences (see, e.g., Levitt and List, 2007; Falk and Heckman, 2009; List, 2009; Croson and Gächter, 2010; Bardsley et al., 2010; Henrich, Heine, and Norenzayan, 2010; Gächter, 2010). Some of this work has raised serious concerns about the relevance of lab findings with regard to the role of social preferences. This paper provides a step in empirically investigating one issue raised in this debate. Our results suggest that using self-selected student samples does not contribute to a systematic overestimation of social preferences. On the contrary, the results of our second study indicates that results obtained from student samples might be seen as a lower bound for the importance of prosocial behavior. Of course, our results do not exclude that laboratory experiments may provide distorted estimates of social preferences for other reasons (such as low stakes, short durations, high degrees of scrutiny). However, we see our paper as a starting point and hope that future research will investigate the empirical relevance of other potential sources for biases.

The paper is organized as follows. Section 2 contains the field study on selection of students into experiments. The question of whether students and nonstudents have different prosocial inclinations is discussed in Section 3. Section 4 concludes.

2. Do Social Preferences Predict Self-Selection?

2.1. Research Design

This section analyzes whether self-selection of students into experiments leads to a misrepresentation of prosocial preferences in the participating part of the student population. We study decisions of students to participate in experiments organized by the experimental economics laboratory of the University of Zurich. Our sample consists of 16,666 undergraduates who registered at the University of Zurich between the fall

term 1998 and the spring term 2004 and for whom registration at the University of Zurich is their first enrollment at a university. For all those students, we know whether and how often they participated in an economics experiment between the fall term 1998 and the fall term 2005. In total 1,783 students participated at least once—that is, the participation rate is about 11%. Conditional on participating at least once, the students participate in 2.5 experiments on average.

To measure the extent of all students' prosocial inclinations we use a naturally occurring prosocial decision at the University of Zurich as a proxy. Each semester, every student has to decide whether or not he or she wants to contribute a predetermined amount to two social funds which provide charitable services (financial support for foreign students (CHF 5) and free loans for needy students (CHF 7), where CHF 1 $\simeq$ \$1.03 (May 2013); for further details see Frey and Meier (2004a, b). Students can therefore give CHF 0, 5, 7, or 12 (both funds together). The level of possible donations is thus very similar to stake sizes typically used in lab settings.

There are several features why these donation decisions constitute an interesting proxy for social preferences. First, the measure does not rely on self-reported survey responses but on actual decisions. Second, donation decisions are made in private and never made public.[2] Third, students are unaware that their behavior is analyzed in a research study. Fourth, and most importantly, all students at the university have to decide about the donations. Thus, our measure is not subject to any selection issue.

However, as with most field measures there are also potential problems. Since the persuasive power of our results critically depends on the quality of our measurement of social preferences, it is important to discuss in detail the different measures we use and how they address potential caveats. Our first measure (*First Field Donation*) only considers a student's donation decision when he or she first registers for a program. This measure has several advantages. First, the university rules require that each student has to show up in person at the registration office for the initial enrollment. This ensures that we know with certainty that this first donation decision has been made by the student him- or herself. Second, as the initial enrollment takes place before the first semester starts, this measure is collected before students have taken any courses at the university, before they have been exposed to any lab recruitment efforts, and before they have participated in any experiment. We can therefore rule out the possibility of reversed causality as participation in experiments cannot have influenced the decision to contribute to the funds. These features make this measurement a particularly clean one.

Our second measure (*Average Field Donation*) exploits information on all donation decisions taken by a student. For each individual, we calculate the average donation amount over all observed contributions. Using several measures per individual has the advantage of reduced measurement error. A potential problem is that the forms for registration renewals can be completed at home. Therefore, we cannot be sure that it is the student him- or herself who fills out the form. However, because students also

2. As researchers we got access to the data through the university administration under the condition that we immediately anonymize the data after matching it with the experimental data base.

have to provide details regarding major and minor study subjects on the same form, it is quite unlikely that another person can perform this task. To further increase the confidence that the variable *Average Field Donation* measures an individual's prosocial inclination we use data collected by Benz and Meier (2008). They perform a modified dictator game in the laboratory using a subsample of the students in our data set as participants. It turns out that individuals with higher average field donations transfer a significantly higher share of their endowment to the recipient (Spearman's Rho $= 0.29$, $p < 0.0001$). This provides direct evidence that our field measure captures the same social motivations as the simple experiments typically used in the laboratory. Finally, it is also reassuring to notice that our two measures *First Field Donation* and *Average Field Donation* are strongly correlated (Spearman's Rho $= 0.73$, $p < 0.001$).[3]

2.2. Results

Panel A in Table 1 reveals that participants differ in various dimensions from nonparticipants. These differences indicate the relevance of self-selection of particular groups of students. In Panel B of Table 1 we investigate whether this selection is also associated with differences in prosocial inclinations. The panel provides descriptive statistics of contributions to the two funds for participants and nonparticipants. The summary statistic does not show any significant difference between participants and nonparticipants. In their first decision, the same proportion of participants and of nonparticipants contributed to at least one of the two funds (75%) and, on average, they donate about the same amount (CHF 8.39 versus 8.45; $p = 0.67$ in a t-test). Figure 1 illustrates that both the participation rate and the number of experiments a student participated in does not significantly depend on individuals' first donation decisions. None of the differences are statistically significant. When we look at all decisions of a student, it turns out that participants contribute on average in 77% of all decisions, while nonparticipants' contribution rate is 76% (n.s.). There is also no substantial difference in the average amount donated (CHF 8.66 versus 8.84; $p = 0.09$ in a t-test; see also the distribution of average donations in Online Appendix Figure A.1). Thus, the raw data analysis does not reveal any significant difference in prosocial inclinations of participants and nonparticipants.

Panel A of Table 2 reports probit estimations, where the dependent variable is an indicator variable for the decision to participate in experiments and the independent variable is either the first donation (columns (1), (2), and (3)) or the average field donation (columns (4), (5), and (6)).[4] We report marginal effects in brackets. Column (1) shows that students who contribute more money in their first decision are not significantly more likely to participate in an experiment than those who don't. The marginal effect is essentially zero. As a consequence of the large number of observations, our effects are quite precisely estimated. The 95% confidence interval

3. As robustness checks we also add estimations relying on a measure that counts how often individuals have contributed to at least one of the two funds. This *Individual contribution rate* correlates highly with the *Average Field Donation* (Spearman's Rho $= 0.92$, $p < 0.001$).

4. For results using contribution to at least one fund, see Online Appendix Table A.2.

TABLE 1. Summary statistics of study 1.

Variable	Nonparticipants Mean	Nonparticipants s.d.	Participants Mean	Participants s.d.	t-test/ χ^2-test[a]
Panel A: Observable characteristics					
Age at registration	21.94	4.21	21.07	2.87	$p < 0.01$
No. of semesters	5.34	3.26	5.97	3.15	$p < 0.01$
Gender (Women=1)	0.57	0.50	0.53	0.50	$p < 0.01$
Nationality (Foreigner=1)	0.08	0.27	0.07	0.25	$p < 0.05$
Computer science	0.04	0.18	0.03	0.16	$p = 0.21$
Economics & Business	0.13	0.32	0.14	0.34	$p < 0.05$
Theology	0.01	0.08	0.003	0.05	$p < 0.05$
Law	0.16	0.36	0.25	0.42	$p < 0.01$
Medicine	0.07	0.26	0.18	0.38	$p < 0.01$
Veterinary medicine	0.03	0.16	0.03	0.16	$p = 0.64$
Arts faculty	0.47	0.49	0.33	0.46	$p < 0.01$
Natural science	0.10	0.30	0.05	0.21	$p < 0.01$
Panel B: Prosocial behavior					
Contributed in first decision (=1)	0.75	0.43	0.75	0.43	$p = 0.80$
First Field Donation	8.39	5.18	8.45	5.16	$p = 0.67$
Individual contribution rate	0.76	0.34	0.77	0.33	$p = 0.20$
Average Field Donation	8.66	4.15	8.84	4.05	$p = 0.09$
No. of observations	14,884		1,783		

Notes: The table presents summary statistics for people who never participated in an experiment and people who participated in an experiment at least once. Panel A reports observable characteristics including the age of the person at registration, the number of semesters for which we observe donations, the individual's gender, the foreigner status, and the individual's field of study. Panel B summarizes our measures for prosocial behavior. "Contributed in first decision" is unity if the individual contributed to at least one of the two charitable funds in his very first decision and zero otherwise. "First field donation" is the amount donated in the very first decision. "Individual contribution rate" is the fraction of all possible decisions in which the individual contributed to at least one of the two funds. "Average field donation" is the average amount that the individual donated in all his decisions.

a. χ^2-tests for categorical variables and t-tests otherwise.

of the marginal effect is [–0.1, 0.1] (in percentage points). This implies that a change in the magnitude of one standard deviation in the first donation decision (s.d. = 5.2) is very unlikely to increase (decrease) the participation rate by more than 0.6 (0.4) percentage points (i.e., an increase (decrease) of 5.6% (3.7%) relative to the average participation rate of 10.7%).

Column (4) reports a regression using the *Average Field Donation* as a proxy for prosocial inclinations. This proxy is potentially influenced by students' experience at the University including their participation in experimental studies. The results are very similar to the ones obtained from using only the first decision: Individuals who contribute on average more to the charitable funds are not significantly more likely to participate in experiments. The marginal effects indicate that the participation rate of students who contribute on average one CHF more is only about 0.1 percentage points higher. This means that for an increase in the average field donation of one standard deviation (s.d. = 4.1), the participation rate increases by only 0.4 percentage points

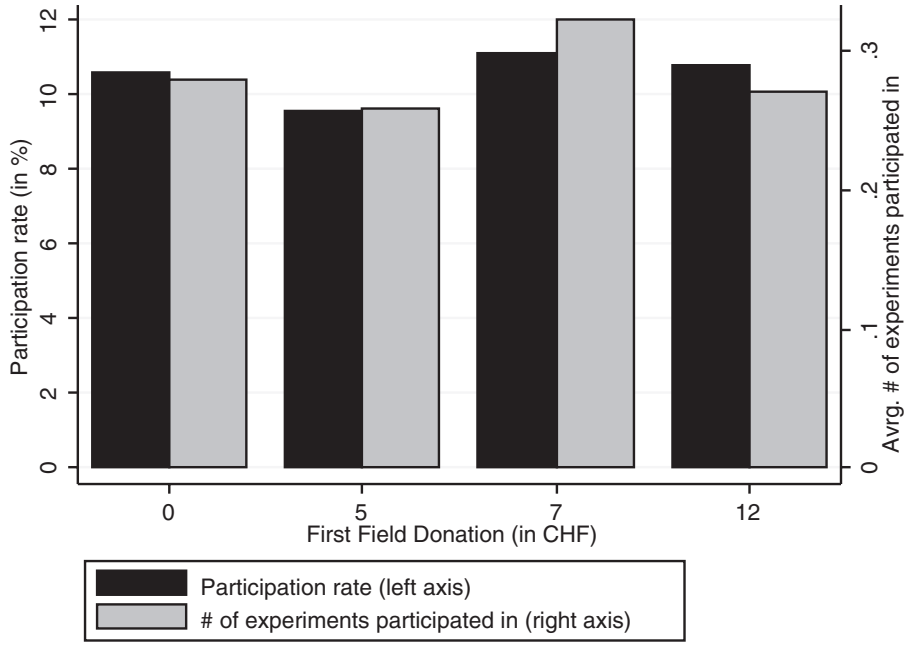

FIGURE 1. First field donation and participation in experiments. The figure shows the participation rate (left axis) and the average number of experiments a student participated in (including students who did never participate, right axis) depending on the first field donation in study 1. Distribution of First Field Donation: 25.20% contribute CHF 0, 4.19% contribute CHF 5, 5.68% contribute CHF 7, and 64.93% contribute CHF 12.

(i.e., an increase of 3.7% relative to the average participation rate of 10.7%). Given the large number of observations the lack of a significant effect is a strong result. The 95% confidence interval of the marginal effect is [–0.02, 0.2] (in percentage points) indicating that it is extremely unlikely that changing the average field donation by one s.d. increases (decreases) the participation rates by more than 0.9 (0.08) percentage points.[5]

In addition to participating for a first time, it is also interesting to investigate if social preferences predict whether a student becomes a regular participant.[6] Column (7) and (8) show Tobit regressions with the number of experiments an individual participated in as dependent variable. The estimations show that both the "First Field Donation" and the "Average Field Donation" are not good predictors for how often somebody participates in experiments (this holds both overall and conditional on participating, see Online Appendix Table A.3 for additional specifications).

5. Cleave, Nikiforakis, and Slonim (forthcoming) use second-mover back transfer in percent of the tripled first-mover investment in a trust game as their measure of social preferences. On average second movers return about 25%. They find that a one percentage point increase in the percentage returned decreases the participation rate by 0.09 percentage points. This is insignificant and the 95% confidence interval is [–0.25, 0.06] (in percentage points).

6. We thank John List for pointing out this second margin of interest to us.

TABLE 2. Participating in experiments depending on prosocial behavior in the field.

Dependent variable	Participating at least once in a laboratory experiment						No. of experiments participated in	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: General selection								
First Field Donation	0.001	0.002	0.002				0.002	
	(0.003)	(0.003)	(0.003)				(0.013)	
	[0.000]	[0.000]	[0.000]					
Avrg. Field Donation				0.005	0.005	0.007*		0.022
				(0.003)	(0.003)	(0.003)		(0.016)
				[0.001]	[0.001]	[0.001]		
Demographic controls	No	Yes	Yes	No	Yes	Yes	No	No
Field of study controls	No	No	Yes	No	No	Yes	No	No
Constant	− 1.252**	− 0.995**	− 1.272**	− 1.289**	− 1.018**	− 1.313**	− 6.347**	− 6.526**
	(0.025)	(0.141)	(0.140)	(0.031)	(0.141)	(0.141)	(0.179)	(0.202)
No. of observations	16,666	16,666	16,666	16,666	16,666	16,666	16,666	16,666
Pseudo R squared	0.00	0.023	0.050	0.000	0.024	0.050	0.000	0.000
Mean of DV	0.107	0.107	0.107	0.107	0.107	0.107	0.275	0.275
Panel B: Subgroups								
	Female	Male	Econ	Non-Econ	Law	Arts		
Avrg. Field Donation	0.002	0.009	0.008	0.005	0.006	0.015**		
	(0.004)	(0.005)	(0.008)	(0.003)	(0.007)	(0.006)		
	[0.000]	[0.002]	[0.002]	[0.001]	[0.001]	[0.002]		
Constant	− 1.294**	− 1.280**	− 1.238**	− 1.301**	− 1.066**	− 1.566**		
	(0.040)	(0.048)	(0.076)	(0.034)	(0.062)	(0.056)		
No. of observations	9,425	7,242	2,104	14,563	2,841	7,506		
Pseudo R squared	0.000	0.001	0.001	0.000	0.000	0.002		
Mean of DV	0.101	0.115	0.120	0.105	0.155	0.077		

Notes: Columns (1)–(6) report probit regressions where the dependent variable is 1 if the subject participated at least once in a laboratory experiment and 0 otherwise. Columns (7) and (8) show Tobit regression where the dependent variable is the number of experiments participated in. Standard errors in parentheses. Marginal effects in brackets. Panel A shows regressions based on the whole population. Demographic controls are "Age at registration", "Number of semesters", dummies for "Gender" and"Foreigner", and cohort dummies for the semester/year in which students registered. Field of study controls for the faculty of a person's major ("Computer Science", "Economics & Business", "Theology", "Law", "Medicine", "Veterinary medicine", "Arts", and "Natural Science"). The full regression results of the regressions are reproduced in Table A.1 in the Online Appendix. Panel B shows coefficients of regressions for different subgroups.

**Significant at 0.01.

As the main purpose of this study is to detect differences between populations (and not to explain these differences if they exist), the estimations without controls are the most important ones. The descriptive statistics reveal many significant differences between the two groups of interest (e.g., gender and major). The question that we want to answer is: do these differences also imply that there is a difference regarding social preferences between these groups? To answer this question, it is important not to include controls (because the observable heterogeneity may exactly be the reason for the difference in social preferences). Therefore, columns (1), (4), (7), and (8) contain our main results.

However, it can be of separate interest whether there is selection for certain groups. To investigate this question, we add two types of controls. In column (2) and (5) we add "demographic" variables (gender, age, foreigner status, number of semesters, cohort dummies). The results don't change. In columns (3) and (6) we additionally control for the field of study. While the marginal effect doesn't change it becomes significant at the 5% level.[7] This indicates that for certain majors, participants may select based on their field donation. Panel B of Table 2 shows separate regressions for different subgroups that might be interesting for research on prosocial behavior. The results show that the marginal effect is bigger for men than for women, but none is significant. The effects also remains insignificant if we consider economists and non-economists separately. If we estimate the effect for the field of studies that are most represented in experiments (law and arts), we find a significant effect for students from the arts faculty.

In sum, our results do not support the hypothesis that participating students have different social preferences than nonparticipants. This suggests that within the group of students the bias due to self-selection on social preferences is likely to be small. While there might be some selection within certain subgroups, these subgroups do not make up a sufficient part of a typical student sample to yield an overall significant effect. However, it is still possible that student participants behave differently than participants recruited from a more general subject pool. We investigate this question in the next section.

3. Do Students Have Different Social Preferences?

3.1. Research Design

We conduct two identical trust experiments using distinct subject pools for the recruitment of participants. Contrary to most existing studies, we use the same recruitment procedure, the same instructions, the same decision process and the same financial incentives for participants in both experiments. Therefore differences in prosocial behavior can only be caused by differences between the two subject pools.

7. See Online Appendix Table A.3 for the corresponding regressions with the number of experiments an individual participated in as dependent variable. Adding controls does not change the results.

All participants in the experiments live in Zurich. However, while one group of our participants was recruited from the student pool at the University of Zurich, the other group was recruited from a representative sample of the population of the city of Zurich (for details on the recruitment procedure of this study, see the Online Appendix).

As participation was voluntary, both our groups of participants are self-selected. In light of our first study it seems plausible to assume the absence of important selection effects with respect to social preferences, but we cannot directly rule out such a possibility with our data. However, our results are informative in any case. Even if sorting takes place our study tells us whether recruiting subjects from the general population yields a different measurement of prosocial inclinations than recruiting subjects from a student pool. This is of practical importance as the vast majority of experiments and surveys relies on voluntary participation.

To measure social preferences, we use a variant of the trust game (Berg, Dickhaut, and McCabe, 1995). Both subjects receive an endowment of CHF 20. The first mover decides how much of his endowment to transfer to the second mover. The transfer can be any amount in steps of 2 CHF—that is, 0, 2, 4, . . . , or 20 CHF. The chosen transfer is tripled by the experimenter and passed to the second mover. Contingent upon the first mover's transfer the second mover decides on a back transfer. This back transfer can be any integer amount between 0 and 80 CHF. The first mover earns his endowment minus his own transfer plus the back transfer of the second mover. The second mover gets his endowment plus three times the first mover's transfer minus the back transfer.[8]

In order to elicit second movers' willingness to reciprocate, we used the contingent response method—(see Brandts and Charness, 2011, for a discussion about the validity of the method). This means that each second mover, before knowing the actual first mover's investment, made a back transfer decision for each of the 11 possible investments (0, 2, . . . , 20) of the first mover. The advantage of the contingent response method is that it allows us to measure each second mover's willingness to reciprocate independently of the transfer which he actually received. This is important, because it enables us to make a clean comparison of the level of reciprocity, even if first movers behave differently between subject pools (for details on the procedure, see the Online Appendix).

3.2. Results

In total we have 1,296 participants in the experiment (295 recruited from the student pool, 1,001 recruited from the general population). Students and nonstudents differ in many socio-demographic dimensions. In particular, we observe that nonstudents are on average older, more likely to be married, less well educated, and more likely to be right-wingers (see Table A.4 in the Online Appendix). In this study we investigate whether students and nonstudents also exhibit different prosocial inclinations. We start by examining trusting behavior of first movers. A simple comparison of first mover

8. First movers were also asked to indicate their expectation about the back transfer of their second mover given their own transfer decision.

TABLE 3. First-mover (FM) and second-mover (SM) behavior in field trust game.

Dependent variable	FM transfer		FM belief		SM back transfer	
	(1)	(2)	(3)	(4)	(5)	(6)
Student	0.299	− 1.486	0.821	0.588	− 2.297**	− 0.118
	(0.611)	(0.797)	(0.977)	(1.467)	(0.483)	(0.904)
FM transfer			1.502**	1.445**	1.597**	1.623**
			(0.053)	(0.062)	(0.036)	(0.039)
Student x FM transfer			− 0.019	0.026	− 0.056	− 0.062
			(0.108)	(0.115)	(0.067)	(0.070)
Socio-demographic controls	No	Yes	No	Yes	No	Yes
Constant	13.17**	5.862*	− 2.675**	− 0.931	2.907**	− 6.602
	(0.287)	(2.589)	(0.452)	(3.302)	(0.285)	(3.779)
No. of observations	652	583	652	583	7,076	6,144
R-squared	0.000	0.178	0.586	0.593	0.488	0.527

Notes: The table investigates differences in first- and second-mover behavior in the trust experiment of study 2. Columns (1) and (2) report OLS estimations with average first-mover transfers as the dependent variable (robust standard errors in parentheses). Columns (3) and (4) report OLS estimations with average expected back transfers of first movers as dependent variable (robust standard errors in parentheses). Columns (5) and (6) report OLS estimations with second-mover repayments as the dependent variable (robust standard errors clustered on individual in parentheses). As repayment decisions are elicited with the contingent response method, we have eleven observations per second mover (one for each possible first mover transfer). "Student" is an indicator variable which is one if the individual has been recruited from the student subject pool and zero otherwise. "FM transfer" is the first-mover transfer. "Student x FM transfer" is the interaction effect of the two. Socio-economic controls include gender, age (and age squared), being an only child, being a foreigner, being married, having obtained the general qualification for entrance to university or technical college, and dummies for political right- and left-wingers. Full estimation results can be found in Table A.5 in the Online Appendix.
*Significant at 5%; **significant at 1%.

transfers between the two groups reveals only a small difference across the two subject pools (13.17 for nonstudents versus 13.47 for students). An OLS regression of first-mover transfers on a student dummy (column (1) of Table 3) reveals that the observed difference of 0.30 is not statistically significant.[9] The 95% confidence interval for this effect is [–0.9, 1.5]. This reveals that it is very unlikely that first-mover transfers of the two groups differ by more than about 10%. While the uncontrolled regression is the most relevant for our comparison of subject pools, it is also of interest to investigate the role of observable differences. Including control variables allows us to compare participants from the student pool to participants from the general populations with similar socio-demographic backgrounds. Adding control variables changes the sign of the student coefficient, but the effect remains insignificant (see column (2)).[10] Results in column (3) and (4) show that the decisions of students and nonstudents are not driven by different beliefs about the behavior of second movers.

We now turn to second-movers' behavior. Figure 2 shows the average second mover back transfers conditional on first-mover transfer. For every possible first-mover transfer students make lower average repayments than nonstudents. All differences

9. All our results are robust if we use Tobit estimates to account for censoring.

10. Controls variables are gender, age (and age squared), being an only child, being foreigner, being married, having obtained the general qualification for entrance to university or technical college, and political opinions. Full estimation results can be found in Table A.5 in the Online Appendix.

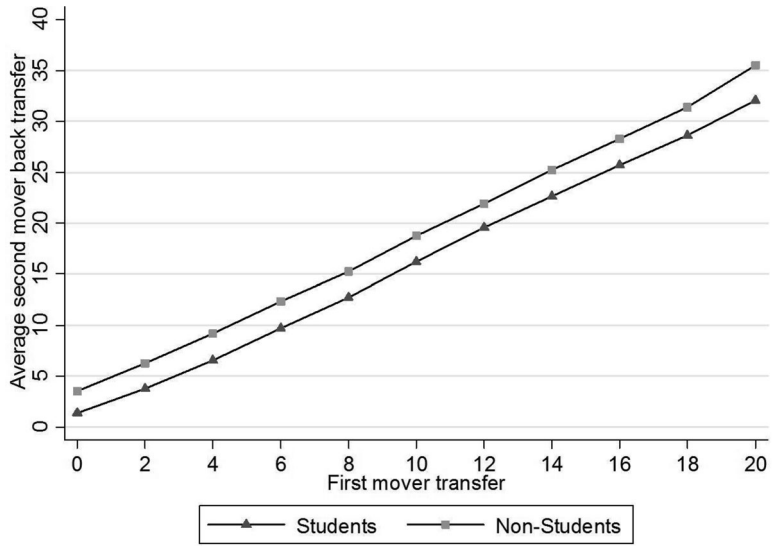

FIGURE 2. Back transfers of students and nonstudents in field trust game. The figure shows average repayments of second movers in the trust game of study 2. The lower line depicts average repayments of participants recruited from the student subject pool of the University of Zurich. The upper line depicts average repayments of participants recruited from a representative sample of the average population of the city of Zurich.

are statistically significant (see Online Appendix Table A.6 for the corresponding p-values). Averaging over all back transfers, students transfer back 15% less than nonstudents. The fact that students transfer back less than nonstudents does not imply that they generally react less sensitively to first-movers' transfers. In fact Figure 2 illustrates that the slope between first-mover transfer and second-mover back transfer is very similar. Put differently, students' and nonstudents' reciprocation pattern is very similar; the only difference being that students reciprocate on a lower absolute level. Column (5) of Table 3 confirms this. It shows an OLS regression with second-movers' back transfers as the dependent variable. We regress back transfers on a student dummy, the first-mover transfer, and the interaction effect between student dummy and first-mover transfer. The coefficient of the student dummy is negative and significant—that is, students transfer back significantly less than nonstudents. However, the interaction effect is close to zero indicating that students and nonstudents exhibit a similar reciprocal inclination as suggested by Figure 2. If we add socio-demographic controls to the regression (see column (6)), the coefficient of the student dummy is no longer significant. This indicates that students are not less prosocial than other participants with a similar socio-demographic background—that is, the difference between the subject pools is driven by the fact that students and nonstudents differ with regard to their socio-demographic background.[11]

11. Table A.5 in the Online Appendix reveals that being married and being a political left-winger significantly increase second-mover repayments.

4. Concluding Remarks

This paper empirically tests whether laboratory experiments with students systematically misrepresent the importance of social preferences. Such an empirical test is critical as experimental methods become increasingly important in economics and experimental results, especially those on social preferences, often challenge insights and policy implications of standard economic models.

Our first study shows that the degree of prosocial behavior in an unrelated field donation does not predict whether (and how often) students participate in experiments. This suggests that self-selection does not significantly bias the social preferences measured in the laboratory. The results of our second study reveal that student participants and nonstudent subjects show very similar behavioral patterns in our trust experiment. While students make less generous repayments, their investment behavior, their beliefs about second-mover behavior, and their reciprocal inclination are very similar to those of participants recruited from the general population.

References

Bardsley, Nicholas, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer, and Robert Sugden (2010). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

Bellemare, Charles and Sabine Kröger (2007). "On Representative Social Capital." *European Economic Review*, 51, 183–202.

Belot, Michèle, Raymond Duch, and Luis Miller (2010). "Who Should be Called to the Lab? A Comprehensive Comparison of Students and Non-students in Classic Experimental Games." Discussion Paper 2010001, University of Oxford.

Benz, Matthias and Stephan Meier (2008). "Do People Behave in Experiments as in the Field? Evidence from Donations." *Experimental Economics*, 11, 268–281.

Berg, Joyce, John Dickhaut, and Kevin McCabe (1995). "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10, 122–142.

Brandts, Jordi and Gary Charness (2011). "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics*, 3, 375–398.

Burks, Stephen, Jeffrey Carpenter, and Lorenz Götte (2009). "Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment." *Journal of Economic Behavior and Organization*, 70, 458–469.

Cleave, Blair L., Nikos Nikiforakis, and Robert Slonim (forthcoming). "Is there selection bias in laboratory experiments? The case of social and risk preferences" Experimental Economics.

Cooper, David and John Kagel (forthcoming). "Other Regarding Preferences: A Selective Survey of Experimental Results." In *The Handbook of Experimental Economics*, Vol. 2, edited by John Kagel and Alvin Roth. Princeton University Press.

Croson, Rachel and Simon Gächter (2010). "The Science of Experimental Economics." *Journal of Economic Behavior and Organization*, 73, 122–131.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2008). "Representative Trust And Reciprocity: Prevalence And Determinants." *Economic Inquiry*, 46, 84–90.

Falk, Armin and James Heckman (2009). "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science*, 326(5952), 535–538.

Fehr, Ernst and John A. List (2004). "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2, 743–771.

Frey, Bruno S. and Stephan Meier (2004a). "Pro-Social Behavior in a Natural Setting." *Journal of Economic Behavior and Organization*, 54, 65–88.
Frey, Bruno S. and Stephan Meier (2004b). "Social Comparison and Pro-Social Behavior: Testing Conditional Cooperation in a Field Experiment." *American Economic Review*, 94, 1717–1722.
Gächter, S. (2010). "(Dis)advantages of Student Subjects: What is your Research Question?" *Behavioral and Brain Sciences*, 33, 92–93.
Henrich, J., S. J. Heine, and A. Norenzayan (2010). "The Weirdest People in the World." *Behavioral and Brain Sciences*, 33, 61–83.
Levitt, Steven D. and John A. List (2007). "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21, 153–174.
List, John A. (2009). "Social Preferences: Some Thoughts from the Field." *Annual Review of Economics*, 1, 21.1–21.20.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

FIGURE A.1. Donations in the Field for Nonparticipants and Participants in Experiments.
TABLE A.1. Participating in Experiments Depending on Prosocial Behavior in the Field.
TABLE A.2. Participating in Experiments Depending on Prosocial Behavior (Alternative Measure).
TABLE A.3. Participating in Multiple Experiments Depending on Prosocial Behavior in the Field.
TABLE A.4. Summary Statistics.
TABLE A.5. First Mover (FM) and Second Mover (SM) Behavior in Field Trust Game.
TABLE A.6. Second Mover Back Transfers Conditional on First Mover Transfer.