

Is There Selection Bias in Laboratory Experiments?

Blair L. Cleave, Nikos Nikiforakis
and Robert Slonim

August 2010

Research Paper Number 1106

ISSN: 0819-2642

ISBN: 978-0-7340-4457-0

Is There Selection Bias in Laboratory Experiments?*

Blair L. Cleave
Department of Treasury and Finance
Victoria, Australia

Nikos Nikiforakis
Department of Economics
The University of Melbourne

Robert Slonim
Department of Economics
The University of Sydney

Abstract

Do the social and risk preferences of participants in laboratory experiments represent the preferences of the population from which they are recruited? To answer this question, we conducted a classroom experiment with a population of 1,173 students using a trust game and a lottery choice task to measure individual preferences. Separately, all 1,173 students were invited to participate in a laboratory experiment. To determine whether selection bias exists, we compare the preferences of the individuals who eventually participated in a laboratory experiment to those in the population. We find that the social and risk preferences of the students participating in the laboratory experiment are not significantly different from the preferences of the population from which they were recruited. We further show that participation decisions across most subgroups (e.g., men vs. women) do not differ significantly. We therefore fail to find selection bias based on social and risk preferences.

Keywords: selection bias; laboratory experiments; external validity; social preferences; risk preferences

JEL Codes: C90, D03

* We would like to thank Jeff Borland who allowed us to run the experiment in his class, Nahid Khan who helped coordinate tutors and Viktoriya Koleva for helping to prepare the experimental material. We would also like to thank Jim Andreoni, Tim Cason, Gary Charness, Rachel Croson, Dirk Engelmann, Ernst Fehr, Simon Gächter, Stephen Leider, John List, Simon Loertscher, Joerg Oechssler and participants at the 2010 Asia-Pacific Meetings of the Economic Science Association at the University of Melbourne, IMEBE 2010 in Bilbao, THEEM 2010 in Thurgau and seminar participants at the University of Melbourne, Monash University, the University of Zürich, the Athens University of Economics and Business, the University of Athens, Royal Holloway University of London for comments. Funding from the Faculty of Business and Economics at the University of Melbourne and the Faculty of Economics and Business at the University of Sydney is gratefully acknowledged.

1. Introduction

Economists have been increasingly using laboratory experiments to study individual preferences (e.g., Andreoni and Miller, 2002; Camerer and Fehr, 2004; Holt and Laury, 2002). The experiments have provided important new insights into the determinants of human behavior by allowing researchers to isolate the effect of preferences from other confounding effects.¹ Accordingly, the findings from this research have had a significant impact on economic theory (e.g., Benabou and Tirole, 2006; Fudenberg and Levine, 2006; Hart and Moore, 2010; Rabin, 1993) and influence on public policy discussions (e.g., Stiglitz, 2010; Thaler and Sunstein, 2008).

While the examination of individual preferences in the laboratory has improved our understanding of human behavior, the typical use of volunteer participants in the laboratory brings to the fore questions about the external validity of experimental findings (e.g., Bardsley et al., 2009; Levitt and List, 2007a, b). One of the most fundamental questions regarding external validity is whether the preferences of volunteer participants are representative of the preferences of the population from which they are recruited, or whether they may differ due to selection bias (e.g., Heckman, 1979; Heckman et al. 1998).

This paper presents the results from a large-scale, controlled experiment investigating whether there is selection bias in economic laboratory experiments. Specifically, the experiment investigates whether the social and risk preferences of volunteer laboratory participants are representative of the preferences of the population from which the participants were recruited. There are two reasons we focus on social and risk preferences. First, both are primitives of economic behavior and the focus of much of the aforementioned experimental research.² Second, it is possible, and indeed has been argued, that social and risk preferences may affect selection into the laboratory. If these preferences affect selection into experiments, then inferences drawn from volunteer laboratory participants may not reflect the preferences of the population they were recruited from.

¹ Some of these insights would be difficult to obtain without laboratory experiments. For example, it can be difficult to ensure that kind actions in the field are not driven by purely selfish reasons; individuals may take kind actions in an attempt to build a good reputation. Laboratory experiments examining preferences typically use simple designs that isolate alternative explanations for behavior to facilitate inference on subjects' motives.

² See Fehr and Gächter (2000) and Harrison and Rutström (2008) for surveys of experimental studies examining social and risk preferences, respectively. As Friedman (2010, p.30) notes, "homegrown preferences over income distribution have taken center stage in experimental economics in recent years." The extent to which pro-social behavior is observed in the laboratory has led to the development of a new generation of economic models providing fundamentally different predictions and insights (for surveys see Fehr and Schmidt, 2006; Sobel, 2005). In this paper, we will say that an individual has social preferences if s/he is willing to sacrifice part of his/her income to increase that of another individual without any potential future monetary or other material benefit from his/her action.

To our knowledge, this is the first attempt to *directly* test for the presence of selection bias by experimentally measuring and comparing the preferences of volunteer laboratory participants to the preferences of the population from which they were recruited. Evidence from psychology experiments suggests that participants in laboratory experiments may be more pro-social than non-participants (e.g., Orne, 1962; Rosen, 1951; Rosenthal and Rosnow, 1973). This could imply that laboratory experiments over-estimate the extent of pro-social behavior in the population. However, the different experimental and recruitment methods used in psychology and economics could result in a different type of selection bias occurring in economic experiments, or indeed no bias at all. It is possible, for instance, that the ubiquitous use of monetary incentives in economic experiments may attract more self-regarding (i.e., less pro-social) individuals. The use of monetary incentives may also affect the sampling of individuals, leading to participants with different tolerances towards risk relative to the population from which they are drawn (e.g., Harrison, Lau and Rutstrom, 2009; Roe, et al., 2009).

To answer our research question, we measured the preferences of a population from which participants are often recruited for economic laboratory experiments – undergraduate students who have had some exposure to economics. Preferences were measured in a classroom experiment using a trust game and a lottery choice task. Over 98 percent of the students we approached agreed to take part in the classroom experiment. We refer to these students as the *population*. This population, consisting of 1,173 students, was invited to participate in a laboratory experiment. To evaluate the existence, extent, and direction of selection bias, we compare the preferences of students who eventually participated in the laboratory experiment (henceforth, *participants*) to those who did not (the *non-participants*). We further examine whether observable characteristics (e.g., gender) affected the decision to participate in the laboratory experiment.

In addition to investigating whether the preferences of participants are representative of the population from which they were recruited, our experiment allows us to explore whether individuals in different subgroups (e.g., men and women) select differently into the laboratory. Laboratory experiments have often been used to investigate whether participants with different observable characteristics – especially gender, but also academic background and nationality – exhibit different types of preferences.³ However, such comparisons could be

³ Examples include comparisons of the behavior of individuals of different gender (e.g., Croson and Gneezy, 2009; Eckel and Grossman, 2008), different academic backgrounds (e.g., Frank, Gilovich, and Regan, 1993), and different nationality (e.g., Henrich, et al., 2001).

biased if members of different subgroups select differently into the laboratory.⁴ In their review of the laboratory evidence on gender differences, Croson and Gneezy (2009) state that research is needed to evaluate whether the observed gender differences in the laboratory could be due to selection bias. Similarly, any observed differences in behavior in the laboratory between students in economics and other disciplines, and between students from different countries could also be due to selection bias.

We find that on average participants are as pro-social and as risk-averse as the population. This result suggests that the social and risk preferences of participants are representative of the population from which the participants were recruited. In addition, we do not find significant differences in participation decisions between subgroups (men vs. women, commerce vs. non-commerce students and domestic vs. international students) based on their social and risk preferences. The one exception we find is that less risk averse women and more risk averse men in the population were more likely to participate in the laboratory experiment. As a result, we find that the difference in risk aversion between women and men is significantly greater in the population than among the participants who select into the laboratory.

The remainder of the paper is organized as follows. Section 2 presents a literature review. Section 3 describes the experimental design with special attention given to describing the recruitment of the population into the laboratory experiment. Section 4 presents the results, focussing on whether social and risk preferences affect selection into the laboratory. Section 5 provides concluding remarks with suggestions for future work.

2. Related Literature

To our knowledge, this is the first study in economics evaluating the impact of non-random selection in laboratory experiments by *directly* comparing the decisions of participants to decisions of the population. At first, this may seem surprising given the long-standing tradition in experimental economics of using volunteers who self-select into experiments as participants. However, there has been a good reason why experimental economists have not generally been concerned about the potential presence of selection bias: participants are usually assigned randomly to treatments and inference is made only regarding the effect of the treatments. This ensures that comparisons *across treatments* are internally valid, as well as

⁴ For instance, Harrison, Lau and Rutström (2009) show that fewer risk-averse people participate in economics experiments when there is greater uncertainty regarding the experimental earnings. If women are more risk averse in the population subjects are recruited from (as they are commonly found to be in laboratory experiments), then men and women may be selecting differently into the laboratory based on their risk preferences.

externally valid so long as selection bias does not result in a sample of participants that responds differently to the treatments than non-participants.⁵

In response to the increasing use of laboratory experiments for the examination of individual preferences, a number of studies have recently appeared investigating the external validity and robustness of laboratory results. One approach examines whether experimental results obtained in the laboratory can be generalized across different laboratory populations. Examples include populations that differ with respect to occupation (e.g., Andersen, Harrison, Lau and Rutström, 2009; Carpenter, Connolly and Meyers, 2008; Cooper, Gu, Kagel and Lo, 1999; Hannah, Kagel and Moser, 2002; Fehr and List, 2004), age (Harbaugh, Krause, Liday and Vesterlund, 2003), and nationality (e.g., Cameron, Chaudhuri, Erkal and Gangadharan, 2009; Henrich et al., 2001; Herrmann, Thöni and Gächter, 2008; Roth, Prasnikar, Okuno and Zamir, 1991). An implicit assumption made in these studies is that any selection bias is identical *across* populations. Our study differs from these studies as we examine whether experimental results can be generalized to the *same* population from which participants are recruited.

A second approach investigating the external validity of experimental results explores whether the results can be generalized across contexts (e.g., laboratory and field) using the same population. For example, Benz and Meier (2008) examine students' giving in the laboratory and towards a university charity; Eckel and Grossman (2000) compare charitable behavior of students in the laboratory ('volunteers') with students in a classroom ('pseudo-volunteers'); Baran, Sapienza and Zingales (2010) examine reciprocal behavior of the same group of MBA students in a laboratory experiment and in a natural field experiment (see also Glaeser et al., 2000). While these studies examine the robustness of behavior *across* contexts, they do not investigate whether the preferences of the laboratory participants are representative of the population from which they are recruited or whether there is selection bias in laboratory experiments.⁶

⁵ We reviewed seven books on experimental economics with extensive references to the experimental literature in economics (Bardsley et al. 2009; Camerer, 2003; Davis and Holt, 1993; Friedman and Sunder, 1994; Holt, 2007; Kagel and Roth, 1995; Plott and Smith, 2008). While all seven books emphasize the importance of random allocation to treatments (some implicitly), only one (Friedman and Sunder, 1994) discusses selection bias. We believe that this omission may be justified given the primary focus on making inferences from treatment effects. In this paper, we do not address the possibility that participants may respond differently to treatments than non-participants. While this form of selection bias would jeopardize the external validity of inferences made from treatment effects, we believe such selection bias is less likely to occur as it would require idiosyncratic selection and treatment interaction effects.

⁶ Another approach for studying external validity is to compare the behavior of laboratory participants with that of a representative sample of the population. In this direction, Bellemare and Kröger (2007) compare the behavior of students in a laboratory trust game to the behavior of a representative sample from the Dutch

A third approach explores whether participants select *out* of experiments after some experience. Casari, Ham and Kagel (2007) find (inter alia) that subjects with higher earnings in a common-value auction experiment are more likely to attend a follow-up auction experiment. Our study differs and complements the study by Casari, Ham and Kagel in two ways. First, we examine whether preferences rather than ability affect selection into experiments. Second, we examine selection at an earlier stage by comparing the behavior of participants and non-participants (i.e., individuals who never participated in a laboratory experiment).

3. Experimental Design and Procedures

This section describes the population in our study, the experimental tasks and treatments, as well as the recruitment process from initial invitation to attendance in the laboratory experiment.⁷

3.1 The population

The experiment was run in the tutorials of Introductory Microeconomics at the Faculty of Business and Economics at the University of Melbourne. Introductory Microeconomics is compulsory for all students in the Faculty of Business and Economics, regardless of whether they major in Economics, Finance, Accounting, Management or Marketing. The course is also taken as an elective by many students outside the Faculty including those majoring in Arts, Science, Engineering and Environments.

Given that economic laboratory experiments are often run using students with different majors and some exposure to economics, we believe that our subject population is similar to the populations commonly used when economists recruit subjects.⁸ There are other reasons for studying this population. Introductory Microeconomics has one of the highest numbers of enrolments at the University of Melbourne. The large and diverse student population provides power to detect selection bias not only over the entire population but also between subgroups. Furthermore, since virtually all students enrolled in the course would have been in their first term at the university, and there had been no recruitment effort during the term, the vast

population who make their decisions at home. While this approach can yield interesting insights, it has the drawback that it can be difficult to invite a representative sample of the population to the laboratory. This approach, therefore, can be seen as a combination of the first two approaches that varies both the subject population and context.

⁷ Supplementary material with the protocols, subject instructions, experimenter instructions and recruitment communications are available at <http://www.economics.unimelb.edu.au/nnikiforakis/>.

⁸ At the time of the classroom experiment, students had only had a one-hour lecture on game theory (see also footnote 10) for more information on the subjects' background.

majority of our population consists of students who have not been exposed to economic experiments.

The Introductory Microeconomics course ran 105 tutorials every week. On the week we conducted the classroom experiment, 1,451 students attended exactly one of the 105 tutorials that were taught by 32 tutors.⁹ Of the 105 tutorials, 17 (with 254 students) were randomly assigned to a control treatment; students in these tutorials were only asked to indicate whether they had interest in attending future laboratory experiments. The remaining 88 tutorials (with 1,197 students) were divided randomly into one of three experimental treatments (see section 3.3).

The classroom experiment was not announced in advance. It was run at the beginning of each tutorial and took approximately 15 minutes to complete. Students who attended the tutorial, but did not wish to participate in the classroom experiment were asked to wait quietly until the experiment was completed. These design choices were made to maximize the likelihood that students would participate in the classroom experiment. Of the 1,197 students in these tutorials, 98 percent (1,173 students) agreed to participate in the classroom experiment.

Demographic information was obtained, with each student's permission, from university administration records and is thus not self-reported. In the control treatment tutorials where a classroom experiment was not run, we did not ask students for permission to access their records in order to be consistent with typical recruitment procedures in experimental economics laboratories. Table 1 presents summary statistics for the other 1,173 students. We refer to the 1,173 students as *the population* that we study. We do not include the 254 students from the control treatment in the population as we have neither demographic nor experimental choice data for them. This population consists of nearly 50 percent female and international students. There were more commerce (73 percent) than non-commerce (23 percent) students.¹⁰ Nonetheless, there were 270 non-commerce students giving us a large sample of each subgroup (male, female, commerce, non-commerce, domestic and international students) within the population.

⁹ More precisely, we instructed tutors to only allow students to participate in the study if the student arrived within five minutes of the beginning of each tutorial. However, among 20 students who arrived past the first 5 minutes four were able to complete the experiment.

¹⁰ By "commerce" we refer to students enrolled at the Faculty of Business and Economics. While we would have liked to examine how economics students select into laboratory experiments, students at the University of Melbourne do not have to select a major until later in their degree. Usually, 25 percent of students at the Faculty of Business and Economics major in economics. Thus, approximately, 20 percent of the individuals in our population are likely to major in economics.

Table 1 – Population characteristics

Observable Characteristics	Number of subjects	Percentage
Female students	572	49 %
Commerce students	854	73 %
International students	587	50 %
Total number of students in classroom experiment	1,173	

3.2 The experimental tasks

We used two common experimental tasks to measure individual preferences.

Trust Game: The trust game we used is a modified version of the original investment game used by Berg, Dickhaut and McCabe (1995). The game involves two players. In our experiment Player A received a \$20 endowment and chose whether to send \$0, \$10 or \$20 to Player B.¹¹ The amount Player A sent was tripled and given to Player B. Player B chose how much to return to Player A from the amount he received. Player A earned $20-x+y$, where x is the amount Player A sent to Player B and y is the amount Player B returned, with support $0 \leq y \leq 3x$. Player B earned $3x - y$ if $x > 0$ and 0 if $x = 0$.

The literature typically assumes that the amount sent measures the level of Player A’s trust and the percent returned measures Player B’s trustworthiness, where the level of trust and trustworthiness increase in the amount Player A sends and in the percent of the amount received that Player B returns, respectively. Potentially, the amount sent and the percent returned can be used to measure an individual’s social preferences. However, for Player A, risk attitudes and beliefs regarding Player B’s behavior make it difficult to interpret the amount sent as reflecting Player A’s social preferences alone. For example, a selfish Player A may send a positive amount if he anticipates that Player B will return more than Player A sent. However, Player B’s decision can be used more directly to measure an individual’s social preferences (specifically, a combination of his preferences over altruism, reciprocation and payoff distribution) since Player B has no payoff uncertainty or strategic incentive in his decision on how much money to return to Player A. We discuss this point in more detail in section 4.

¹¹ In Berg, et al. (1995) Player A can send any whole dollar amount of their \$10 endowment to B. Time constraints meant that we had to use the strategy method in eliciting the response of Player B which requires individuals to state how they would react to each action of Player A. Accordingly, it was practical to restrict Player A’s choices.

The game was played once. Given our interest in the amount returned, we had each student make decisions both as Player A and B in order to observe the amount returned by all participants in the trust game. Furthermore, to save time in the tutorial, when acting as Player B students were asked to indicate how much they would return if Player A sent \$10 and if Player A sent \$20; this allows us to calculate earnings and pay students outside of the tutorial time (discussed below). We thus collected three decisions for each participant: the amount sent as Player A and the amounts returned if they received \$30 and \$60 as Player B. Students received no feedback at any point regarding the decision of any other participant during the experiment. They also did not know which role they would be randomly assigned to when they made their choices, nor did they ever learn who they were paired with, which could have been any student across any of the tutorials running the trust game. All instructions and procedures were common knowledge.¹²

Table 2 – Lottery Choice Task

Lottery Choice	Payoff X (Probability 50%)	Payoff Y (Probability 50%)	Expected Value	Standard Deviation	CRRA Range*
1	\$22	\$22	\$22	0.00	$2.74 < r$
2	\$30	\$18	\$24	6.00	$0.91 < r < 2.74$
3	\$38	\$14	\$26	12.00	$0.55 < r < 0.91$
4	\$46	\$10	\$28	18.00	$0.37 < r < 0.55$
5	\$54	\$6	\$30	24.00	$0.00 < r < 0.37$
6	\$60	\$0	\$30	30.00	Risk Seeking

*CRRA: $U(\$x) = [x^{(1-r)}]/(1-r)$ for $r > 0$ & $r \neq 1$; $U(\$x) = \ln(x)$ for $r = 1$.

Subjects were provided with the information in the first three columns on the left. They did not receive information regarding a lottery's expected value and standard deviation.

Lottery Choice: The lottery choice task is an Ordered Lottery Sequence (OLS) first used by Binswanger (1981). Our OLS task is a slightly modified version used by Eckel and Grossman (2008). Students were presented with six lotteries and had to choose one. Each lottery l had two payoffs, $\$X_l$ and $\$Y_l$, each with a 50 percent probability of being selected. Table 2 shows the payoffs, expected value and two indicators of risk for each lottery. Lottery 1 is risk free with a certain payoff of \$22. The expected payoff for lotteries 2 through 5 increased by \$2 with each lottery. Lotteries 5 and 6 had an identical expected payoff of \$30. The payoffs increased in risk from Lottery 1 through 6 (as measured by the variance in the payoffs or assuming CRRA preferences). Thus, choosing Lottery 1 reflects the greatest risk aversion, while lotteries 2 through 5 reflect the second to fifth most risk averse preferences. A risk

¹² If it is common knowledge that both players receive utility only from their own monetary payoff, all equilibria of the trust game (using the strategy method) require Player A to send \$0. Allowing for trembles, the unique equilibrium also requires Player B to return \$0 for every amount sent.

neutral individual will be indifferent between choosing Lottery 5 or 6 and a risk seeking individual will choose Lottery 6.

3.3 Experimental treatments in the tutorials

Time constraints meant that we could not run both the trust game and lottery task within a single tutorial. Therefore, students participated either in a control condition (no classroom experiment), the trust game or the lottery task during their tutorial. In every treatment, tutors gave each student a flyer asking them whether they would be interested in being included in a new database of volunteers for laboratory experiments.¹³ Interested students had to tick a box with the word ‘yes’ on the flyer and provide their university e-mail address. Students were told that they would later receive information via e-mail regarding how to sign up in the new database. In the control condition, students were only asked to complete the flyer and nothing further was done. We included the control treatment in order to evaluate whether taking part in the class experiment affected student’s participation in future laboratory experiments.

In the experimental treatments, tutors ran either the trust game or the lottery task. Tutors were told neither the purpose of the classroom experiment nor the relationship between the flyer and classroom experiment. Thus, the classroom experiment was run double-blind since neither the experimenters nor the subjects knew the goal of the experiment. To ensure that tutors would not intentionally (or unintentionally) influence decisions in the experiment, they were told that the flyer was part of a separate recruitment drive. Tutors were explicitly instructed not to encourage or discourage anyone to answer ‘yes’ or “no” on the flyer. At no stage were students deceived since they were given the opportunity to participate in future experiments if on the flyer they indicated that they were interested in doing so.

The three experimental treatments were: “Flyer-Trust,” “Flyer-Lottery” and “Trust-Flyer.” In “Flyer-Trust,” students were asked to complete the flyer prior to playing the trust game. In “Flyer-Lottery,” students were asked to complete the flyer prior to making their choice in the lottery task. In “Trust-Flyer” students were given the flyer to complete after playing the trust

¹³ The flyer was the same as the one used for ‘normal’ recruiting at the University of Melbourne. It contained the following information: “The Economics Department runs experiments regularly for research purposes. Individuals are paid for their participation and no experience in economics is necessary. While payments depend on participants’ decisions and the type of the experiment, *average* payments tend to be \$25 per hour. Furthermore, by participating you will be helping promote research at the University. If you would like to receive invitations to take part in economics experiments in the future, please write your university email address below. Note that individuals who sign up for inclusion in the database are *randomly* invited on a weekly basis. This means that each individual in the database will have the same chance to be invited to participate in experiments. If at any point you wish to stop receiving invitations to experiments, you can do this easily by sending an email to economics-experiments@unimelb.edu.au.” The flyer included a field for interested students to write their email address. At the time of the experiment \$1 Australian was worth approximately \$0.90 U.S and the minimum hourly rate in Australia was approximately \$14 per hour.

game. In all three treatments, students were not informed of the second task until everyone had completed the first task. Treatments were randomly allocated across the tutorials.

3.4 The classroom experiment procedures

Running the experiment during the tutorials had several advantages. First, tutorial attendance is a course requirement, with students having to attend eight of eleven tutorial classes in order to pass the course. The attendance requirement ensured that a high percent of the students enrolled in the course would attend the tutorial in which the experiment was run. Second, the number of students attending tutorials was small (averaging 14 students, with a minimum of 6 and a maximum of 18). This allowed tutors to answer any questions privately, and ensure that there was no communication between students during the experiment.

As previously mentioned, the experiments were run by the tutors. Each tutor was given detailed, written information about the experimental procedure including precise instructions about what to say when handing out the flyers and before administering the experiment. Upon arriving at the tutorial, tutors handed out the flyer to the students (except in “Trust-Flyer”). Once everyone completed the flyer, the tutors collected all of them before beginning the experiment to ensure that no student could change his decision after participating in the classroom experiment. The instructions to the students stressed that their decisions would not be revealed to any other student, even if they were chosen to receive payment. It took approximately 3 minutes to complete the flyer and another 12 minutes to complete the classroom experiment.

The classroom experiment was run during the week of May 25th-29th, 2009. On June 2, as we had informed students, we held a public drawing where 160 participants (13.6 percent of the population) were selected for payment and were randomly assigned roles in the trust game.¹⁴ After the draw, a list of the identity numbers of the students chosen for payment was posted on the course’s webpage. Due to final exams and the subsequent winter vacation, students selected for payment were notified by e-mail on August the 6th and they received their payments in private between August the 6th and 13th. No one other than the experimenter and the randomly-selected student knew what choices the student made. Students never learned who they were paired with in the trust game.

¹⁴ While typically all subjects are paid in laboratory experiments, this was not possible in our experiment due to its large scale. To compensate for the low probability of being selected for payment, expected payments were high relative to students’ opportunity cost of their 10 minutes participation in the experiment. For instance, in the trust game Player A could guarantee \$20 for himself and in the lottery task students could guarantee \$22. Moreover, the low payment probability should not affect our results regarding selection bias. We can test this by comparing participation rates in the control and in the other treatments. As we will see, we find no difference in participation rates between the control and any of the treatments (see Figure 1 and footnote 16 below).

3.5 Recruitment into laboratory sessions

Most experimental databases are populated with volunteers who are invited to participate in experiments over an extended period of time. In universities with large databases (such as the database at the University of Melbourne with more than 2,500 volunteers), it can take a while before subjects are invited for an experiment. For this reason, we decided to wait 3 months before sending an e-mail to the students who had expressed an interest in participating in economics experiments (from their flyer decisions). The email, sent on August 17th, reminded the students that they had expressed an interest in experiments during the microeconomic tutorials and invited them to register at a new, purpose-made, on-line, volunteer database. Registering involved (i) reading and agreeing with the rules governing the use of the database, and (ii) providing socio-demographic information such as their gender, nationality, age, and major area of study. Registering took about five minutes.

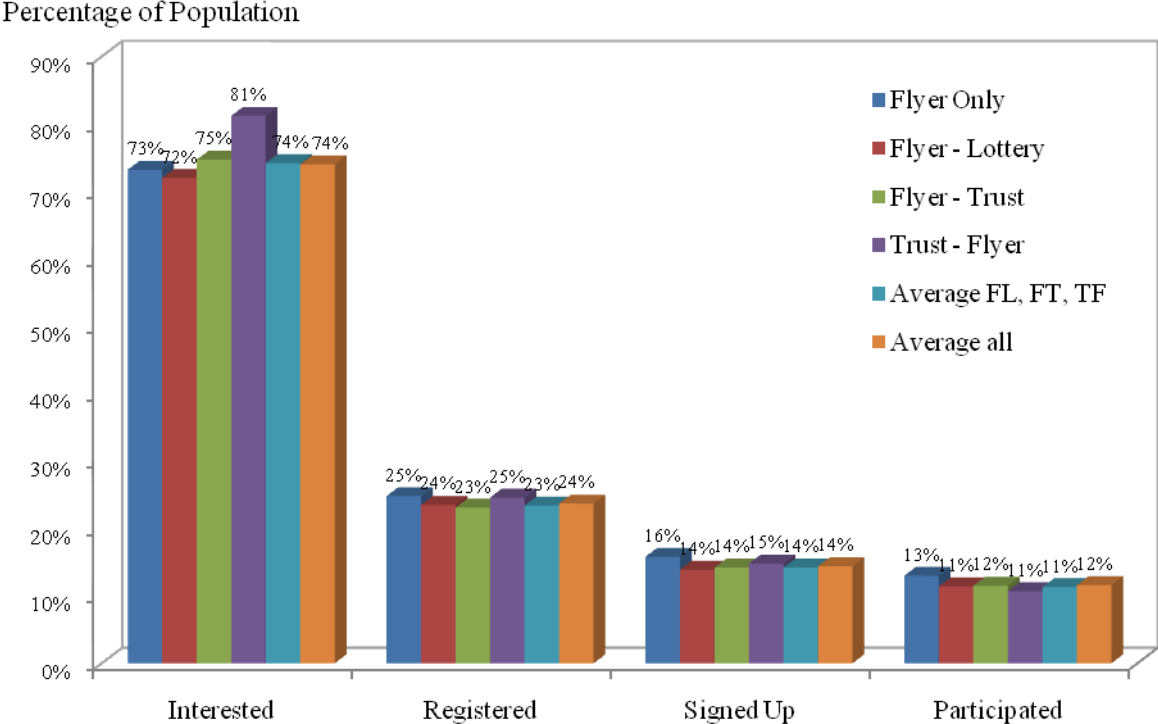
On October 8th, the students who had registered in the database received an e-mail informing them that an experiment was being run. The e-mail included information about the dates and times of the sessions, and the web-address they could click on to register. In total, we included eight experimental sessions that were spread over two weeks and at different times of the day. Of the eight laboratory sessions, three never filled up, and two of these three sessions were chronologically the last two of the laboratory sessions. This suggests that most (if not all) students wanting to participate were not being excluded due to space limitations.

Figure 1 shows the attrition of the population starting from the flyer response to ultimate participation in the laboratory. Most students (74 percent) initially expressed interest in participating in laboratory experiments (the flyer decision).¹⁵ However, only one-third of those expressing interest originally (and 24 percent of the overall population) registered in the on-line database to become eligible to participate. This may reflect the low cost of expressing an initial interest during class (i.e., simply ticking a ‘yes’ on the flyer) relative to reading laboratory guidelines and filling out an online form with multiple fields. Of the students who registered into the database, 58 percent (14 percent of the overall population) signed up to take part in any of the laboratory sessions. Finally, only 12 percent of the population actually came to any of the laboratory sessions. We are unaware of any evidence regarding attrition in

¹⁵ Figure 1 shows that a higher proportion of students in the Trust-Flyer treatment marked ‘yes’ on the flyer than in the other treatments. This suggests that after immediately participating in the classroom experiment student interest was higher than among those who had not participated in a classroom experiment. However, this initial difference in interest disappears when subjects are asked to register into the database. Thus, the effect of having first participated in the experiment in the “Trust-Flyer” treatment was fleeting.

previous experimental economics laboratory studies. This level of attrition indicates that there could be non-trivial inference concerns for the study of population preferences.

Figure 1 – Population Attrition



‘Interested’ denotes the percentage of the population who expressed an interest in receiving an email about signing up to the recruitment database; ‘Registered’ denotes the percentage of the population who registered on the database; ‘Signed Up’ denotes the percentage of the population who signed up for a laboratory experiment; ‘Participated’ denotes the percentage of the population who participated in the laboratory experiment.

Figure 1 shows that neither participation in the classroom experiment (comparing the control to all treatments) nor the order in which the flyer was handed out had an effect on participation rates – across the three treatments and the control condition a virtually identical 12 percent of the students from the tutorials eventually participated in the laboratory experiment.¹⁶ Figure 1 also shows that the majority of those who self selected out of participating in the laboratory experiment did so by not registering in the database; 76 percent of the population chose not to register into the database and thus never received any further information about laboratory experiments.

¹⁶ To test whether participation rates differed across treatments and between the treatments and the control, we ran Probit regressions with the dependent variable equal to 1 if the student participated in a laboratory session and equal to 0 otherwise, and we included dummy variables for each treatment. We estimated the model with robust standard errors at the level of the tutor. For every pair-wise comparison we found no significant difference (p -value > .20 in all cases) between treatments, between each treatment and the control, and between the control and the combined treatments. Thus, we conclude that running the classroom experiment did not affect the likelihood that students would ultimately attend a laboratory session.

4. Results

We first examine whether selection into laboratory experiments can be explained by observable characteristics. We then investigate whether selection into laboratory experiments can be explained by social and risk preferences. We conclude by examining whether there is selection bias within subgroups of the population based on the individual preferences.

4.1 Observable Characteristics and Selection: Main Effects

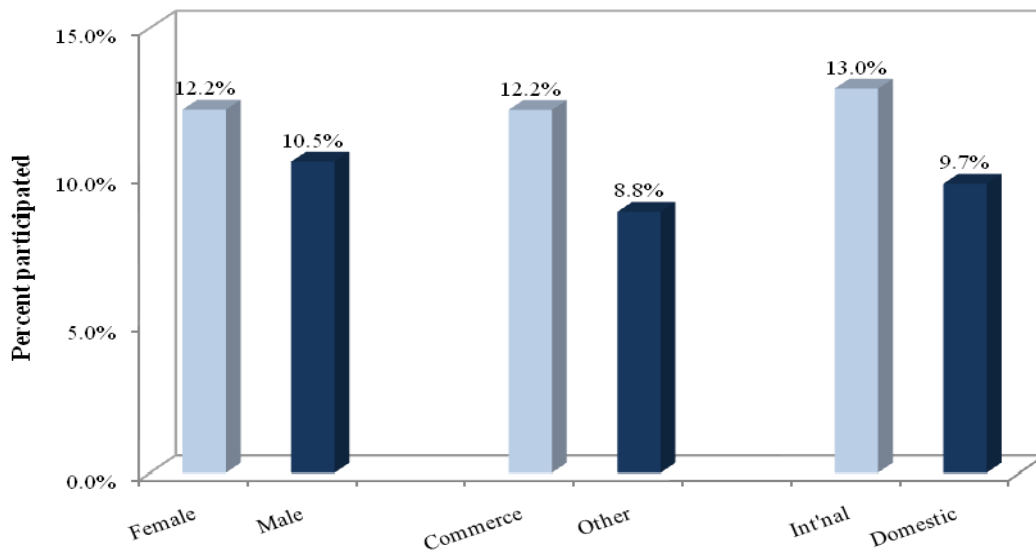
Figure 2 shows that on average women were 1.7 percent more likely than men to participate in the laboratory experiment, while commerce and international students were 3.4 and 3.3 percent more likely to select into the laboratory experiment than non-commerce and domestic students, respectively. To test whether these differences are statistically significant, we estimate the following probit model:

$$(1) \quad y_i = \alpha + \beta_1 \text{female}_i + \beta_2 \text{commerce}_i + \beta_3 \text{international}_i + \sum \delta_j X_{ji} + \varepsilon_{ii}$$

where y_i equals 1 if student i participated in the laboratory experiment and equals 0 otherwise, female_i , commerce_i and international_i are dummy variables indicating whether the student is female, a commerce major and international, respectively. X_{ji} is a vector of dummy variables controlling for (i) the day of the week the student attended the tutorial, (ii) whether the flyer was administered after the classroom experiment, and (iii) whether the student was randomly selected to get paid for the classroom experiment. Although tutors were given identical training and instructions to follow for all sessions, it is possible that some of them had distinct effects on student's likelihood to select into the laboratory. We thus report robust errors clustered at the level of the tutor.

Table 3 presents the estimates from Model 1. Column 1 presents estimates without the additional controls or tutor effects, Column 2 adds tutor random effects and Column 3 adds the control variables. In all three regressions, the estimates for female, commerce and international students do not reach the 10-percent level of significance, indicating that we cannot detect any significant selection bias for these observable characteristics. Column 3 indicates that the inclusion of the other control variables neither improves significantly the model fit (the log-likelihood does not increase significantly), nor has an effect on the estimates for any of the other variables. Although the exclusion of these covariates does not qualitatively affect the results presented at any point in the paper, we always include them for completeness.

Figure 2 – Percentage of population participating in laboratory experiment by observable characteristics



The last column of Table 3 shows the marginal effect of each observable characteristic on selection. Although none are statistically significant, the point estimates indicate that commerce and international students were on average 2.7 and 2.5 percent more likely to select into the laboratory than non-commerce and domestic students (p -values = .12 and .20), respectively. We summarize:

Result 1: *On average, we do not detect any significant selection bias based on observable characteristics.*

Table 3 - Effects of observable characteristics on selection into the laboratory experiment

	(1)	(2)	(3)	Marginal Effects
Female	0.057 (0.098)	0.057 (0.127)	0.053 (0.127)	0.010 (0.024)
Commerce	0.159 (0.116)	0.159 (0.100)	0.153 (0.106)	0.027 (0.018)
International	0.133 (0.100)	0.133 (0.106)	0.135 (0.104)	0.025 (0.020)
Constant	-1.426*** (0.112)	-1.426*** (0.122)	-1.611*** (0.134)	
Tutor Effects	No	Yes	Yes	Yes
Fixed Effects	No	No	Yes	Yes
Log likelihood	-412.00	-412.00	-408.72	-408.72
Observations	1173	1173	1173	1173

Probit estimates with robust standard errors at the tutor level in parentheses.

Fixed effects include day effects, order effects and random selection for payment effects.

* p -value < .10; ** p -value < .05; *** p -value < .01

4.2 Social and Risk Preferences and Selection: Main Effects

As mentioned in section 3, our cleanest measure of social preferences is the percent returned by the second movers in the trust game. Nevertheless, for completeness, we will also examine whether the amount sent is correlated with selection into the laboratory experiment.

Percent Returned: Figure 3a and Columns 1-2 in Table 4 investigate whether there is selection bias based on the average percent returned, r , by each student ($r = \frac{1}{2} \sum_{x=10}^{20} \frac{y_x}{3x}$).

Column 1 re-estimates Model 1 for the sub-sample of students who gave a response to both return decisions.¹⁷ These estimates are qualitatively the same to the estimates from the full sample (Column 3, Table 3) except that in this sub-sample women are marginally significantly more likely than men to participate in the laboratory experiment (p -value $<.10$).

Figure 3a shows the percent of students who selected into the laboratory based on the percent they returned. We classify the students' choices into six categories for the graphical presentation, with each category having a distinct interpretation for social preferences: students who on average returned (a) $r=0$ (suggesting no social preferences), (b) $0 < r < 1/3$ (suggesting minimal social preferences by returning an amount less than was sent), (c) $r=1/3$ (suggesting moderate social preferences by returning the exact amount sent), (d) $1/3 < r < 1/2$ (suggesting substantial social preferences by returning more than the amount sent, but keeping more for themselves), (e) $r=1/2$ (suggesting high social preferences by equally dividing the amount received) and (f) $r > 1/2$ (returning over half the amount received). Only four percent of the students (25/602) returned more than one-half the amount received.

Figure 3a shows a slight downward trend in selection into the laboratory the more the students returned. Twelve and 14 percent of students who returned nothing (i.e., $r=0$) and between nothing and one-third (i.e., $0 < r < 1/3$), respectively, selected into the laboratory. Approximately, 10 percent of students who returned from one-third to one-half of the amount received (i.e., $1/3 \leq r \leq 1/2$) selected into the laboratory. Only four percent of the students who returned over one half of the amount received selected into the laboratory. Column 2 in Table 4 confirms the direction of this relationship. The marginal effects in Column 2m reveal that

¹⁷ Some students specified a return amount for only one of the possible actions of Player A. Past evidence using the strategy method (e.g., Bellemare and Kroger 2007; Garbarino and Slonim, 2009) along with the current data show that the higher the amount sent by Player A, the higher the percent returned by Player B. Thus using either decision alone when subjects only indicated how much they would return for only one of the two amounts sent will be biased compared to using the average percent returned. We thus exclude from the analysis in this section observations from 61 individuals who did not indicate how much they would return for both amounts that could have been sent. However, our results are qualitatively not affected if we include these observations in our analysis.

for an additional one percent returned, a student was 0.9 percent less likely to select into the laboratory. This implies that a student who returned nothing was 3.33 percent and 5 percent more likely to select into the laboratory than a student who returned one-third or one-half of the amount received, respectively. However, this estimated effect does not reach a conventional level of significance (p -value = .16).

Amount Sent: Figure 3b and Columns 3-4 in Table 4 investigate whether there is selection bias based on the amount sent in the trust game. Column 3 in Table 4 re-estimates Model 1 for the sub-sample of students who played the trust game. These estimates are qualitatively the same as the estimates from the full sample (Column 3, Table 3).

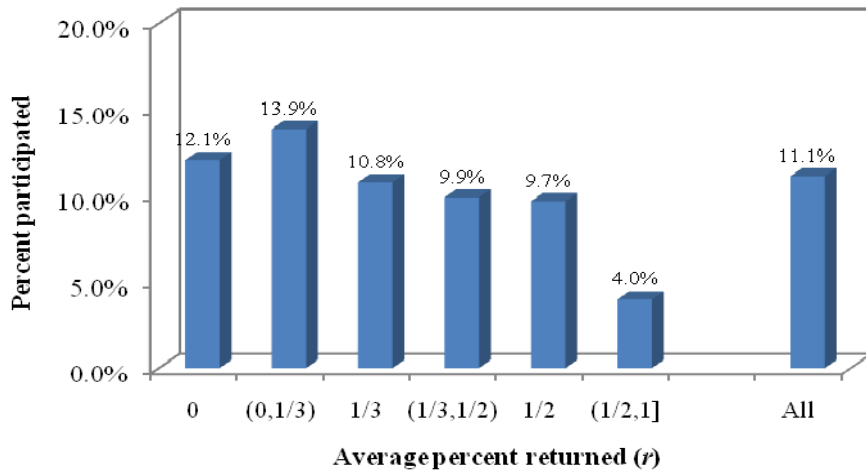
Figure 3b shows a negative relationship between the amount students sent in the classroom experiment and their likelihood to participate in the laboratory experiment. Nearly 16 percent of students who sent \$0 in the trust game selected into the laboratory. By contrast, less than 9 percent of the students who sent \$10 or \$20 selected into the laboratory. Column 4 in Table 4 adds the amount sent as an explanatory variable to Model 1. The estimate on amount sent shows that on average the more a student sent, the less likely the student was to select into the lab. This relationship is significant at the 1-percent level and the marginal effects in Column 4m indicate that for every additional \$10 sent during the classroom experiment, a student was, in absolute terms, 4.3 percent less likely to select into the lab. In relative terms, as can be seen in Figure 3b, the marginal effects indicate that students who sent nothing were twice as likely to participate in the laboratory experiment as the students who sent \$20.

Lottery choice: Figure 3c and the last three Columns of Table 4 investigate whether there is selection bias based on risk preferences. Column 5 re-estimates Model 1 for the sub-sample of students who received the lottery choice. Similar to the full sample, there is again no significant difference in who selects into the lab in the sub-sample of students who completed the lottery task based on the observable characteristics. Figure 3c further shows no discernable pattern in who selects into the lab based on the lottery choice. The estimates in Column 6 confirm that on average the effect of choosing a more risky lottery has no significant effect on the likelihood of selecting into the laboratory (p -value > .40). We summarize the main result from this subsection.

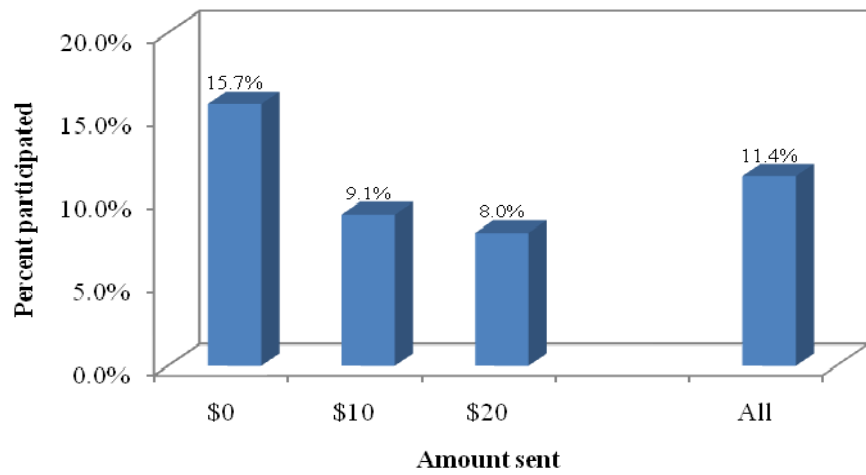
Result 2: *Students who sent the least in the trust game were more likely to select into the laboratory experiment. However, we do not detect any significant difference in the likelihood to participate in the laboratory experiment based on either the lottery choice or the percent returned in the trust game.*

Figure 3– Percentage of population participating in the laboratory experiment by choices in the classroom experiment

3a. Percent Returned



3b. Amount Sent



3c. Lottery Choice

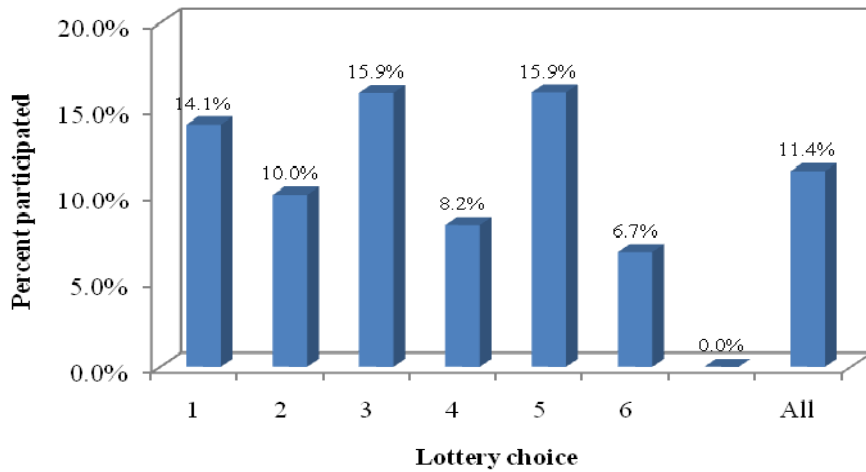


Table 4 - Effects of social and risk preferences on selection into the laboratory experiment

	(1)	(2)	(2m) Marginal Effects	(3)	(4)	(4m) Marginal Effects	(5)	(6)	(6m) Marginal Effects
Female	0.292* (0.167)	0.311* (0.172)	0.057* (0.032)	0.181 (0.156)	0.201 (0.159)	0.037 (0.030)	-0.078 (0.167)	-0.111 (0.186)	-0.021 (0.035)
Commerce	0.162 (0.153)	0.141 (0.168)	0.025 (0.028)	0.165 (0.155)	0.108 (0.162)	0.019 (0.028)	0.162 (0.173)	0.155 (0.175)	0.027 (0.028)
International	0.009 (0.136)	0.039 (0.135)	0.007 (0.025)	0.038 (0.127)	0.083 (0.127)	0.015 (0.023)	0.222 (0.162)	0.199 (0.144)	0.037 (0.026)
Ave. Percent Returned		-0.513 (0.430)	-0.093 (0.079)						
Amount Sent					-0.023*** (0.009)	-0.0043*** (0.002)			
Lottery Chosen								-0.046 (0.051)	-0.009 (0.010)
Constant	-1.58*** (0.15)	-1.44*** (0.20)		-1.56*** (0.15)	-1.37*** (0.16)		-1.77*** (0.26)	-1.58*** (0.34)	
Tutor Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-likelihood	-205.21	-204.21	-204.21	-228.83	-225.84	-225.84	-176.08	-175.59	-175.59
Observations	602	602	602	658	658	658	510	510	510

Probit Estimates with robust standard errors at the tutor level in parentheses.

Fixed effects include day effects, order effects and random selection for payment effects.

* p -value < .10; ** p -value < .05; *** p -value < .01

Result 2 implies that on average the amount returned and the lottery choices of laboratory participants should be representative of the choices of the population, whereas the amount sent among laboratory subjects will be less than in the population. Table 5 presents the mean and standard deviation of the participants' choices across the tasks along with the population (true) means we want to estimate. The t-tests in Table 5 examine whether participants' mean choices differ significantly from the population means.

Table 5 reveals that the average percent returned and lottery choice for the participants is 11 and 7 percent lower than that for the population. However, as anticipated, neither difference reaches the level of significance (p -value > .10). Table 5 also shows that laboratory participants sent nearly 25 percent less than the population. The average amount participants sent is significantly lower than the population average (p -value = 0.02). Thus, inference on the average percent returned and lottery choice based on choices of the laboratory participants does not differ significantly from the population's preferences. In contrast, inference on the amount sent of the laboratory participants differs from the population due to selection bias.

Table 5 – Comparison of choices of participants and population

	Percent Returned	Amount Sent	Lottery Choice
Participants	22.8% (2.32%) [67]	\$6.13 (0.83) [75]	3.40 (0.22) [58]
Population	25.9% [602]	\$8.13 [658]	3.64 [510]
<i>t</i> -test: Participants Average = Population Average	$t = -1.33$ p -value = 0.19	$t = -2.42$ p -value = 0.02	$t = -1.14$ p -value = 0.26

Entries are averages, standard deviation is in parentheses and the number of observations is in squared brackets

4.3 Discussion of Result 2

To the extent that the *amount sent* reflects social preferences, Result 2 implies that laboratory experiments may *under-estimate* the percentage of pro-social individuals in the population. However, as mentioned in Section 3.2, there are several reasons to be cautious about interpreting the amount sent as a clean measure of social preferences, where social preferences imply that an individual's utility does not depend solely on his/her material payoff, but also on the material payoff of other relevant individuals (Fehr and Fischbacher, 2002). First, the amount sent may partly be based on preferences other than social preferences such as risk aversion (e.g., Ashraf, Bohnet and Piabkov, 2006; Cox, 2004; Karlan, 2005) and betrayal aversion (Bohnet and Zechhauser, 2006). Second, the amount sent is likely to be affected by beliefs regarding the amount that will be returned (e.g., Ashraf, Bohnet and Piabkov, 2006; Garbarino and Slonim, 2009). This implies that selection into the laboratory of the subjects sending the least amount in the trust game could indicate selection based on risk aversion and betrayal aversion preferences or beliefs, rather than only social preferences.

The percent returned in the trust game is a cleaner measure of social preferences. It may reflect preferences over distributional outcomes, altruism or reciprocity. Although there are important distinctions between these motivations, all imply that the individual's utility depends on the payoffs of another person, and consequently represents social preferences. Result 2 indicates that with respect to the percent returned, we are unable to detect a significant selection bias on average in social preferences in who participates and who does not participate in the laboratory experiment. Finally, Result 2 also indicates that with respect to the lottery choice, we are also unable to detect a significant selection bias on average in risk preferences in who participates and who does not participate in laboratory experiments. We thus find little support *on average* for selection bias with respect to social or risk preferences.

4.4 Social and Risk Preferences and Selection Bias among Subgroups

Figures 4-6 and the regressions in Table 6 disaggregate the analysis to examine whether there is selection bias based on the average percent returned, amount sent or lottery choices that differs between the three subgroups: male and female students, commerce and non-commerce students, and international and domestic students. We will therefore make nine pair-wise comparisons (three choices by three subgroups) to examine whether there is differential selection across subgroups.

Table 6 includes the interaction of each choice in the classroom experiment (percent returned, amount sent and lottery choice) with each observable characteristic (gender, major, and international status) to test whether selection differs significantly between the subgroups on each behavior. The figures show the percent who participated for each subgroup by the average percent returned (Figure 4), the amount sent (Figure 5) and lottery choice (Figure 6). Panels a, b and c in each figure show the percent who participated by gender, commerce student status and international student status.

Figure 4 shows no discernable difference in selection based on the average percent returned between men and women (Figure 4a) or between international and domestic students (Figure 4c). The insignificant estimates for the interaction terms in Columns 1 and 3 of Table 6 confirm that the difference in participation rates between men and women and between international and domestic students does not depend on the average percent returned. However, among students who returned greater than one third, the commerce students appear to be relatively more likely to participate in the laboratory experiment than the non-commerce students. The significant interaction term in Column 2 confirms that commerce students were more likely than non-commerce students to participate in the laboratory experiment the more they returned.

Figure 5a shows that women who sent nothing were 12 percent more likely to participate in the laboratory experiment than men while women who sent either \$10 or \$20 were only 1 percent more likely to participate. Figures 5b and 5c reveal no obvious difference in participation rates between commerce and non-commerce students or international and domestic students based on the amount sent. The interaction terms in Columns 4-6 in Table 6 show that none of these differences reach the level of significance, indicating that we cannot detect any differences in selection bias on the amount sent between the different subgroups.¹⁸

¹⁸ Although the Female*Amount Sent interaction term does not reach the level of significance, Figure 4a and the coefficient for 'Female' in Column 1 of Table 6 suggest that among students who sent nothing, women may be relatively more likely than men to participate in a laboratory experiment.

Table 6 - Effects of social and risk preferences by observable characteristics on selection into the laboratory experiment

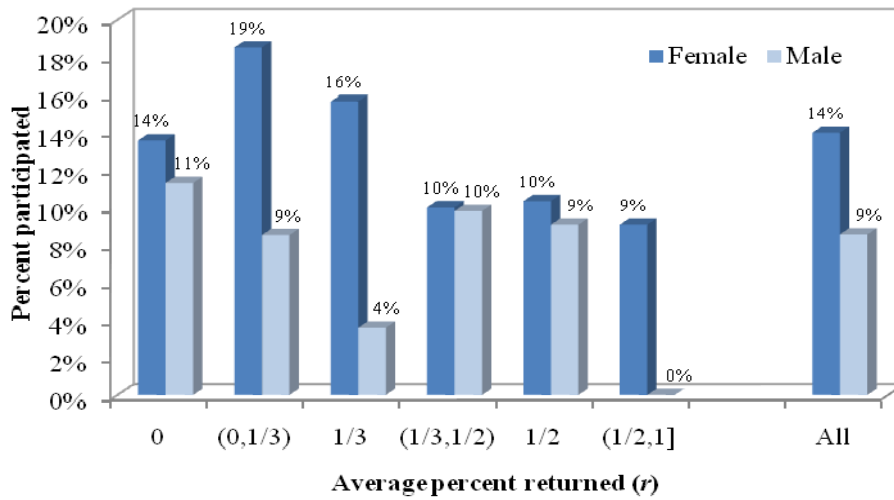
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.266 (0.193)	0.325* (0.176)	0.309* (0.172)	0.365* (0.208)	0.201 (0.160)	0.202 (0.160)	-0.954*** (0.305)	-0.108 (0.187)	-0.108 (0.186)
Commerce	0.143 (0.168)	-0.158 (0.244)	0.138 (0.169)	0.107 (0.163)	0.083 (0.222)	0.102 (0.161)	0.166 (0.184)	-0.128 (0.242)	0.152 (0.174)
International	0.041 (0.136)	0.029 (0.14)	0.182 (0.171)	0.081 (0.126)	0.081 (0.129)	0.019 (0.177)	0.196 (0.152)	0.202 (0.143)	0.508 (0.321)
Ave. Percent Return (APR)	-0.602 (0.556)	-1.473** (0.748)	-0.298 (0.503)						
Female * APR	0.183 (0.493)								
Commerce * APR		1.231* (0.658)							
International * APR			-0.021 (0.014)						
Amount Sent				-0.012 (0.013)	-0.026 (0.016)	-0.028** (0.013)			
Female * Amount Sent				-0.023 (0.018)					
Commerce * Amount Sent					0.003 (0.019)				
International * Amount Sent						0.009 (0.015)			
Lottery Choice							-0.155** (0.064)	-0.104 (0.082)	0.001 (0.046)
Female * Lottery Choice							0.240*** (0.081)		
Commerce * Lottery Choice								0.078 (0.076)	
International * Lottery Choice									-0.085 (0.065)
Constant	-1.43*** (0.211)	-1.20*** (0.252)	-1.49*** (0.209)	-1.45*** (0.183)	-1.35*** (0.186)	-1.33*** (0.169)	-1.16*** (0.311)	-1.38*** (0.401)	-1.77*** (0.366)
Tutor Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	602	502	597	658	658	658	510	510	510

Probit estimates with robust standard errors at the tutor level in parentheses. Fixed effects include day effects, order effects and random selection for payment effects.

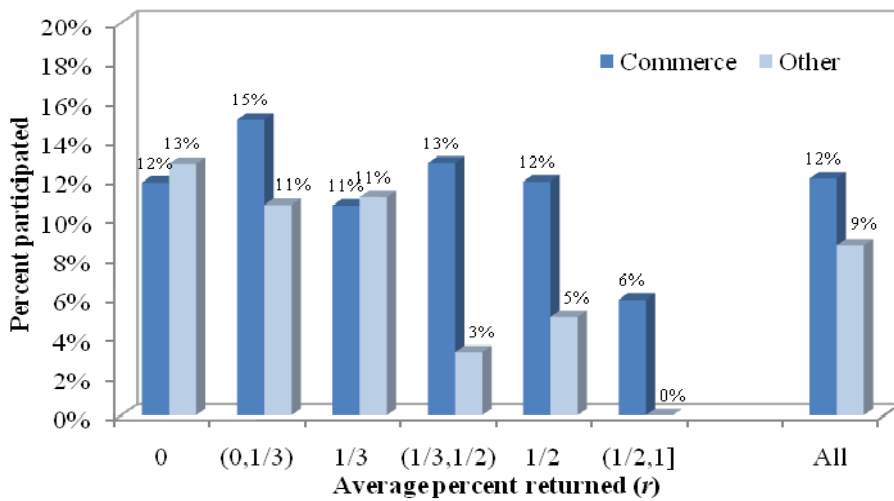
* p -value < .10; ** p -value < .05; *** p -value < .01

Figure 4 – Percentage of population participating in the laboratory experiment by average percentage returned in the trust game

4a. Female and Male



4b. Commerce and Other



4c. Domestic and International

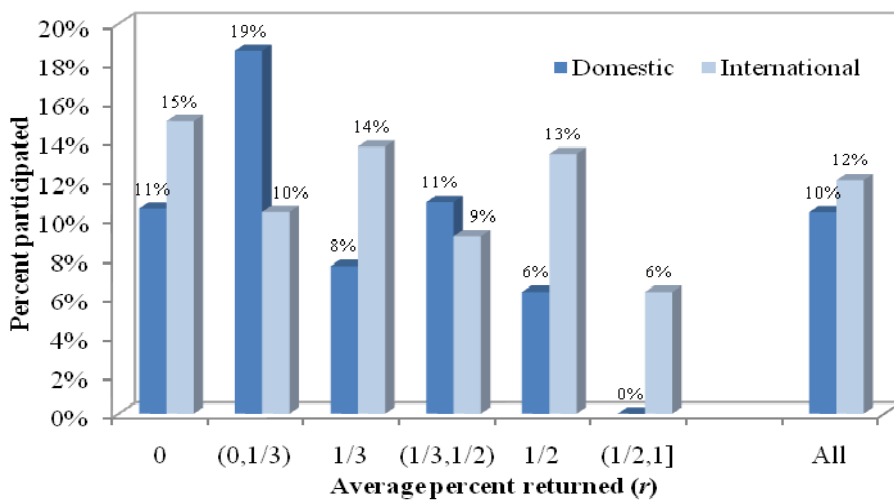
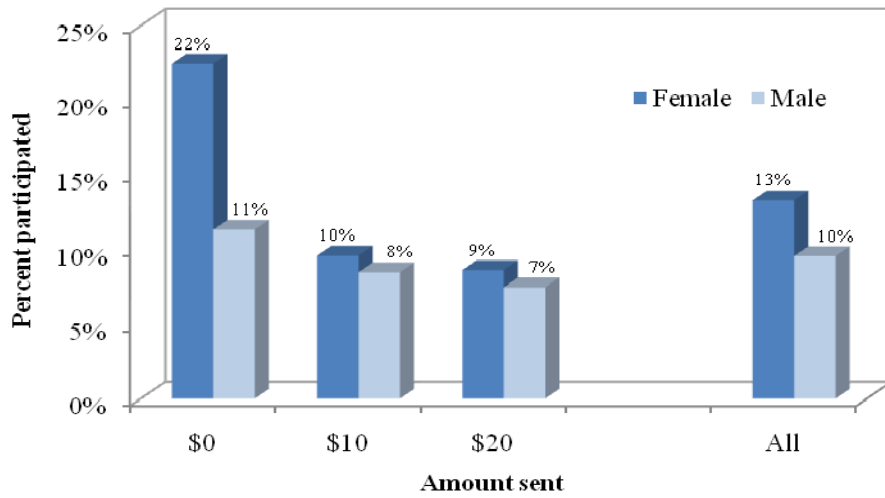
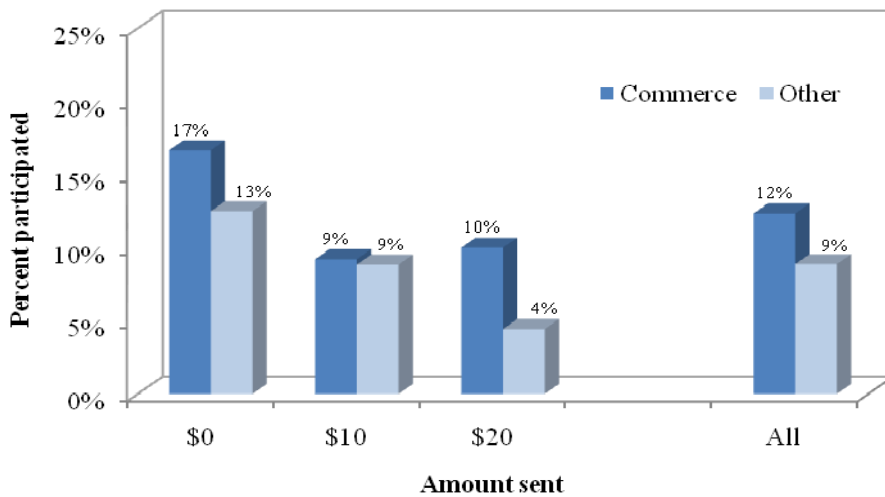


Figure 5 – Percentage of population participating in the laboratory experiment by amount sent in the trust game

5a. Female and Male



5b. Commerce and Other



5c. Domestic and International

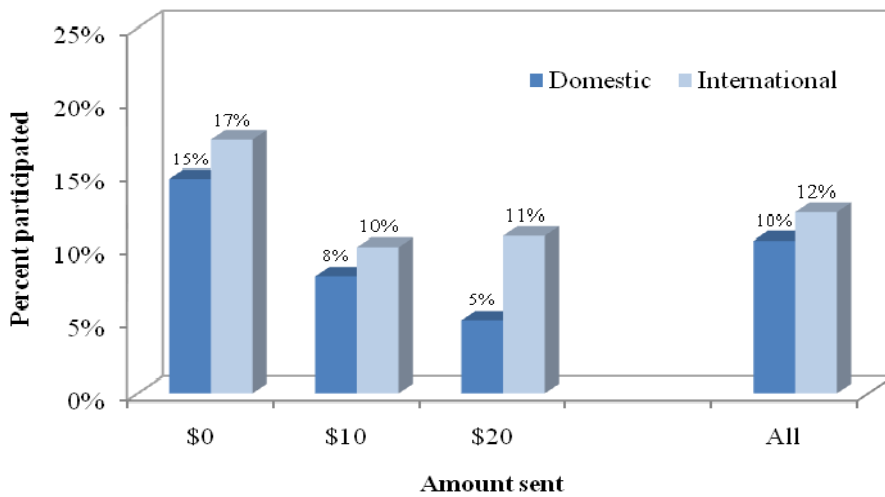
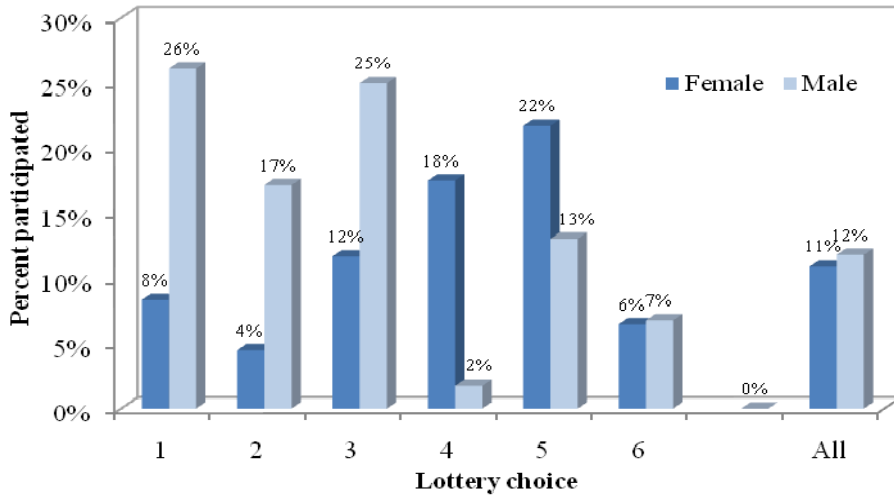
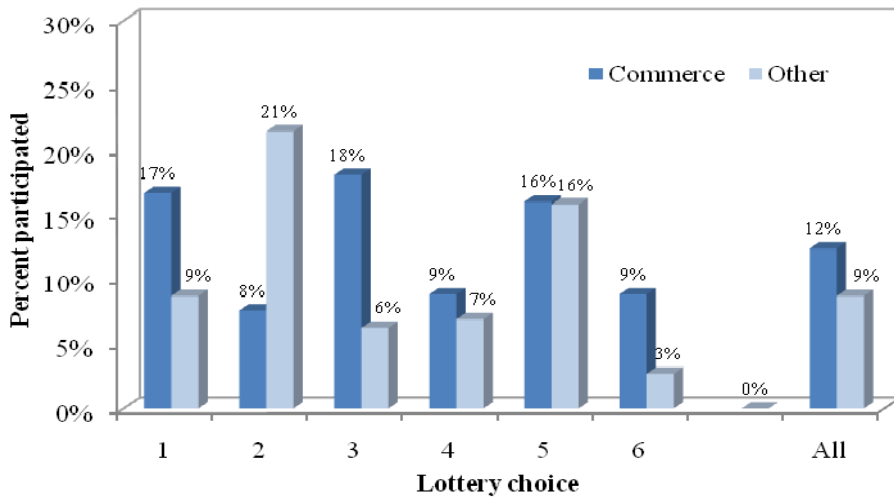


Figure 6– Percentage of population participating in the laboratory experiment by lottery chosen

6a. Female and Male



6b. Commerce and Other



6c. Domestic and International

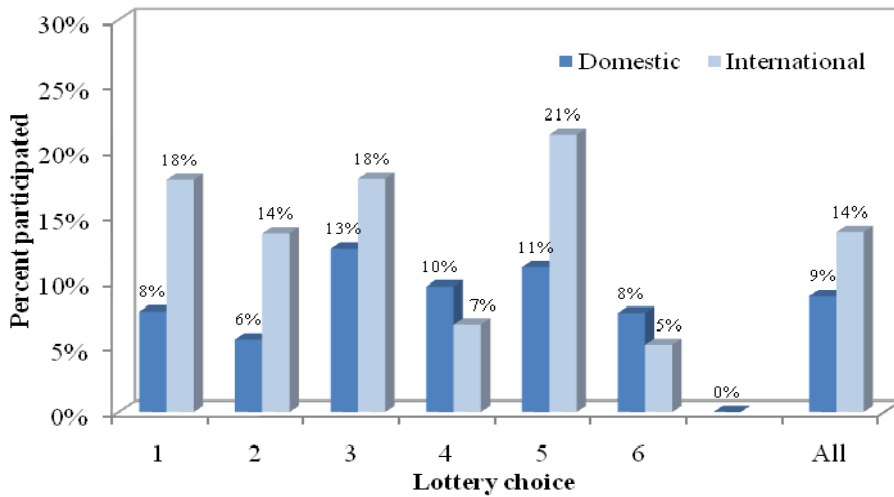


Figure 6a reveals that, although men and women were overall equally likely to participate in the laboratory experiment, among the students who chose the least risky lotteries (Lotteries 1-3) men were more likely to participate. In contrast, among the students who chose the more risky lotteries (Lotteries 4-6) women were the most likely to participate. Figures 6b and 6C show no discernable pattern in participation among commerce and non-commerce students or among international and domestic students based on lottery choices. The interaction estimates in Columns 7-9 of Table 6 confirm these observations. Column 7 shows that among students who chose the risk-free lottery, women were significantly less likely to participate (negative significant estimate on the main effect on Female), and among the students who chose increasingly risky lotteries men were less likely to participate (negative significant estimate on lottery choice) while women were relatively and absolutely more likely to participate (positive significant estimate on the Female*Lottery Choice). The marginal effects (not shown in the table) indicate that the likelihood women select into the laboratory relative to men increases by 4.4 percent for a one level shift towards a riskier lottery choice. The marginal effects overall indicate that while women were 12.9 percent *less* likely than men to select into the laboratory among those who chose the least risky lottery, women were 9.2 percent *more* likely than men to select into the laboratory among those who chose the most risky lottery. Thus, the relatively more risk averse men and relatively less risk averse women in the population are the ones who participated in the laboratory experiment. Finally, the interaction estimates in Columns 8 and 9 indicate that commerce and non-commerce student participation decisions did not differ by the lottery choices they made. Similarly, there is no significant difference in the participation decision between international and domestic students based on their lottery choices. We summarize:

Result 3: *Of the nine pair-wise tests examining whether there was selection bias between subgroups, we only detect one significant and one marginally significant difference in selection. Women were significantly more likely than men to participate in the laboratory experiment the riskier the lottery they chose, and commerce students were marginally significantly more likely to participate than non-commerce students the more they returned in the trust game.*

4.5 Discussion of Result 3

We identified one significant and one marginally significant instance of selection bias between subgroups (out of a total of nine pair-wise comparisons). These differences could jeopardize inference regarding differences in the population from laboratory data. To demonstrate the implications of this selection bias consider the following. Suppose we ran the same trust and lottery experiments in the laboratory as we ran in class and assume that the

students who selected into the laboratory would make the same choices in the laboratory as they did in class. The important question for inference is which, if any, results based on the behavior of the laboratory participants would not generalize to the population?

Table 7 presents the results of this exercise. The first column presents estimates of the determinants of the average percent returned from a Tobit regression censored at the minimum (0) and maximum (100) potential percent returned. Column 3 presents estimates of the determinants of amount sent from an ordered Probit regression with the dependent variable taking on the three values for the amount sent: \$0, \$10 and \$20. Column 5 shows estimates of the determinants of the lottery choice from an ordered Probit model with the dependent variable taking on values from 1 (Lottery 1) to 6 (Lottery 6). Columns 1, 3 and 5 present estimates using only the laboratory participants. Below each estimate, we report robust standard errors.

Columns 2, 4 and 6 in Table 7 present the equivalent (true) parameters using the whole population. The critical inference question here is whether the estimates from the participants in the laboratory sample deviate from the true population means. We are thus interested in whether the laboratory estimates are significantly different from the population means rather than different from 0. Thus, below each estimate in Columns 1, 3 and 5 we report the results of two tests. The first p -value indicates whether the parameter estimates from the laboratory sample is different than 0. We present this test since this is what is reported in a typical paper in which the true population values are unknown. The second p -value indicates whether the parameter estimates from the laboratory sample are significantly different than the true population values shown in Columns 2, 4 and 6. If there is significant selection bias in the choices between subgroups, then the laboratory sample estimates will differ from the population true values. By construction, there is no other reason that the inference from the laboratory sample should deviate from the true population values other than selection bias. However, it is possible that we may not detect an incorrect inference if the laboratory sample size is sufficiently small and thus does not provide us with enough power to reject the estimates from the true population values.

Table 7 shows that drawing inferences about the population from the students who eventually participated in the laboratory experiment would have led to only one incorrect conclusion about the population (shaded area). In particular, we would have incorrectly concluded that there is no gender differences in risk preferences when in fact women are more risk averse than men in the population. Given that a number of laboratory studies have recently explored the risk preferences of men and women (see Croson and Gneezy 2009 for a

review), further study of how the non-random selection of participants affects inference on gender differences in risk preferences seems warranted. Nonetheless, this is the only inference from the laboratory participants that is not externally valid to the population from which the laboratory participants were recruited.¹⁹

Table 7: Inference on the effect of observable characteristics on preferences of laboratory participants compared to the population

	Percent Returned ¹		Amount Sent ²		Lottery Choice ³	
	Participants	Population	Participants	Population	Participants	Population
Female	4.4%	3.3%	-0.188	0.179	0.056	-0.587
	(4.8%)		(0.316)		(0.262)	
<i>Estimate=0</i>	<i>p=.369</i>		<i>p=.552</i>		<i>p=.831</i>	
<u><i>Estimate=Population</i></u>	<i>p=.820</i>		<i>p=.246</i>		<i>p=.014</i>	
Commerce	-1.9%	-4.7%	-0.571	-0.405	0.187	-0.104
	(5.6%)		(0.235)		(0.342)	
<i>Estimate=0</i>	<i>p=.739</i>		<i>p=.015</i>		<i>p=.584</i>	
<u><i>Estimate=Population</i></u>	<i>p=.622</i>		<i>p=.481</i>		<i>p=.395</i>	
International	4.5%	6.1%	0.562	0.270	-0.691	-0.273
	(6.8%)		(0.282)		(0.282)	
<i>Estimate=0</i>	<i>p=.512</i>		<i>p=.046</i>		<i>p=.014</i>	
<u><i>Estimate=Population</i></u>	<i>p=.813</i>		<i>p=.301</i>		<i>p=.138</i>	
Constant	25.4%	26.7%				
	(5.2%)					
Test: All 3 Estimates = Population Estimates	F(3,24)=0.014 p=.935		$\chi^2(3)=3.83$ p=.283		$\chi^2(3)=9.08$ p=.028	
Tutor Effects	Yes		Yes		Yes	
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Log-likelihood			-67.22		-98.98	
R ²	0.15					
Observations	67	602	75	658	58	510

1 Tobit Regressions censored at Percent Return lower limit equals 0 and upper limit equals 100

2 Ordered Probit Regressions

3 Ordered Probit Regressions

Bold: True population value

Shaded Cell: Laboratory sample estimate differs significantly from population

Each cell contains the parameter estimate and robust standard errors in parentheses. Tutor random effects are included. Fixed effects include day effects, order effects and random selection for payment effects.

¹⁹ Recall from Column 2 in Table 6 that the interaction between commerce students and the average percent returned was marginally significant. This indicated that commerce students in the population were more likely than non-commerce students to participate in the laboratory experiment the more they returned. We see this result directionally in Columns 3 and 4 of Table 7. The inability to detect an incorrect inference in the presence of the observed selection bias is likely due to insufficient power since (i) this selection bias was only marginally significant and (ii) only a subset of the 67 laboratory participants were non-commerce students.

5. Conclusion

One of the main purposes of empirical investigation is to explore how broadly theories and the underlying assumptions can be generalized. It is generally hoped that empirical findings can be generalized from one population and one setting to another. However, a more fundamental question is whether selection bias prevents inference from a sample of volunteers to the population from which they are drawn.

Understanding individual preferences has been an increasingly central theme in experimental economics in recent years. The investigation of individual preferences has had a significant impact on economic theory as evinced, for example, by a new generation of economic models abandoning the assumption that individuals are selfish (see Fehr and Schmidt, 2006; Sobel, 2005). However, the non-random selection of participants into laboratory experiments implies that selection may affect the external validity of the results.

We conducted a large, controlled experiment to investigate whether the social and risk preferences of volunteer participants are representative of the population from which the participants were drawn. We first measured individuals' preferences in a classroom experiment with 1,173 students. Separately, we invited all 1,173 students to attend a laboratory experiment. Our main result is that participants' social and risk preferences are found to be representative of those in the population. Specifically, we found that participants return on average the same amount in the trust game and choose the same lottery as non-participants. We also found that participants sent less than non-participants in the trust game. While this suggests there is selection bias based on the amount sent, many factors other than social preferences affect this decision, such as risk aversion and beliefs, making it difficult to determine the type of preferences (or beliefs) that might be driving this selection.

The absence of selection bias in our data suggests that inferences on social and risk preferences from laboratory experiment can be externally valid to the population the participants are recruited from. This finding is without doubt good news for experimental economists studying individual preferences and for theorists who use insights from laboratory experiments in their models. However, more experiments are needed to evaluate whether selection on social and risk preferences creates inference problems more generally (e.g., across countries and institutions).

Selection in laboratory experiments is likely to be a function of the recruitment method used. Although we designed our recruitment method to be similar to the ones used in many economics laboratories, different methods could affect selection. This has several implications

for future research. First, we believe that systematic research is needed to evaluate how different recruitment methods affect selection. This will facilitate the comparison of results from different laboratories. Second, given that replication is essential in empirical research, experimental studies should discuss the recruitment method used (i.e., the information provided for the recruitment of volunteers into databases and later for the recruitment of volunteers in experiments) along with the experimental instructions.²⁰ One possibility is that each laboratory standardizes its recruitment procedure and posts details online.

Third, experiments similar to ours are needed to evaluate selection bias in different populations that are of interest. We studied a population that is often targeted when recruiting volunteers for participation in laboratory experiments. However, this is not the only population that is of interest to economists measuring individual preferences. Different factors may determine selection across populations. Some of them (e.g., earnings, level of education, age) will be easier to control than others which may also affect selection (e.g., individual preferences).

Finally, experiments are needed to evaluate whether selection based on other preferences occurs (e.g., inter-temporal choice preferences). Until more empirical work clarifies the nature and extent of selection bias in laboratory experiments, researchers need to be cautious about the implicit and explicit inferences they draw from experimental data regarding the underlying preferences of a population.

²⁰ Most economic journals require that authors of experimental papers submit the experimental instructions and provide details for the experimental procedures followed, but not for the recruitment procedure. An exception is *Econometrica*, which also requests information about the recruitment procedures.

References

- Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and Elisabet Rutström (2009). Preference Heterogeneity in Experiments: Comparing the Field and Laboratory, forthcoming *Journal of Economic Behavior and Organization*.
- Andreoni, James and John Miller (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, *Econometrica*, 70 (2), 737-753.
- Ashraf, Nava, Iris Bohnet, and Nikita Piankov (2006). Decomposing Trust and Trustworthiness, *Experimental Economics*, 9(3).
- Baran, Nicole, Paola Sapienza, and Luigi Zingales (2010). Can We Infer Social Preferences from the Lab? Evidence from the Trust Game, CEPR Discussion Paper No. DP7634.
- Bardsley, Nicholas; Cubitt, Robin; Loomes, Graham; Moffatt, Peter; Starmer, Chris & Sugden Robert (2009). *Experimental Economics: Rethinking the Rules*, Princeton University Press; New Jersey.
- Bellemare, Charles, and Sabine Kröger (2007). On Representative Social Capital, *European Economic Review*, 51(1), 181-202.
- Benabou, Roland and Jean Tirole (2006). Incentives and Prosocial Behavior, *American Economic Review*, 96 (5), 1652-1678.
- Benz, Matthias and Stephan Meier (2008). Do people behave in experiments as in the field? Evidence from Donations, *Experimental Economics*, 11 (3), 268-281
- Berg, Joyce, John Dickhaut and Kevin McCabe (1995). Trust, Reciprocity, and Social History, *Games and Economic Behavior*, 10(1) 122-142.
- Binswanger, Hans P. (1981). Attitudes toward risk: Theoretical Implications of an Experiment in Rural India. *Economic Journal*, 91, 867-890.
- Bohnet, Iris, and Richard Zechhauser (2006). Trust, risk and betrayal, *Journal of Economic Behavior and Organization* 55, 467-484.
- Camerer, Colin (2003). *Behavioral Game Theory*, Princeton University Press.
- Camerer, Colin, and Ernst Fehr (2004). Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists, in *Foundations of Human Sociality*, edited by J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, Oxford University Press, Oxford 2004.
- Cameron, Lisa, Ananish Chaudhuri, Nisvan Erkal, and Lata Gangadharan (2009). Propensities to Engage in and Punish Corrupt Behavior: Experimental Evidence from Australia, India, Indonesia and Singapore, *Journal of Public Economics*, 93 (7-8), 843-851.
- Carpenter, Jeff, Cristina Connolly and Caitlin Myers (2008). Altruistic Behavior in a Representative Dictator Experiment, *Experimental Economics*, 11(3), 282-298.
- Casari, Marco, John Ham and John Kagel (2007). Selection bias, demographic effects and ability effects in common value auctions experiments, *American Economic Review*, 97 (4), 1278-1304.

- Cooper, David J., John Kagel, Wei Lo and Qing Liang Gu (1999). Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review*, 89 (4), 781-804.
- Cox, James (2004). How to Identify Trust and Reciprocity, *Games and Economic Behavior*, 46 260-281.
- Croson, Rachel, and Uri Gneezy (2009). Gender Differences in Preferences, *Journal of Economic Literature*, 47(2) 448–74.
- Davis, Douglas, and Charles Holt (1993). *Experimental Economics*, Princeton University Press.
- Eckel, Catherine, and Phillip Grossman (2000). Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method in Dictator Experiments, *Experimental Economics*, 3, 107–120.
- Eckel, Catherine, and Phillip Grossman (2008). Forecasting Risk Attitudes: An experimental Study Using Actual and Forecast Gamble Choices, *Journal of Economic Behavior & Organization*, 68(1) 1-17.
- Fehr, Ernst and John List (2004). The Hidden Costs and Rewards of Incentives – Trust and Trustworthiness among CEOs, *Journal of the European Economic Association*, 2, 741-771.
- Fehr, Ernst, and Urs Fischbacher (2002). Why Social Preferences Matter – The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives, *Economic Journal*, 112, C1-C33.
- Fehr, Ernst, and Simon Gächter (2000). Fairness and Retaliation – The Economics of Reciprocity, *Journal of Economic Perspectives*, 14, 159-181.
- Fehr, Ernst, and Klaus M. Schmidt (2006). The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories, in *Handbook on the Economics of Giving, Reciprocity and Altruism*, Vol. 1, edited by Serge-Christophe Kolm and Jean Mercier Ythier, Amsterdam: Elsevier.
- Frank, Robert; Thomas Gilovich and Dennis Regan (1993). Does Studying Economics Inhibit Cooperation? *Journal of Economic Perspectives*, 7(2) 159–71.
- Friedman, Daniel (2010). Preferences, Beliefs and Equilibrium: What Have Experiments Taught Us? *Journal of Economic Behavior and Organization* 17, 29-33.
- Friedman, Daniel, and Shyam Sunder (1994) *Experimental Methods: A Primer for Economists*, Cambridge University Press.
- Fudenberg, Drew, and David K. Levine (2006). A Dual-Self Model of Impulse Control, *American Economic Review*, 96 (5), 1449-1476 .
- Garbarino, Ellen, and Robert Slonim (2009). The Robustness of Trust and Reciprocity Across a Heterogeneous Population, *Journal of Economic Behavior and Organization*, 69(3) 226-40.
- Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman and Christine L. Soutter (2000). Measuring Trust, *The Quarterly Journal of Economics*, 115 (3) 811-846.

- Greiner, Ben (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. *University of Cologne, Working Paper Series in Economics* 10.
- Hannah, Lynn, John Kagel and Don Moser (2002). Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences and Effort Requests on Behavior, *Journal of Labor Economics*, 20 (4), 923-951.
- Harbaugh, William T., Kate Krause, Steven G. Liday and Lise Vesterlund (2003). Trust in Children, in *Trust, Reciprocity and Gains from Association* edited by Elinor Ostrom and James Walker, Russell Sage Foundation, New York.
- Harrison, Glenn W., Morten Lau and Elisabet Rutström (2009). Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments, *Journal of Economic Behavior and Organization* 70, 498–507.
- Harrison, Glenn W., and Elisabet. Rutström (2008). Risk Aversion in Experiments, in J. C. Cox and G. W. Harrison (eds.) *Research in Experimental Economics*, 12, 41–196, Emerald Group Publishing Limited.
- Hart, Oliver, and John Moore (2010). Contracts as Reference Points, forthcoming in *The Quarterly Journal of Economics*.
- Heckman, James (1979). Sample Selection Bias as a Specification Error, *Econometrica*, 47 153–61.
- Heckman, James, Hidehiko Ichimura, J. Smith and Petra Todd (1998). Characterizing Selection Bias Using Experimental Data, *Econometrica* 66, 1017-98.
- Henrich, Joseph, Robert Boyd, Sam Bowles, Colin Camerer, Herbert Gintis, Richard McElreath and Ernst Fehr (2001). In Search of Homo economicus: Experiments in 15 Small-Scale Societies, *American Economic Review*, 91(2) 73-79.
- Herrmann, Benedikt, Christian Thöni and Simon Gächter (2008). Antisocial Punishment Across Societies, *Science*, 319(5868), 1362-7.
- Holt, Charles (2007). *Markets, Games, and Strategic Behavior*, Pearson Education Inc.
- Holt, Charles, and Susan Laury. 2002. "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644--1655.
- Kagel, John, and Alvin Roth (1995). *Handbook of Experimental Economics*, Princeton University Press.
- Karlan, Dean (2005). Using Experimental Economics to Measure Social Capital and Predict Financial Decisions, *American Economic Review*, 95(5), 1688-1699.
- Levitt, Steven D., and John A. List (2007a). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World, *Journal of Economic Perspectives*, 21(2), 153-174.
- Levitt, Steven D., and John A. List (2007b). Viewpoint: On the Generalizability of Lab Behaviour to the Field, *Canadian Journal of Economics*, 40(2), 347-370.

- Orne, Martin (1962). On the Social Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications, *American Psychologist*, 17(10), 776-83.
- Plott, Charles R., and Vernon L. Smith (2008). *Handbook of Experimental Economics Results*, North Holland, New York.
- Rabin, Matthew (1993). Incorporating Fairness into Game Theory and Economics, *American Economic Review*, 83 (5), 1281-1302.
- Roe, Brian E., Timothy C. Haab, David Q. Beversdorf, Howard H. Gu, and Michael R. Tilley (2009). Risk-attitude Selection Bias in Subject Pools for Experiments Involving Neuroimaging and Blood Samples, *Journal of Economic Psychology*, 30 (2), 181-189.
- Rosen, Ephraim (1951). Differences between Volunteers and Non-Volunteers for Psychological Studies, *Journal of Applied Psychology*, 35, 185-193.
- Rosenthal, Robert W., and Ralph L. Rosnow (1973). *The Volunteer Subject*, New York: John Wiley and Sons.
- Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study, *American Economic Review*, 81, 1068-1095.
- Sobel, Joel (2005). Interdependent Preferences and Reciprocity, *Journal of Economic Literature*, 43(2), 392-436.
- Stiglitz, Joseph, E. (2010). *Freefall: America, Free Markets, and the Sinking of the World Economy*, New York: W. W. Norton & Company.
- Thaler, Richard, and Cass Sunstein (2008). *Nudge*, Yale University Press.