Who should be called to the lab?

A comprehensive comparison of students and non-students in classic experimental games

Michele Belot[*], Raymond M. Duch[†], and Luis M. Miller[‡]

February 27, 2012

Abstract

This study exploits the opening of the experimental lab in Oxford to compare the behavior of students and non-students in a number of classic experimental games, some of which are primarily other-regarding (Trust Game, Dictator Game, and Public Goods Game) and others which have game forms that are strategically challenging (Beauty-contest and Second-price Auction). We find that students are more likely to behave as homo-economicus agents than non-students. In none of the five

[*]Nuffield College, University of Oxford, michele.belot@nuffield.ox.ac.uk

[†]Nuffield College, University of Oxford, raymond.duch@nuffield.ox.ac.uk

[‡]University of the Basque Country, luis.m.miller@gmail.com

games is there evidence of significant differences in comprehension between students and non-students. Our findings suggest that students are atypical in fundamental determinants of behavior, such as the ability to reason strategically and social preferences. Experiments using students are likely to provide a lower bound estimate on deviations from homo-economicus behavior in the general population.

Keywords: laboratory experiments, external validity

JEL classification: C72 C91

1 Introduction

Most subject pools in experiments conducted in social sciences are drawn from undergraduate student populations (Morton and Williams 2009; Peterson 2001). Based on a survey of 60 laboratory experimental economics papers published in major experimental economic journals, Danielson and Holm (2007) found that only four did not use students as subjects.

Students are usually perceived as a "convenience sample" and are usually preferred to non students for methodological reasons. We expect them to understand the instructions better, be more at ease in a laboratory environment and be better able to reason in the abstract than other people. For all these reasons, students may be better able to behave "normally" (as they would outside the lab) than other people. They are also considered to be a more homogenous population. These advantages allow experimental researchers to obtain insights at relatively low costs and with relatively small sample sizes. But the question is how externally valid are the results found with student samples? More specifically, do students differ in fundamental determinants of behavior in economic games? For example, do they differ in their ability to reason strategically, in their attitudes towards risk or in their pro-sociality? And does this matter for their behavior in experimental games? If that is the case, then the results from experiments conducted on student samples may have limited external validity. If on the other hand differences between students and non-students are an artefact of the lab (because for example students are better able to understand the instructions), then we would have a stronger case for limiting our experimental sample to student populations.

We invited 63 students and 65 non-students to participate to the first experiment conducted in the Nuffield Centre for Experimental Social Sciences. Both subject pools participated in sessions conducted in the exact same manner. The idea here is simply to compare two different samples, neither of which can be claimed to be representative of the whole population, at least *ex ante*. The subjects all played the same five classic experimental games; some of which are primarily other-regarding (Trust Game, Dictator Game, and Public Goods Game)and others which have game forms that are strategically challenging (Beauty-contest and Second-price Auction). We first identify the extent to which behavior differs across these games and across subject pools. Then, we explore whether variations in behavior are an artefact of the lab or reflect differences in fundamentals driving behavior.

Our paper provides three main contributions to the literature (discussed in the next section). First, the experiment is conducted in a new laboratory with no reputation. This is important because differences in students and non-students observed in a lab operating for a period of time could be due to differential selection effects or learning. Students may be better informed about the type of experiments conducted in the lab than others and know better "how to play the games". Second, we compare behavior across a range of very different games, rather than concentrating on one type of games. This will allow us to provide a broad picture of the relative importance of strategic reasoning, risk and social preferences. Finally, we present original new methods to assess whether the differences are an artefact of the lab or not. We introduce a measure of understanding of the instructions, we evaluate learning effects and finally, we have a measure of external validity.

We find substantial differences in behavior in most games. In games that involve other-regarding preferences, students are two to three times more likely to choose the homo-economicus equilibrium strategy than non-students. In contrast, the differences in behavior are smaller and not always significant in games that do not engage other-regarding preferences. We find no evidence that these differences are due to differences in levels of comprehension of the games. The differences are most dramatic for the simplest of the games (the Dictator Game) and controlling for a measure of understanding of the instructions does not change the results. The differences also subsist when subjects have time to learn. The Public Goods Game was repeated 10 times and the differences are as pronounced at the beginning as at the end of the experiment. On the other hand, behavior is correlated with individual covariates such as age and cognitive ability (in games involving cognitive reasoning), suggesting that students differ in fundamental determinants of behavior.

We provide a mild test of external validity of our results. We implement a non-obtrusive test of cooperative behavior that takes place outside of the, potentially contrived and abstract, lab environment. Again, differences in other-regarding behavior persist between subject pools. Hence, neither result suggests that our subject pool differences are an artefact of the experimental design.

Summarizing, our results show students differ from non-students across a range of games and show that these differences are not an artefact of the lab. We cannot claim that either of these samples is more or less externally valid. The conclusion from our work is that attempts to generalize findings outside the lab either have to be specific to the group under investigation, or otherwise need to be considered as lower bounds of deviations from homo economicus behavior and a full representative sample will be needed to obtain averages that are valid for the general public.

2 Related literature

Despite the large reliance on students in laboratory work in economics, there is only a limited number of methodological studies comparing students to non-students (Carpenter et al. 2005; Carpenter et al. 2008; Anderson et al. 2010; Falk et al. 2012). Other social sciences (e.g. psychology, marketing) with a long tradition of experimental work have identified important behavioral differences between students and non-students, for example in personality tests, perceptions and attitudes, etc. A comprehensive meta-analysis conducted by Peterson (2001) suggests that student samples tend to be much more homogeneous than non-student samples and that the treatment effect sizes differ significantly across these two types of samples. But we know relatively little about differences in behavior in economic games.

There are a few studies comparing pro-social behavior between students and other groups of the population. Carpenter et al. (2005) compares students and workers in both a laboratory and field setting and finds that workers were more generous in the Dictator Game while students made more generous offers in the Ultimatum Game. A comparison of contributions of students versus community members in a charitable donation version of the Dictator Game found that students were much less generous (Carpenter et al., 2008). More recently, Anderson et al. (2010) find that self-selected students appear considerably less pro-social than self-selected adults in a prisoner's dilemma game. Cappelen et al. (2011) show that a nationally representative adult sample behave more pro-socially than a student sample. Falk et al. (2012) find that non-students make significantly more generous repayments in a trust game. These authors conclude that results from student samples might be seen as a lower bound for the importance of prosocial behavior.

Next to these, there are a couple of studies comparing behavior in games that do not engage other-regarding preferences. A classic study is the comparison of newspaper and lab Beauty-contest experiments (Bosch-Domenech et al., 2002) which concludes that the lab and field generate comparable results. Depositario et al. (2009) find no difference in the bidding behavior of students and non-students in a uniform price auction.

Hence there is some evidence that the behavior of student and non-student subject pools differ. But the evidence remains very limited, focuses on one or two games and is usually obtained in labs that are already operating. Also, we do not know whether these differences could be an artefact of the laboratory environment or not. We have implemented a set of experiments designed to calibrate the overall magnitude of these subject pool differences and to assess whether, and how, these differences vary across types of games. The design allows us not only to explore whether game type conditions subject pool differences in behavior but also test whether these differences are an artefact of the lab or not.

3 Experimental design

We exploited the opening of the lab at the Nuffield Centre for Experimental Social Sciences in Oxford. This was the first experiment carried out in the lab. Hence, the subjects have not yet been exposed to experiments carried out in the lab. Accordingly, past participation in CESS lab experiments or information about CESS experiments cannot explain outcomes in this experiment.

At the time of the experiment, the CESS subject pool consisted of 1,000 students and non-students: 75 percent are students and 57 percent are females. Half of the students are freshmen and they come from more than 30 different disciplines. Half of the non-students are private employees and there is also a significant number of workers, self-employees, public employees, and unemployed.

Students were recruited mainly on the freshers fair in Oxford, while non students were recruited in various ways. We e-mailed non-academic staff working at the university, we contacted dozens of local shops in Oxford, placed advertisements in a local newspaper and local pubs. The advertisement mentioned that we were looking for people from all walks of life to participate in decision-making experiments and surveys in social sciences. We also mentioned that participants should expect to earn between £10 and £15 per hour. The typical gross pay of students working for Oxford university is £12, and the average salary of an administrator in the UK in 2008 was £16,994, which corresponds to an hourly rate of £8.5.[1]. Because the participants are informed about their expected payment, we would not expect systematic differences in the opportunity cost of participating. These two populations are both local and both prepared to participate to experiments. Thus, we are comparing two pools that could realistically be used in experiments conducted at Oxford.

Table 1 describes the six sessions of the experiment; two with students, two with non-students and two with a mixed population (a total of 128 subjects).[2] All sessions were conducted at 5.30 or 6 pm. Each session lasted for about one hour and a half. Upon arrival, subjects were assigned randomly to a desk in the lab. We read out the instructions.[3] We presented the experiment as consisting of two parts. The first part involved 6 different situations, which we asked subjects to treat independently of each other. We read the instructions for the first situation; subjects were instructed to take a decision after we read

[1]Source: http://www.mysalary.co.uk/

[2]We assigned students and non-students to sessions with only students, only non-students or a mix of both. We found no evidence of differential play according to the session type. The analysis presented here pools the results across sessions

[3]The instructions can be found in the appendix.

the instructions. We then moved on to the second situation and so on. Subjects received no feedback about their payoffs in each of these situations until all six situations were completed (except in the last situation, which featured a repeated interaction). Earnings for this part were determined by selecting one of the six situations randomly (subjects were informed of the payment scheme at the very beginning). The second part of the experiment was a test of cognitive ability consisting of 26 questions. Subjects received £0.20 for each correct answer.[4]

Session	**1**	**2**	**3**	**4**	**5**	**6**
Population	Student	Mixed	Mixed	Non-student	Student	Non-student
No. of subjects	20	24	24	24	20	16
Date	19/02/09	20/02/09	23/02/09	24/02/09	25/02/09	27/02/09
Situation paid	Auction	Guessing	Trust	Guessing	Auction	Risk
Av. earnings	£7.9	£8.7	£18.5	£8.4	£8.1	£12.2

Table 1: Session characteristics

The situations were presented in the following sequence for all subjects (identified only by their number): 1) Trust game (TG), 2) Guessing game (GG), 3) Dictator game (DG), 4) Second price sealed bid auction, 5) Elicitation of risk preferences, 6) Repeated public good

[4]We separated the experiment into two parts because they were fundamentally different. The first part, experimental games, was administered using z-Tree (Fischbacher, 2007) and different situations involved a combination of one's own decisions, others' decisions and chance. The second part, the cognitive ability test, was administered using a different program (MS Access) and it did not involve strategic interaction nor chance moves. In this latter situation, participants' payoffs depended only on their individual answers.

game. We conducted a single-blind two-person dictator game; a two-person binary trust game; a guessing game involving all the participants in a session; a second-price auction with all the subjects in one session bidding for a single unit of a commodity under a sealed-bid procedure; and a standard four-person ten-period public good game (see Table 2).

For each game, we asked them to describe an example of a possible outcome, immediately after the instructions were read. The goal was to obtain a measure of understanding without priming them in any particular way.

In the second part, students were asked to do a short 12-minute version of an IQ test, consisting of 4 components: Numerical computation (calculation), numerical reasoning (e.g. logical series), abstract reasoning (figures and series) and verbal ability (e.g. analogies). Subjects had a limited amount of time for each component (2/3 minutes), and they received £0.20 for each correct answer.[5]

We also elicit risk preferences in the first part of the experiment (situation 5). Since we expect non-students to be quite heterogeneous in their mathematical ability, we chose a simple strategy method to elicit risk preferences (Dave et al., 2010). They are presented

[5]We selected 26 questions in total from an on-line psychometric test battery: www.psychometric-success.com. We sampled the questions according to their levels of difficulty, including very easy, easy and more difficult questions. The level of difficulty is typically increasing in such tests, so we selected questions from the beginning, the middle and the end, to ensure we could capture well differences in cognitive ability across subjects.

with eight successive choices between a safe amount (£5) and a lottery. The lottery gave them a 50% chance of earning a positive amount (which decreased by £1 in each case from £14 in the first case to £7 in the last case) and a 50% chance earning nothing. We told them that we would select one of the cases randomly to determine the payoff for this situation.

Additionally, we have information on the subjects' age, gender, professional status, ethnicity and cognitive ability.

4 Experimental results

4.1 The sample

Our subject pools differ along important dimensions (table 2). Students obviously all share the same professional status and are homogenous in age. They are younger on average than the non-students. Non-students show much more heterogeneity in age and in professional status. Students have a much higher average cognitive ability score[6], but there is a fair amount of variation both among students and non-students. In terms of risk preferences, measured by the number of lottery choices, we find that non-students tend to be less risk averse on average.

[6]the test score ranges between 0 and 520

	Students			Non-students		
	Mean	Std. Dev.	Min-Max	Mean	Std. Dev.	Min-Max
Age	23	4.7	18-35	34	11	18-70
Cognitive ability	415	71.8	240-520	350	91	120-520
% female	54			63		
Number of lottery choices	2.6	2.5	0-8	3.6	2.9	0-8

Table 2: The sample

Summarizing, our subject pools do look different in important dimensions. Students are younger and smarter. These differences could be an advantage, insofar as they may help students understand better the instructions and the games played. But the differences could also compromise the external validity of the results if they directly affect their ability to reason strategically or are correlated with risk or other-regarding preferences (Anderson et al., 2010).

4.2 Game behaviour across subject pools

We concentrate our analysis here on the Nash equilibrium strategy based on monetary payoffs -the homo-economicus equilibrium strategy, assuming selfishness and rationality and common knowledge of rationality. For the Guessing Game, the Nash equilibrium is that subjects should report 0. But this is not very informative – only one subject does so in our sample. Alternatively, it is common for this game (Camerer et al., 2002) to define equilibrium strategies according to depths of levels of reasoning. Level 0 subjects are non-sophisticated subjects and simply pick a number at random. Level 1 subjects best respond to Level 0 subjects whose random picks will average around 50, thus report a guess equal to 2/3 of 50. Level 2 subjects best respond to Level 1 subjects, thus report a guess equal to $(2/3)^2$ of 50. More generally, the level of depth of reasoning, k, commands a best response equal to $50 * (2/3)^k$. We will consider the probability of reasoning of level 1 or more. The Nash equilibrium strategy for the other games are the following: Dictator Game – donate zero; Trust Game – send zero and send back zero; Public Goods Game – contribute zero. For the second-price sealed bid auction, the equilibrium is defined as bidding the private value or the private value +£0.1 (Kagel and Levin, 1995).

Table **??** reports the percentage of subjects playing the homo-economicus equilibrium strategy, as well as the average decision in each of the five games. In all games with the exception of the auction game, we find that students are more likely to behave as homo-economicus individuals.

Our dictator game results for our student sample are in line with previous literature. Dictator games in which the recipients are anonymous result in average donations between 10 and 15 percent (Hoffman et al., 1994; Hoffman et al., 1996; Eckel et al., 1996). The average donation for our student sample was 16 percent which is similar to these other findings (which, again, are typically based on student subject pools). Fifty-seven percent of our student samples donated nothing which again is consistent with other findings – Eckel et al., 1996, for example, find that 63 percent of their student subject pool

Game	Decision	**EB-NS**	**EB-S**	**p-value**	**D-NS**	**D-S**	**p-value**
Dictator	£0 - £10	17	57	p<0.001	3.5	1.6	p<0.001
Trust - sender	£0 - £10	18	65	p<0.001	8.2	3.5	p<0.001
Trust - receiver	£0 - £15	13	44	p=.005	13.1	8.4	p=0.005
Public Good r1	20 tokens MRC=0.4	9	24	p=0.026	11.1	9.9	p=0.323
Public Good r10	20 tokens MRC=0.4	29	52	p=0.008	5.4	4.8	p=0.151
Guessing	£20 if closest to 2/3 of average (0-100)	32	56	p=0.008	45.9	38.3	p=0.051
Auction	Private vale: £4	29	27	p=0.908	4.8	3.7	p=0.205
Auction	Private vale: £6	46	31	p=0.420	5.1	4.9	p=0.730
Auction	Private vale: £8	13	62	p=0.005	5.6	7.7	p=0.008
Auction	Private vale: £10	18	36	p=0.175	7.3	7.7	p=0.311

Note: p-values obtained from two-sample test of proportions and two-sample ranksum test.

EB-(N)S: % equilibrium behavior of non(students). D-(N)S: actual decision of (non)students.

Table 3: Aggregate behavior in the five games

donate nothing when the recipient is anonymous. Seventeen percent of our non-students donated nothing which is less than a third of the student subject pool result. Duch et al. (2004) similarly find a reluctance of general population samples to make no donation – in their representative sample from Benin only eight percent made this choice. The average donation of our non-students is 35 percent of the maximum; about twice the average donation of the students. This 2-1 non-student-student ratio is roughly the difference in donations that Carpenter et al. (2005) find between Middlebury students and Kansas City workers.

In contrast, for the binary trust game, our student subjects are outliers; only 35 percent of our student trustors trust. In their base-line treatment Berg et al. (1995) find that 90 percent of first movers invest; in a similar trust experiment Ortmann et al. (2000) and Eckel et al. (2004) find that 80 percent of first movers invest. In a binary trust experiment very similar to ours, Casari et al. (2009) report that 73 percent of their student subjects (designated as trustors) trusted by investing. Our non-student participants behavior resemble these results: 82 percent of first movers were trusting. Thus, it could be that the first participation makes students particularly less trusting.

Rates of reciprocation in binary trust games typically are in the range of 50-60 percent. Casari et al. (2009), for example, report a rate of reciprocation of 60 percent for their student subjects. Fifty-six percent of our non-student trustees reciprocate while 87 percent of non-student second-movers reciprocated. In the case of reciprocation we find that our students are roughly consistent with reciprocation results in other studies while our non-student subjects here are very much a positive outlier.

Our third other-regarding game is the public good game – this is the one game that involved repeated play (10 rounds). As table 3 indicates, students are considerably more likely to behave in the homo economicus fashion by free-riding: in the first round of the public goods game, 24 percent of the students contributed nothing to the public good while only 9 percent of non-students were strictly non-cooperative. Gachter et al. (2004) conduct a similar, although one-shot, public good game with students and non-students. Their results suggest, consistent with ours, that non-students are more cooperative than students: non-students contributed an average of 10.2 tokens (out of 20) while students contributed 8.8. The average contribution of our non-students in the first round was 11.1 (out of 20) compared to 9.9 for students; in the last round the average contributions, respectively, were 5.4 and 4.8.[7] On balance the patterns in both studies are similar and provide quite convincing evidence that students are less cooperative than non-student subject pools.

All three of our games involving other-regarding preferences demonstrate, consistent with virtually all experimental literature, that subjects, in general, are more trusting and more cooperative than classic theory would predict. And of course there have been considerable efforts to incorporate this "kindnes" into models of economic behavior (Fehr et al., 1999). The three empirical results presented above leave little doubt that students are dramatically more non-cooperative and less trusting than non-student subjects.

Comparing these other-regarding games in terms of their complexity already suggests that the differences across subject pools are not simply the result of confusion. The dictator game is extremely simple; the trust game is somewhat more complex in that first movers

[7]The differences between their results and ours might be associated with the fact that theirs is a one-shot game and ours is a repeated public goods game. Although one might speculate that this would increase the free-riding in the Gaechter et al. (2004) experiment.

should consider the likely behavior of second movers; and finally the public good game requires more complex strategic calculations related to collective action and free riding. But there is no evidence that non-students are less cooperative in the very simple, as opposed to more complex, other-regarding games.

Our two other games – the Auction and Beauty-contest games – involve no (or little) role for other-regarding preferences, since there is only one person winning the game and the others earn nothing. In these games students are more likely to make the homo-economicus equilibrium choice; but the magnitude of these differences is much smaller than it was for other-regarding games. Table 3 suggests that for the Beauty-contest the ratio of student equilibrium choices to non-students equilibrium choices is roughly 2 to 1 compared to the 3-1 ratio for the other regarding games. And for the second-price sealed bid auction game the ratio is even smaller 1.5 to 1.

Turning to the beauty contest game, the equilibrium strategy here is 0, yet typically the percentage of subjects choosing 0 in experiments is less than 10 percent (Camerer 1997; Nagel 1995) – only one non-student subject in our experiment selected 0. On balance, students exhibit higher levels of iterative reasoning although the differences are less stark than with other-regarding games. Thirty percent of our non-student subjects made choices consistent with at least level 1 iterative reasoning compared to 56 percent for the student subjects. Although in the case of, roughly, level 2 iterative reasoning, students and non-students are more similar (21 and 15 percent, respectively). The average guess of our student subjects is 38 which is consistent with similar guessing games results (in which the average is multiplied by two-thirds). Nagel (1995) finds that the average guess is 35. Camerer (1997) conducted multiple Beauty-contest experiments (with a multiple of .70 applied to the mean) and found the average guesses, for mostly student subject pools, ranged between 31 and 40. Our non-student subject pool exhibited somewhat less iterative rationality with an average guess of 46.

In the second-price sealed bid auction, individuals should bid their private value (Kagel et al., 1993). Thirty-seven percent of our student subjects bid their private value. Again this is consistent with other experimental results; Kagel et al. (1993) find that 37 percent of students bid their private value in a second-price sealed bid auction. Our non-students are somewhat less likely to behave rationally and bid their private value – 26 percent of non-students behave in this fashion. But these differences are not statistically different. And an analysis of average bids for students and non-students suggests that both students and non-students on average have bids that approximate their private value.

Summarizing, the evidence so far suggests that students behave differently both in strategic and non-strategic games, but the differences are smaller in games that are more complex (repeated public good game and second price sealed bid auction).

4.3 Are these differences an artefact of the lab?

Are these populations truly different or is this an artefact of the lab experimental setting?

The first question is whether these differences are due to differences in the level of understanding of the instructions and the games. The results presented earlier showed that the large differences in behavior are observed in the simplest games, such as the dictator game and the trust game, while differences are smaller in the more cognitively demanding games, such as the public good game, guessing game and the auction.

To investigate this further, we constructed a measure of understanding of the instructions. Immediately after reading the instructions for each game, subjects were asked to give an example of possible decisions and of the payoffs that would result from these decisions (see the Appendix for the detailed instructions). For example, for the trust game, they were asked to give an example of (1) the decision of the first player (send 0 or £10), (2) the decision of the second player (send 0 or £15 back to the first player) and the resulting earnings of each player (called 1 and 2).

Table 4 presents the percentage of subjects who understood the instructions for each game, that is, those who provided a correct example for all decisions and payoffs involved. Generally, students do understand better the instructions than non-students, but the difference is larger in games that involve more calculations or are cognitively more demanding (public good game, guessing game). Differences in understanding are unlikely to explain the large differences observed in the simple games, such as the dictator game and the trust game.

Levels of understanding are positively correlated with our measure of cognitive ability

	students	non-students	p-value
dictator game	100	98	.33
trust game	97	89	.10
public goods game	83	66	.02
guessing game	91	68	.09
auction	92	73	.01

Note: p-value obtained from two-sample test of proportions.

Table 4: % who understand the instructions

(0.13, significant at 1% level), negatively correlated with age (-0.14, , significant at 1% level) and positively correlated with gender, women scoring slightly higher (0.06, significant at 10% level).

Table 5 reports the estimated coefficient of the difference between students and non-students if one limits the sample to those who understood the instructions. We do not find large changes in the estimated difference between students and non-students if we limit the sample to those who understood.

We have additional evidence of the role of comprehension by looking at learning effects in the repeated public good game. If the student/non-student differences for the other-regarding games were the result of comprehension we might expect these differences to erode with repeated play. Our repeated play of the public goods game suggests that both students and non-students become more homo-economicus over the repeated play but

	Student	dummy
	(1)	(2)
Dictator game	.40	.43
	(.11)***	(.11)***
Trust game - sender	.46	.44
	(.11)***	(.12)***
Trust game - receiver	.31	.33
	(.11)***	(.11)***
Public goods game	.15	.20
(round 1)	(.06)**	(.06)***
Guessing game	.23	.28
	(.09)***	(.10)***
Auction	.12	.15
	(.08)	(.09)*
Sample of those who understand only	No	Yes

Note: *** denotes significance at the 1% level, ** at the 5% level, and * at the 10%.

Table 5: Understanding of the instructions and equilibrium behavior - Differences between students and non students

differences in cooperative behavior persist. Our student results closely tract the results of Andreoni (1988; 1995) who conducted similar 10-round public goods games with student subject pools: In the first round of his PGG experiment, 20 percent contribute nothing and in round 10, 45 percent contribute nothing Andreoni (1995). Figure 1 compares the non-cooperation in the Andreoni (1988; 1995) experiments with the evolution of non-cooperation by students and non-students in our experiment.

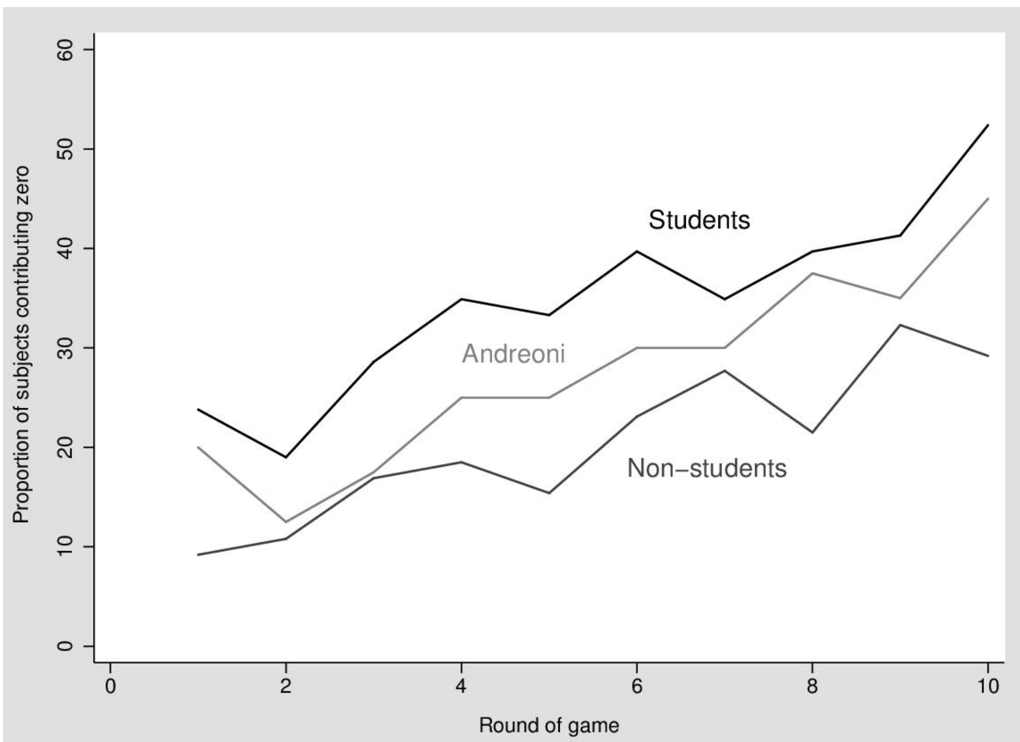

Figure 1: Percentage of Subjects Contributing Nothing by Round in Repeated Public Goods Game

Table 6 reports the results for a simple test for differences in learning behavior. We regress the proportion of subjects contributing zero on the student subject dummy, number of rounds, and the interaction of the student dummy and number of rounds. There is a

	Probability of contributing 0 (probit estimates marginal effects)
student	.158
	(.080)**
Number of rounds	.028
	(.007)***
Student x Number of rounds	-.002
	(.008)

Note: *** denotes significance at the 1% level, ** at the 5% level, and * at the 10%.

Table 6: Learning

clear positive trend: On average the percentage of subjects contributing zero increases by 3 percentage points in each round. But we fail to find any evidence of differences in trends between the two subject pools. Rates of learning by students and non-students are similar suggesting that both subject pools respond similarly, over iterative play, to the non-cooperative incentives of the game.

The results presented here suggest that the differences between students and non-students cannot be explained by differences in comprehension and may instead be driven by differences in fundamentals driving behavior, such as their ability to reason strategically and their preferences. The large differences observed in the simple other-regarding games suggest that non-students are more pro-social than students. And the difference observed in strategic games that are more cognitively challenging suggest a role for the ability to reason strategically, although the differences between students and non-students are smaller once the game becomes too cognitively demanding, such as in the public good game and the second price sealed bid auction.

To shed more light on the differences in fundamentals driving behavior, we now turn to the role of risk attitudes. There are two games where risk possibly plays a role as well - the trust game and the public good game. Students are on average less risk averse, but the difference is small and not statistically significant. In principle, risk aversion could explain lower levels of trust and lower contributions in the public good game. We include our measure of risk aversion as a control variable. Note that the measure is noisy and shows a high degree of inconsistency in responses to the risk items given the simplicity of the question: About a third of the subjects (in both pools) switch back and forth (at least once) between the 'safe' and the 'risky' alternative.[8] With this caveat in mind, we restrict our sample to those who gave consistent answers and pool the decisions from both games. For the sake of brevity, we do not report the results in a table.[9] The estimated student dummy hardly changes when controlling for risk aversion, and the measure of risk aversion does not have a significant effect on behavior. Thus, differences in risk aversion - as measured in this standard and simple way - do not help explaining differences in

[8]Non-unique switching points is not uncommon in these types of experiments. Holt et al (2002) elicit risk preferences in a similar fashion and report 19.9 percent of the subjects exhibited non-unique switching behavior.

[9]The results are available upon request from the authors.

behavior in the games involving risk.

Since the two populations are quite different from each other in many dimensions, it is perhaps not so surprising that they differ in fundamentals driving behavior. The student sample has better cognitive skills, is younger and are by definition students while the others are not. Our goal here is not to identify which demographics or socio-economic characteristics matter most to explain these differences. Our sample is too small to study this in detail. But we do have information on cognitive ability, age and gender, and can at least investigate to what extent these characteristics are correlated with behavior and help explaining differences. There is a body of empirical literature suggesting that IQ, age and gender are correlated with behavior in a number of economic games. Burks et al. (2009) find that higher IQ increases the probability of playing the non-cooperative strategy in a Prisoner's dilemma game. Sutter et al (2007) find that trust increases almost linearly with age from early childhood to early adulthood, but stays rather constant between different adult age groups. A large number of experimental studies in economics and psychology have documented gender differences in preferences (Croson and Gneezy, 2009).

Table 7 presents the results including individual covariates. We find that controlling for our measure of cognitive ability does reduce slightly differences between students and non-students; in particular in the trust game and the guessing game. Thus, cognitive ability does seem to matter for understanding the game and the strategic components of the game. Regarding age, we find that older people donate more in the dictator game, suggesting more pro-social preferences. In the guessing game, age explains a large part of the difference between students and non-students. Older subjects tend to report higher guesses. One possibility is that age captures a form of ability, not captured by our measure of cognitive ability. This hypothesis seems to be supported by the fact that we do have a strong negative correlation (-.48), significant at the 1% level, between cognitive ability and age. Thus, older people may be less able to reason strategically.

These results imply that students may well be a convenient sample, but they are quite atypical. The last question we ask is whether the laboratory itself differentially triggers other-regardedness in non-students versus students. Levitt et al (2007) suggest that the "nature and extent of scrutiny associated with the lab" might induce subjects to conform to social norms against wealth maximizing choices (Levitt et al, 2007). Students might be more comfortable than non-students with the scrutiny associated with the lab environment. The experience of a laboratory experiment may also be much more out of the ordinary for a non-student, and affect their behavior in a systematic way. If that is the case, their behavior in the lab may be more externally valid than the behavior of non-students.

We have an indirect measure of whether our student/non-student differences in other-regardedness are replicated outside of the laboratory environment. Two weeks after the experiment, subjects received a thank-you e-mail that also asked them how they had been informed about the lab. This requests a form of cooperation that is only beneficial to us; is not personal in any way; and in principle involves little cost. This test enables us to obtain a different measure of other-regarding preferences outside the lab and in a natural

Probit estimates			
(marginal effects)			
	(1)	(2)	(3)
Dictator game	.41	.31	.31
	(.12)***	(.16)*	(.16)*
Trust game - sender	.50	.55	.55
	(.12)***	(.14)***	(.14)***
Trust game - receiver	.22	.38	.38
	(.12)*	(.14)**	(.14)***
Public good game	.10	.10	.10
(round 1)	(.07)	(.08)	(.08)
Guessing game	.19	.06	.06
	(.09)**	(.11)	(.12)
Auction	.11	.21	.21
	(.09)	(.10)*	(.10)*
Control for IQ	Yes	Yes	Yes
Control for age	No	Yes	Yes
Control for gender	No	No	Yes

Note: *** denotes significance at the 1% level, ** at the 5% level, and * at the 10%.

Table 7: Demographics and behavior - Probability of homo-economicus decision Differences between students and non students (Student dummy)

context (this kind of e-mail is relatively common practice and there is no reason to expect differences in familiarity between students and non-students). Of course, there could be systematic reasons why students may be more or less likely to respond, other than pro-social preferences. But our prediction would be that these other reasons (opportunity cost of time, frequency of computer usage and checking of e-mails) would increase the probability they will respond rather than decrease it. We find systematic differences in the response rates between students and non-students. 53% of the students replied to the e-mail against 72% of the non-students (this difference is significant at the 2% level). Thus, the student/non-student differences we reported earlier do not seem to be an artefact of the laboratory setting. Overall, we find that our student population tends to be more pro-social than the non-student population.

5 Discussion and conclusion

Our results suggest that in experiments that invoke other-regarding considerations, student and non-students subject pools will make choices that are very different - students are more likely to behave as a homo-economicus agent than is the case for non-students. The student subject pool was dramatically less kind and less trusting than the non-student subjects. Fairly consistently across experiments, non-students were more other regarding by a ratio of three-to-one. Our results suggest these differences were not the result of features of the experimental design such as comprehension or the unnatural nature of the lab experiment context. Rather we are inclined to conclude that non-students are more inequity averse and trusting.

Student and non-student subject pools can differ quite significantly in terms of characteristics that likely matter for the decisions they take in a typical economic experiment, such as cognitive skills and age for example. They are better able to reason strategically and they are less pro-social.

To the extent that experimenters are concerned with identifying subjects with motivations conforming to the classic assumptions of game theoretic models, they should focus on student subject pools and avoid non-students. This is particularly the case with games that incorporate other regarding features, such as trust or dictator games. Experiments using students are likely to provide a lower bound estimate of other-regardedness in the general population. In the case of games in which cognitive skills are critical to the internal validity of a particular experiment, the results are more nuanced and the student/non-student differences are not as large. Still, we find evidence of differences in cognitive skills that may result in differences in the ability to reason strategically.

If in contrast researchers are concerned by the external validity of the results, using a more heterogenous population may be desirable. If a population differs in fundamentals driving behavior such as preferences and the ability to reason strategically, then most economic models would predict that treatment effects - changes in behavior in response to changes in the incentive structure - will differ as well. Our results show that important socio-demographic characteristics (being a student, cognitive ability, IQ, age, etc.) are correlated with important fundamentals driving behavior. In that context, using a population that is more heterogeneous along these dimensions may be desirable to obtain externally valid results, but of course this will come at the cost of statistical power and will require larger samples to identify treatment effects.

References

Anderson, J., Burks, S., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D., Potter, R., Rocha,K., Rustichini, A. (2010). 'Self Selection Does Not Increase Other-Regarding Preferences among Adult Laboratory Subjects, but Student Subjects May Be More Self-Regarding than Adults', IZA Discussion Paper No. 5389.

Andreoni, J. (1988). 'Why Free Ride? Strategies and Learning in Public Good Experiments', *Journal of Public Economics*, vol. 37, pp. 291-304.

Andreoni, J. (1995). 'Cooperation in Public Goods Experiments: Kindness or Confusion', *American Economic Review*, vol. 85, pp. 891-904.

Berg, J., Dickhaut, J. and McCabe, K. (1995). 'Trust, Reciprocity, and Social-History', *Games and Economic Behavior*, vol. 10, pp. 122-142.

Bosch-Domenech, A., Montalvo, J., Nagel, R. and Satorra, A. (2002). 'One, Two, (Three), Infinity...: Newspaper and Lab-Beauty Experiments', *American Economic Review*, vol. 92, pp. 1687-1701.

Burks, S., Carpenter, J. Götte, L. and Rustichini, A. (2009). 'Cognitive Skills Affect Economic Preferences, Social Awareness, and Job Attachment', *Proceedings of the National Academy of Science*, vol. 106(19), pp. 7745-7750.

Camerer, C. (1997). 'Progress in Behavioral Game Theory', *The Journal of Economic Perspectives*, vol. 11, pp. 167-188.

Camerer, C., Ho, T. and Chong, J. (2002). 'Sophisticated EWA Learning and Strategic Teaching in Repeated Games', *Journal of Economic Theory*, vol. 88, pp. 104-137.

Cappelen, A., Nygaard, K., Sorensen, E. and Tungodden, B. (2011). 'Social Preferences in the Lab: A Comparison of Students and a Representative Population', CESifo Working Paper Series No. 3511.

Carpenter, J., Connolly, C. and Myers, C. (2008). 'Altruistic Behavior in a Representative Dictator Experiment', *Experimental Economics*, vol. 11, pp. 282-298.

Carpenter, J., Burks, S., Verhoogen, E. (2005). 'Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games', *Research in Experimental Economics*, vol. 10, pp. 261-290.

Casari, M., Ham, J. and Kagel, J. (2007). 'Selection Bias, Demographic Effects, and

Ability Effects in Common Value Auction Experiments', *American Economic Review*, vol. 97, pp. 1278-1304.

Croson, R. and Gneezy, U. (2009). 'Gender Di erences in Preferences', *Journal of Economic Literature*, vol. 47, pp. 448-74.

Danielson, A. J. and Holm H. J. (2007).'Do You Trust Your Brethen? Eliciting Trust Attitudes and Trust Behavior in a Tanzanian Congregation', *Journal of Economic Behavior and Organization*, vol. 62, pp. 255-271.

Dave, C., Eckel, C., Johnson, C. and Rojas, C. (2010). 'Eliciting risk preferences: When is simple better?', *Journal of Risk and Uncertainty*, vol. 41(3), pp. 219-243.

Depositario, D., Nayga, R., Wu, X. and Laude, T. (2009). 'Should Students be used as Subjects in Experimental Auctions', *Economics Letters*, vol. 102, pp. 122-24.

Duch, R. and Palmer, H. (2004). 'Its Not Whether You Win or Lose, but How You Play the Game: Self-Interest, Social Justice, and Mass Attitudes toward Market Transition', *American Political Science Review*, vol. 98, pp. 437-452.

Eckel, C. and Grossman, P. (1996). 'Altruism in Anonymous Dicatator Games', *Games and Economic Behavior*, vol. 16, pp. 181-191.

Eckel, C. and Wilson, R. (2004). 'Is Trust a Risk Decision?', *Journal of Economic Behavior and Organization*, vol. 55, pp. 447-466.

Falk, A., Meier, S. and Zehnder, C. (2012) 'Do Lab Experiments Misrepresent Social Preferences? The case of self-selected student samples, forthcoming in Journal of the European Economic Association.

Fehr, E. and Schmidt, K. (1999). 'A Theory of Fairness, Competition and Cooperation', *Quarterly Journal of Economics*, vol. 114, pp. 817-68.

Fischbacher, U. (2007). 'z-Tree: Zurich Toolbox for Ready-made Economic Experiments', *Experimental Economics*, vol. 10, pp. 171-78.

Gachter, S., Herrmann, B. and Thoni, C. (2004). Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence', *Journal of Economic Behavior and Organization*, vol. 55, pp. 505-531.

Hoffman, E., McCabe, K. and Smith, V. (1994). 'Preferences, Property Rights, and Anonymity in Bargaining Games', *Games and Economic Behavior*, vol. 7, pp. 346-380.

Hoffman, E., McCabe, K. and Smith, V. (1996). 'Social Distance and Other-Regarding Behavior in Dicatator Games', *American Economic Review*, vol. 86, pp. 653-60.

Holt, C. and Laury, S. (2002). 'Risk Aversion and Incentive Effects', *American Economic Review*, vol. 92, pp. 1644-1655.

Kagel, J. and Levin, D. (1993). 'Independent Private Value Auctions: Bidder Behavior in First-, Second- and Third-Price Value Auctions with Varying Numbers of Bidders', *The*

Economic Journal, vol. 103, pp. 868-79.

Levitt, S. and List, J. (2007). 'What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?', *Journal of Economic Perspectives*, vol. 21, pp. 153-74.

Morton, R. and Williams, K. (2009). *From Nature to the Lab: Experimental Political Science and the Study of Causality*, New York: Cambridge University Press.

Nagel, R. (1995). 'Unraveling in Guessing Games: An Experimental Study', *American Economic Review*, vol. 85, pp. 1313-1326.

Ortmann, A., Fitzgerald, J. and Boeing, C. (2000). 'Trust, Reciprocity, and Social History: A Re-examination', *Experimental Economics*, vol. 3, pp. 81-100.

Peterson, R. A. (2001). 'On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis', *Journal of Consumer Research*, vol. 28, pp. 450 461.

Sutter, M. and Kocher, M. (2007). 'Trust and Trustworthiness across Different Age Groups', *Games and Economic Behavior*, vol. 59, pp. 364-82.

APPENDIX – EXPERIMENT INSTRUCTIONS

Dear participants,

Welcome and thank you for participating to our experiment. The experiment will last for about one and a half hour. Please do remain quiet from now on until the end of the experiment. You will have the opportunity to ask questions in a few minutes.

The experiment consists of two main parts. Throughout the whole experiment, you will receive instructions on the computer screen and be told what to do.

<u>PART I</u>

In this first part, you will be asked to make choices in six different situations. Each of these six is an entirely different and independent situation. You should treat each as being independent of the others. We will describe each situation successively and explain precisely what you have to do.

In each situation your earnings will depend on your choices, possibly the choices of others and chance. To make sure you understand what you have to do, you will be asked to think of an example for each situation before you take the actual decision. This is only to make sure you understood, and these examples have no implications at all for your earnings. When each situation is over, there will be a very short pause and then we will introduce the next situation. You will only be informed about the outcomes of all six situations **at the end of PART I**. We will show you a screen with the outcomes corresponding to each of the six situations.

The computer will then "roll a die" to pick one of the six situations, which will determine your actual earnings. That is, **your actual earnings will depend on the choices you made in <u>one</u> of the six situations only** – this single situation will be selected at random by the computer for all participants in this room.

<u>PART II</u>

In this second part, you will be asked to answer questions and your earnings will depend on your answers (your earnings in this part will depend only on <u>your</u> answers). Again, you will be informed about the outcome of this part only **at the end of PART II**.

We ask you to remain quiet during the whole experiment. Those who do not respect the silence requirement will be asked to leave the experimental room. Once the experiment is finished, please remain seated. You will be called up successively by the number on your table; you will then receive an envelope with your earnings and you will be asked to sign a receipt.

Important note: The CESS lab has a **strict "no deception" policy**. That means that **under no circumstances** will participants to experiments be deceived. All the information you will receive from us is true. For example, if we tell you that you have been paired to another participant in the room, this is indeed the case.

Finally, note that your participation is considered voluntary and you are free to leave the room at any point if you wish to do so. In that case, we will only pay you the participation fee of £4.

Please leave these instructions on your table when you leave the room.
You can take notes on these pages if you wish to do so.

If you have any questions, please raise your hand now.

PART I

SITUATION 1

We have formed pairs of people at random in the room. Each pair is made up of **person 1** and **person 2**. None of you will know with whom you have been paired. The experimenter is the only one who knows who is paired with whom.

Person 1 will receive £10 and then can choose whether to transfer these £10 to **person 2** or not to transfer any money and keep the £10.

- If **person 1** decides to keep the £10, **person 2** earns nothing.
- If **person 1** decides to transfer the £10 to **person 2**, the money will be tripled by us before it is passed on to **person 2**. That is, **person 2** will receive £30.

Person 2 will be able to choose whether to transfer £15 back to **person 1** or not to transfer anything back in the event her/his account is credited with £30.

- If **person 2** transfers £15 back, both **person 1 and person 2** will earn £15.
- If **person 2** does not transfer £15 back, **person 2** will earn £30 and **person 1** will earn nothing.

To make sure you understand the instructions, could you please provide an example (of your choice) of the possible decisions person 1 and 2 could make and the pay-offs they would get:

Decision of Person 1: ______________________________

Decision of Person 2: ______________________________

Earnings of Person 1: £ _____

Earnings of Person 2: £ _____

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return to the computer and follow the next instructions on the screen.

SITUATION 2

Each of you has to guess a number between 0 and 100. We will calculate the average of the guesses of all participants in the room and multiply this number by two-thirds. The person whose guess is closest to this number will receive £20 (in case of a tie, the computer will select randomly one of the winners and he/she will receive £20).

To make sure you understand the set-up, could you please provide an <u>example</u> of a possible outcome for this situation:

Average of all guesses: ____

The winner would be the person with the guess closest to what number? ____

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return to the computer and follow the next instructions on the screen.

SITUATION 3

You have been paired randomly with another person in the room. Each pair is made up of **person 1** and **person 2**. None of you will know with whom you have been paired. The experimenter is the only one who knows who is paired with whom. **Person 1** will receive £10 and will have the opportunity to give a portion of her (his) £10 to **person 2**. **Person 1** can transfer any (round) amount between 0 and £10.

____

To make sure you understand the set-up, could you please provide an <u>example</u> (of your choice) of a possible outcome for this situation:

Decision Person 1: Transfer £ ___ to person 2

Earnings Person 1: ______

Earnings Person 2: ______

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return to the computer and follow the next instructions on the screen.

SITUATION 4

You will now have the opportunity to bid for an amount of money. There are four possible amounts of money you could bid for: £4, £6, £8 or £10. The computer will determine at random for each participant which of these four amounts you could bid for. This will appear on the computer screen. Thus, some of the participants will be able to bid for £4, some for £6, some for £8 and some for £10. All participants in this room bid <u>in the same auction</u>.

You will be asked to introduce a bid. The person with the highest bid (among *all* participants in this room) will win the amount of money that has been allocated to that person. Bids will remain private and will only be known to the experimenter.

Important: The winner will have to pay the amount of money corresponding to the bid of the **second highest bidder**. In case of a tie, the computer will select randomly one of the highest bidders and this will determine who has won the auction.

Thus, if you are the highest bidder of the auction; you will earn:

Earnings	=	Your Prize (£4, £6, £8 or £10)	Second highest bid

If you are not the highest bidder you will not win – nor will you have to pay anything.

Note that the laboratory has a "no bankruptcy policy", thus, bids that would result in negative earnings beyond the £4 show-up fee will not be permitted.

To make sure you understand the instructions, could you please provide an example of a possible outcome for this situation? Suppose there were four participants who could win respectively 4, 6, 8 and 10. What possible bids could be submitted?

Bid submitted by person 1 (for £4): ______

Bid submitted by person 2 (for £6): ______

Bid submitted by person 3 (for £8): ______

Bid submitted by person 4 (for £10): ______

Who would be winning the auction? Person __

Earnings of the person winning the auction: £ ______ - £ _______ = £ ________
(Prize) (second highest bid)

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return to the computer and follow the next instructions on the screen

SITUATION 5

We will now propose eight different choices between a fixed amount of money and an all-or-nothing lottery. The lottery will work as follows. The computer will roll a six-sided die. The general rule is that you will earn nothing if the die indicates 1, 2 or 3; and earn something (as indicated for each case) if the die indicates 4, 5 or 6. Finally the computer will pick at random a number between 1 and 8 to determine which of 8 cases applies to determine your earnings in this situation.

You will be asked to indicate your preferred option in each of the cases.

For example:

Case 1: Choice between

A: £5 with certainty or
B: £ 0 if the die shows 1, 2 or 3; £14 if the die shows 4, 5 or 6.

To make sure you understand the set-up, could you please provide an <u>example</u> (of a possible outcome in this case:

Your choice (A or B): _______

Outcome of the lottery (die: 1,2,3,4,5 or 6): _______

Your earnings: £ _______

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return now to the computer and follow the next instructions on the screen

SITUATION 6

We have formed groups of four people at random in the room. You have been assigned to one of these groups. None of you will know who is in his/her group. The experimenter is the only one who knows who is in which group.

Each member of the group has to decide on the division of 20 tokens. You can put these 20 tokens on your private account or you can invest them fully or partially into a project. Each token you do *not* invest will automatically be transferred to your private account.

Your income from the private account:
For each token you put on your private account, you will receive exactly one point. Nobody except you earns something from your private account.

Your income from the project
We will add up the contributions made by the four members of your group to the project. Each member will then receive an income from the project calculated as follows:

Income from the project = Sum of all contributions of the members of the group $\times$ 0.4

Your total income

Your total income is the sum of your income from the private account and your income from the project.

We will ask you to make such a decision 10 times in a row. We will add up the points you have earned in each period and calculated your final earnings using the following exchange rate: £1 = 25 points.

Note that your group will remain the same throughout this situation.

To make sure you understand the set-up, could you please provide an <u>example</u> (of your choice) of a possible outcome for this situation:

Your contribution to the project: _______ tokens

Sum of all contributions to the project (including yours): _______ tokens

Your income from the private account: _______ points

Your income from the project: _______ points

Your total income: _______ points

If you have any questions about the set-up, please raise your hand now and wait for the experimenter to come to you.

Please return to the computer and follow the next instructions on the screen.

<u>PART II</u>

We will now ask you 4 sets of multiple-choice questions. Each set has to be completed within a certain amount of time. You can skip questions and come back to them later within the same set. Once a set is completed, it is not possible to change or come back to the answers you have submitted. You will earn 20p per good answer submitted.