Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Norms of Punishment in the General Population

Stefania Bortolotti
Marco Casari
Francesca Pancotto
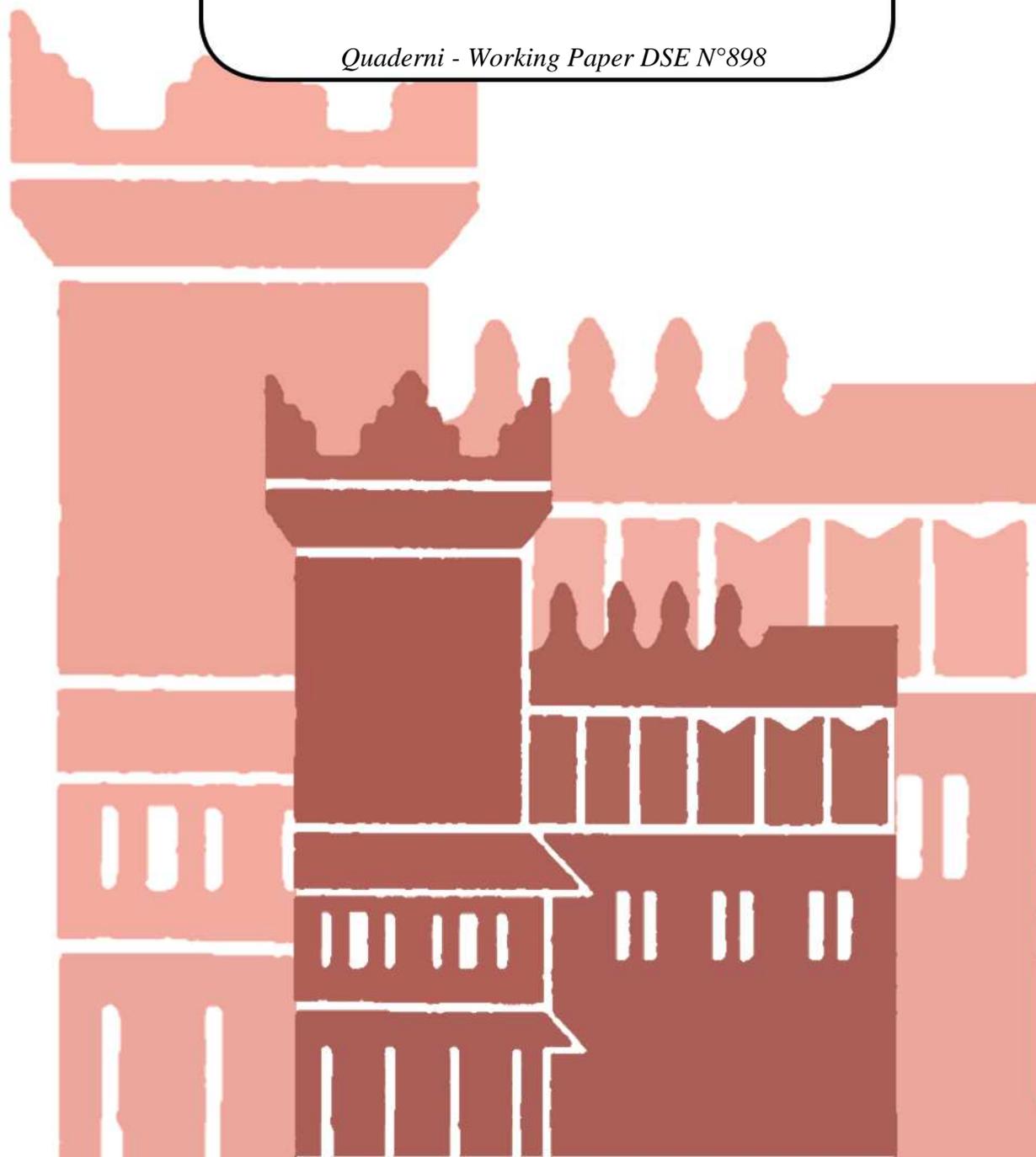
Norms of Punishment in the General Population

Stefania Bortolotti[*]
University of Bologna

Marco Casari[†]
University of Bologna

Francesca Pancotto[‡]
University of Modena and Reggio Emilia

August 6, 2013

Abstract

Norms of cooperation and punishment differ across societies, but also within a single society. In an experiment with two subject pools sharing the same geographical and cultural origins, we show that opportunities for peer punishment increase cooperation among students but not in the general population. In previous studies, punishment magnified the differences across societies in peoples ability to cooperate. Here, punishment reversed the order: with punishment, students cooperate more than the general population while they cooperate less without it. Our results obtained with students cannot be readily generalized to the society at large.

Keywords: Public goods game · representative sample · experiments · external validity

JEL classification: C72 · C90 · Z13

1 Introduction

Peer punishment of free riders has been identified as an effective mechanism to sustain cooperation (Ostrom et al., 1992). There is, however, a wide variation in punishment norms across societies: punishment opportunities enable some societies to overcome collective action problems, and lead other societies into feuds and revenge that harm cooperation (Herrmann et al., 2008; Henrich et al., 2010). Here we show experimentally that, even within the same culture, punishment has a beneficial or a detrimental effect on cooperation depending on the sample of

---

[*]Bortolotti: Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; stefania.bortolotti@unibo.it, Phone: +39051 209 8135.

[†]Casari (corresponding author): Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; marco.casari@unibo.it, Phone: +39051 209 8662, Fax: +39051 209 8493.

[‡]Pancotto: University of Modena and Reggio Emilia, Viale Allegri, 9, 40121, Italy; francesca.pancotto@unimore.it, Phone: +39 0522 523 264.

the population involved.

Consider, for instance, a society of young and old citizens. Cooperation and punishment behaviors can develop in different ways for two reasons: first, different group-specific norms are followed by young and old citizens; second, the same individuals may behave differently when facing only people from the same age group or when interacting in a mixed group. For instance, youngsters may follow one norm when interacting among peers, but change it when they relate with elderly people. When deciding whether to punish or not, a young person may have no hesitations if the target is another young person (i.e., in-group), but she may refrain from punishing an old person (i.e., out-group). Therefore, the propagation of group-specific norms in a society depends both on the relative size of each group and on the interaction of in- and out-group norms.

For the above reasons, one may wonder whether experiments conducted with college students to study other-regarding preference or social norms are informative about the society at large. To this end, we have measured cooperation levels in a task with and without opportunities for peer punishment with two distinct subject pools originating from the same geographical area. One sample was drawn from the student population of a large, public University (*Student* treatment). The other sample was drawn from the general adult population (*Representative* treatment) and stratified according to three categories – gender, age, and employment status – and quotas were defined so to reproduce the composition of the national population.

Economic experiments conducted with a representative sample of the population are rare. Recruiting such samples is indeed a hard task because of logistic and technical issues. In addition, payments must be higher to compensate participants' opportunity costs. Table 1 summarizes experiments comparing a student sample and a sample of the general population.[1]

We contribute to the present literature by ensuring high methodological standards and comparability across participant pools. Following Harrison et al. (2002); Bellemare and Kroger

---

[1]Papers comparing students and non-students with the aim of controlling for self-selection in the participation to experiments are beyond the scope of this paper (for a review, see Exadaktylos et al., 2013). We also exclude from this review studies comparing students and a specific sub-sample of the population (i.e., CEO, elderly, prisoners).

Table 1: Studies comparing college students with the general population

	Stratified sample	Same procedures for both samples	Lotteries	Dictator	Ultimatum	PGG	PGG w/pun	Trust Game	Beauty Contest	Country
TWO DISTINCT SAMPLES										
This study	**Y**	**Y(lab)**	**1**			**R**	**R**			**Italy**
Cappelen et al. (2010)	Y	Y(lab)		1				R		Norway
Bellemare and Kroger (2007)	Y	N						1		Netherlands
Falk et al. (2012)	Y	Y(mail)						1		Switzerland
Carpenter et al. (2008)	N*	N		1						US
Gächter et al. (2003)	N*	N					1			Russia
Gächter et al. (2004)	N*	N				1				Russia
Belot et al. (2010)	N	Y(lab)	1	1		R		1	1	UK
Bosch-Domenech et al. (2002)	N	N							1	Germany/US/Spain
ONLY ONE SAMPLE										
Harrison et al. (2002)	Y	Interview	1							Denmark
Ermisch et al. (2009)	Y	Interview						1		Britain
Exadaktylos et al. (2013)	N*	Interview		1	1			1		Spain
Bellemare et al. (2008)	N*	Internet		1	1					Netherlands
Bellemare et al. (2011)	N*	Internet			1					Netherlands
Dohmen et al. (2008)	N*	Interview						1		Germany
Egas and Riedl (2008)	N*	Internet				R	R			Netherlands

Notes: We consider a sample to be *stratified* if it has been selected according to pre-specified categories and target quotas. N^* indicates a representative sample that has not been selected ex-ante according to target quotas. In the cells relative to each task, 1 indicates a one-shot game and R a repeated game.

(2007); Falk et al. (2012); Cappelen et al. (2010), we specified ex-ante stratification variables and quotas. To the best of our knowledge, this study, along with Cappelen et al. (2010), is the only experiment conducted in the laboratory to compare a student sample with a stratified sample. We conducted the experiment by following the same procedures for both students and the general population. Unlike most of the previous comparisons that focused on one-shot experiments, we collected repeated measures of cooperation –with and without punishment– to investigate whether differences in contribution norms evolve over time or remain stable. Finally, our study, together with Falk et al. (2012), restricts participation of the student and representative samples to those subjects resident in a given region to increase comparability across samples.

We report three main findings. First, without punishment, the general population cooperated more than college students. Second, the introduction of the opportunity to punish increased cooperation among college students but not in the general population. Third, this result did not stem from lack of punishment as the general population sample punished more than the student sample. Punishment did not promote cooperation among the general population because it was frequently directed toward cooperators rather than free riders. If one was to classify the Italian

society according to the impact of peer punishment, opposite conclusions would then be drawn depending on whether the experiment was run with a representative sample or with college students.

The remainder of the paper is organized as follows. Section 2 describes the characteristics of the subject pools, the experimental tasks, and procedures. Section 3 presents the main results on cooperation and punishment. Finally, Section 4 includes a discussion of the results and conclusions.

2 Participants and design

The experiment comprises two treatments –*Representative* and *Student*– that vary only according to the composition of the participant pool. All participants, regardless of the treatment, were born within the region Emilia-Romagna (Italy). This information was common knowledge and could help subjects to form more accurate expectations about norms and others' behavior. This present restriction was explicitly stated during the recruitment process and publicly announced by the experimenter at the beginning of each session.

The *Representative* sample was recruited among the general adult population by two professional companies, both unaware of the goal of the research. The companies contacted people by phone –through telephone directories and private databases– or through a local recruiter. Recruiters were provided with a script to approach potential participants.[2] To be eligible, subjects had to: (a) be at least 18 years old; (b) be born within the province of Ravenna;[3] and (c) be resident within the province of Ravenna. The sample was stratified according to age (18-39, 40-59, 60 or older), sex, and employment status (employed, housewives or retired, others – including students and unemployed). The target quotas for each category were defined according to the composition of the Italian population.[4] To favor a wide participation, subjects received a 30 Euros gasoline vouchers as show-up fee in addition to the earnings gained through sessions.

---

[2] For a detailed description of the recruitment process, see Appendix B.

[3] Ravenna is one of the eight provinces of Emilia-Romagna .

[4] These data were collected as part of a wider research project to investigate social norms across various locations in Italy, where Ravenna was selected as one of the provinces of interest. For sample stratification, we referred to the figures of the National Institute of Statistics concerning inhabitants in January 1st, 2009 (source Istat: http://demo.istat.it/pop2009/index1.html).

The *Student* sample was recruited among the students of the University of Bologna. The University of Bologna has around 90,000 students with campuses in four of the eight provinces of the region Emilia-Romagna. Only students that were born in Emilia-Romagna were invited and could take part to the study.[5] Invitations were sent to subjects present in the ORSEE (Greiner, 2004) database of the Bologna Laboratory for Social Sciences (BLESS) at the time of the experiment.[6] This sample taps into a standard participant pool of college students, which is roughly balanced between humanities and science majors and with relatively few students of the faculty of Economics and Business, and a limited exposure to other economic experiments.[7]

Table 2 reports the socio-demographic characteristics of the two samples. Whereas the gender composition is similar, age and employment compositions differ widely. In the *Student* sample the vast majority of participants is aged between 18-39, while in the *Representative* sample most participants belong to the 40-59 category (44.7 percent) and the remaining subjects are equally distributed across 18-39 and 60 or above. Among the *Representative* sample, half are employed and about 13 percent are students. The overwhelming majority of participants self-reported in the questionnaire to be at least second generation natives of the region. Participants share deep rooted local norms: as a matter of fact, about 87 (84) percent of the participants in the *Representative* sample (*Student* sample) have one or both parents who were born in the region.[8]

Each session included a series of repeated Public Goods Games with and without punish-

---

[5]Because of the limited number of students that were born within the province of Ravenna and present in the ORSEE database, we decided to include among the potential participants subjects that were born in all the provinces of Emilia-Romagna, all of them shared similar socio-economic characteristics. As pointed out by Harrison and List (2004), there are at least two factors, which may restrict the generalizability of lab results obtained with students: *(i)* there is an endogenous sample selection among students participating to experiments; *(ii)* students are not informative about the general population. Since we are mainly interested in *(ii)*, we do not take any additional precaution to limit endogenous sample selection among students. In the same spirit, we did not exclude from the database the small proportion of non-students that used to take part to experiments; we however retain the term 'Student' for brevity.

[6]Since we are mainly interested in assessing to what extent results obtained with a standard participant pool can be extended to the general public, we opted for a group of participants as similar as possible to the one commonly involved in standard lab experiments. To this end, we sampled participants from the ORSEE database rather than from the general college population of the University of Bologna.

[7]The database includes a small fraction of non-students, most of them were former students living in the area. We had 18 non-student participants (17 percent). Thanks to a questionnaire we know that 14 were 32 years old or younger and that 9 hold a college degree. About 1/3 of them hold a college degree and were looking for their first job.

[8]The figures for the representative sample refers to the province of Ravenna.

Table 2: Socio-demographic characteristics of the two samples

	Representative sample	Student sample
Male	51.5%	55.8%
Age		
18-39	24.3%	95.2%
40-59	44.7%	4.8%
60 or above	31.1%	0.0%
Employment Status		
Employed	47.6%	8.6%
Unemployed	10.7%	7.7%
Students	13.6%	82.7%
Housewife or retired	28.2%	1.0%
Education Level		
8th grade or lower	18.5%	1.0%
High school	47.5%	55.8%
College, Master, or Phd	34.0%	43.3%
Rootedness		
Elementary school in the region (county)	86.4%	97.1%
Mother born in the region (county)	69.9%	72.1%
Father born in the region (county)	63.1%	70.2%
Sessions		
	02/03/2011	23/02/2011
	04/03/2011	24/03/2011
Dates (dd/mm/yyyy)	05/03/2011	24/03/2011
	01/10/2011	16/06/2012
N. of participants	108	104

Notes: Self-reported answers from a post-experimental computerized questionnaire. Due to a software failure, questionnaire answers for one *Representative* session (02/03/2012) were collected on the phone a few weeks after the session. Five participants did not answer the phone; as a result, for the representative sample, questionnaire data are available for 103 out of 108 subjects.

ment.[9] PGG-Standard and PGG-Punishment were repeated for eight periods each. Before each period, participants were divided into groups of $N = 4$ under a strangers matching protocol. Interaction was anonymous and there was no possibility of building an individual reputation: a subject could not identify whether the same participant was in his/her group in the following periods, hence ruling out the possibility of building a reputation.

In the PGG-Standard, each subject received an endowment of $w_i = 20$ tokens and had to

---

[9]Each session included a total of five parts presented in a fixed order: 1) choice over lotteries; 2) PGG-Standard; 3) PGG-Punishment; 4) PGG-Standard; 5) PGG-Threshold. Subjects received feedback on part 1 only at the end of the session. For the comparison of norms of cooperation across subject pools, we focus only on parts 2 and 3. Instructions for all five parts are in Appendix C.

simultaneously decide how to allocate these tokens between a group account (x) and a private account (w_i-x). Each group comprised $N = 4$ members and contributions to the group account could only take four levels, $x_i = \{0, 6, 14, 20\}$. Individual earnings were determined as follows:

$$\pi_i^1 = w - x_i + a \sum_{j=1}^{N} x_j$$

where the marginal per per capita return (MPCR) of the public good was $a = 0.5$. At the end of each period, a subject could observe individual contributions and earnings for each group member. Earnings cumulated from one period to the next.

PGG-Punishment was identical to PGG-Standard but for the addition of a second stage in which subjects had the opportunity to reduce, at a cost, the earnings of the other group members. After receiving feedbacks about individual contributions, every subject could assign $p_i = \{0, 1, 2\}$ deduction points to each group member; a deduction point had a cost of 1 token for the punisher and reduced the earnings of the targeted subject by $b = 4$ tokens. Punishment decisions were simultaneous and earnings were computed as follows:

$$\pi_i = \pi_i^1 - b \sum_{j \neq i}^{N} p_j^i - \sum_{j \neq i}^{N} p_i^j$$

At the end of each period, a subject could observe the deduction points he/she received and his/her final earnings. The punisher's identity was not revealed.

In a one-shot interaction, it is a dominant strategy for rational self-interested subjects to contribute zero in both PGG-Standard and PGG-Punishment, because the marginal per capita return of the public good is below 1 and above $1/N$, and to assign zero deduction points in PGG-Punishment. Given the finite repetition of the game, by backward induction, the same strategy applies to all periods. Group surplus is instead maximized when everyone contributes the whole endowment and never punishes.

The study comprised eight experimental sessions, equally divided across treatments for a total of 212 subjects. Session participants ranged between 20 and 32 and the laboratory hardware and set-up were identical across subject pools and locations. The same experimenter read the instructions in all sessions. *Representative* sessions were held in Faenza in a large hotel

conference room in the city center, where we deployed the mobile Bologna Laboratory for Experiments in Social Sciences (BLESS). *Student* sessions took place in Bologna at the permanent BLESS laboratory.[10]

In an effort to make the task more intuitive, we largely relied on graphical elements. To facilitate elderly people unfamiliar with computers, all choices were made by simply touching the screen (see sample screens in the Appendix C) and there was indeed no need to type or use a mouse. At the end of the session, subjects filled in a questionnaire. The average *Student* (*Representative*) session lasted about 90 (120) minutes. Subjects were paid in private at the end of the session. The experiment paid 1 Euro for every 40 tokens earned. There was no show-up fee in *Student* sessions and a 30 Euros gasoline voucher in *Representative* sessions, under the assumption of a lower opportunity cost for students than for the general adult population. Average per-capita earnings were 19.50 Euros in *Student* sessions and 17 Euros (plus the show-up fee) in *Representative* sessions.

3   Results

We report five main results; we first consider aggregate behavior (Results 1, 2, 3) and then present the evolution of contributions and punishment norms over time (Results 4 and 5).

In the PGG-Standard, how do observed contribution levels in the student population compare to the ones observed in the representative population?

Result 1 *The representative sample cooperates more in the standard Public Goods Game than the student sample.*

The average cooperation level over the eight periods was 9.1 in the *Representative* and 6.8 in the *Student* treatment. Support for Result 1 is provided by an OLS regression, where the dependent variable is the contribution level of a subject in a period (Table 4, Model 1).[11] The

---

[10]Upon arrival, subjects were seated at a visually separated desk, and no form of communication was allowed during the experiment. A paper copy of the relevant instructions was handed out before each part and read loud by the experimenter. Before PGG-Standard and PGG-Punishment, subjects had to answer a computerized quiz to ensure their understanding. Everyone had to answer all questions correctly before proceeding. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

[11]We also run ordered probit regressions to take into account that the dependent variable was not continuous but could take on only four values. Ordered probit regressions yielded qualitatively and quantitatively similar results.

Table 3: Average contributions to the public good

Public Good Game	All periods			First Period		
	Representative	Student	Pooled	Representative	Student	Pooled
Standard	9.11 $\sim$	6.79	7.97	11.04$\sim$	10.54	10.79
	$\sim$	$\wedge^{**}$		$\sim$	$\sim$	
Punishment	8.88$<^{**}$	12.77	10.78	9.07 $\sim$	10.52	9.78
Total	8.99	9.78		10.06	10.52	

Notes: Average individual contributions to the public good, divided by subject pool and stage game. Significance level from two sided Mann-Whitney ranked-sum tests are reported for each pairwise comparison. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. $\sim$ indicates that the we failed to find any statistically significant difference. We treated each session as an independent observation when pooling all periods. For first period data, we consider each subject as an independent observation.

main explanatory (dummy) variable *Representative sample* has a positive and highly significant coefficient hence suggesting that the general public cooperates more than college students. To account for subjects' understanding, we also included the dummy *Low understanding* taking into account the number of mistakes in the control questions and the time taken to answer correctly all questions.[12] Instead, when following a very conservative approach and considering each session as an independent observation, we fail to find any statistically significant difference in contributions across subject pools in PGG-Standard (Mann-Whitney ranked-sum, p=0.149, $N_R = N_S = 4$, two-sided).

In PGG-Punishment, how do contribution levels observed in a student population compare to contributions in the general population?

Result 2 *With punishment, the representative sample cooperates less than the student sample. The opportunity of peer punishment enhances cooperation levels in the student sample but not in the representative sample.*

The introduction of peer punishment reverses the effect: the general population contributes less as compared to students. Average cooperation in the PGG-Punishment was 8.9 in the *Representative* and 12.8 in the *Student* treatment. The difference across subject pools in the PGG-Punishment is highly significant according to an OLS regression on individual contribu-

---

[12]The dummy takes value 1 for subjects in the last decile of the distribution according to either the number of mistakes or the total answering time.

tions (Table 4, Model 3). The negative coefficient of the explanatory variable *Representative sample* lends support to the evidence that students are more cooperative then the general public in the PGG-Punishment. The difference is also statistically significant according to a Mann-Whitney ranked-sum test (p=0.021, $N_R = N_S = 4$, two sided). Moreover, the opportunity of peer punishment enhances cooperation levels in the student sample but not in the representative sample (Mann-Whitney ranked-sum test, p=0.021, $N_{S-PGG-std} = N_{S-PGG-punish} = 4$, p=0.773, $N_{R-PGG-std} = N_{R-PGG-punish} = 4$, two sided).[13]

Table 4: Treatment effect on contributions.

| | Dep.var.: Contribution | | | |
| | PGG Standard | | PGG Punishment | |
	Model 1	Model 2	Model 3	Model 4
Representative Sample	2.316***	-0.047	-3.887***	-1.847*
	(0.668)	(0.914)	(0.742)	(0.959)
Low understanding	-0.021	-0.021	-1.473	-1.473
	(0.815)	(0.816)	(1.005)	(1.006)
Period		-0.919***		0.423***
		(0.093)		(0.104)
Period x Representative		0.525***		-0.453***
		(0.129)		(0.134)
N.obs.	1696	1696	1696	1696
R-squared	0.024	0.071	0.072	0.080

Notes: Linear probability regression on individual contribution levels, individual-level random effect and robust standard errors (in parentheses). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10 % level, respectively.

---

[13]Per-period profit decreases from PGG-Standard to PGG-Punishment for both subject pools. In the *Representative* treatment the earnings drop was more pronounced; subjects earned, on average, about 9.1 tokens less in each period. The loss was of only 1.5 tokens among students.

Figure 1: Contributions to the public good over time

(a) Standard

(b) Punishment
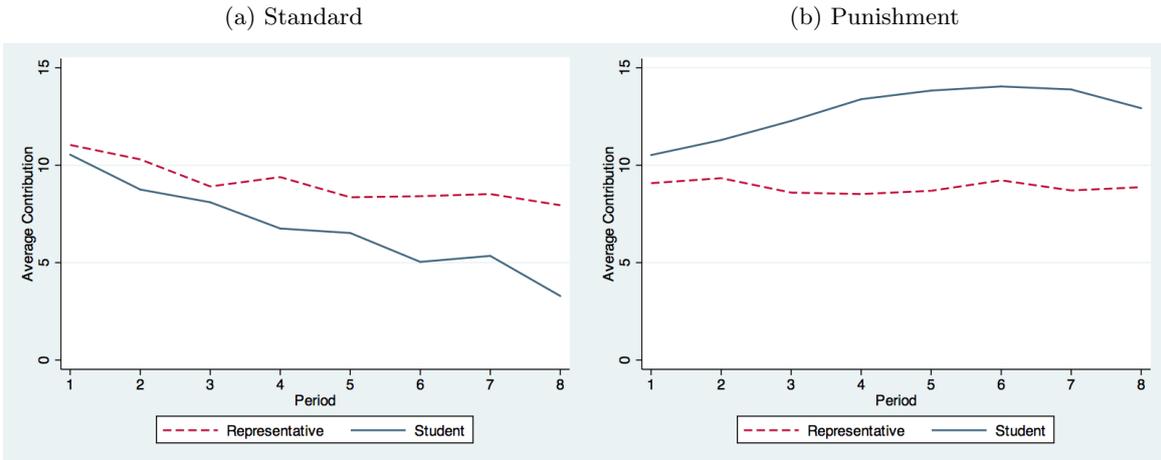
We now consider individual decisions over time to grasp a better understanding of the underpinning dynamics of cooperation. As a matter of fact our experiment offers repeated measures of cooperation, this allows us analyzing the initial contribution levels as well as the dynamics of contribution and punishment over time (Figure 1).

Result 3 *Cooperation in the initial period is indistinguishable between representative and student samples both with and without opportunities to punish.*

Table 3 and Figure 1 provide support for Result 3. In the PGG-Standard, individual contributions in the first period are not significantly different across subject pools (Mann-Whitney ranked-sum, p=0.614, $N_R = 108$, $N_S = 104$, two-sided). The same conclusion holds for the PGG-Punishment (p=0.169, $N_R = 108$, $N_S = 104$, two-sided). We also regressed contributions in the first period over the dummy *Representative sample* (see Table A-1 in Appendix A) and it turns out that differences across treatments are not statistically significant for both PGG-Standard (Model 1) and PGG-Punishment (Model 2).

As shown in Figure 1, differences across treatments emerged over time. While in the first period, the two pools are indistinguishable, within the last period of the PGG-Standard, the representative sample shows a cooperation level more then twice as large as the student sample (7.9 and 3.3, respectively). In particular, cooperation among students unravels rather quickly while the general population manages to sustain a more stable contribution level. Support for this finding is provided in Table 4 (Model 2). The negative coefficient for *Period* reasserts the presence of a declining trend in the PGG-Standard, while the positive coefficient in the interac-

tion term indicates that the decline in the *Representative* treatment is less pronounced than in the *Student* treatment. The dynamics in the PGG-Punishment were exactly the opposite (see Model 4); contributions tend to increase over time and the upward trend is more marked in the *Student* than in the *Representative* treatment.

What drives these different trends in cooperation across games and subject pools? To answer this question, we will focus on individual decisions to contribute and punish in the last part of this Section. We first consider whether the reaction to others' contributions – i.e., conditional cooperation – is the same across treatments. Are adjustment dynamics the same in our two participant pools?

Result 4 *Current contributions depend less on observed past contributions in the representative sample than in the student sample.*

We consider an indirect measure of conditional cooperation (Fischbacher et al., 2001; Kocher et al., 2008) and test how current contributions adjust to previous contributions made by others.[14] Here we mostly focus on the PGG-Standard that in our view provides a cleaner test of conditional cooperation. Indeed in the PGG-Punishment previous contributions are likely to be connected with punishment, and not just with cooperative behavior.[15]

Table 5 (Models 1 to 3) lends support to Result 4 for the PGG-Standard. In all specifications, the dependent variable is the contribution level at time t for each subject. In the first two models we consider each sample separately and the regressor of interest is the sum of other group members' contributions in period $t-1$ (*Others' contributions in t-1*).[16] In PGG-Standard, *Others' contributions in* $t-1$ has a positive and highly significant impact in the student sample but is not significant in the representative sample (Models 1 and 2, respectively in Table 5). This result is confirmed also in the pooled sample (Model 3).[17]

---

[14]Conditional cooperation is commonly defined as the willingness to contribute to the common pool based on the expectation that others will contribute as well. We consider an indirect measure and assume that a subject's belief about future group members' contributions depends on their past contributions. Our strangers-matching protocol weakens this relation compared to a partner-matching protocol. Alternatively, one could have used the strategy method to directly elicit conditional cooperation.

[15]If high cooperators are more likely to punish than free riders, there should be a correlation between the punishment received by a subject and others' contributions in the previous period.

[16]We also control for time trend and low understanding as in Table 4.

[17]As a robustness check, we run the same regressions using a GMM system methodology to control for potential

Table 5: Conditional cooperation and observed contributions.

Dep.var.:	PGG-Standard			PGG-Punishment		
	Representative	Students	Pooled sample	Representative	Students	Pooled sample
Contribution	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Others' contributions in t-1	0.017	0.081***	0.094***	0.059***	0.088***	0.093***
	(0.020)	(0.018)	(0.018)	(0.016)	(0.016)	(0.015)
Period	-0.293***	-0.630***	-0.445***	-0.009	0.153	0.066
	(0.107)	(0.112)	(0.078)	(0.100)	(0.117)	(0.076)
Low understanding	1.516	-1.219	0.194	-0.872	-2.017	-1.423
	(1.246)	(1.065)	(0.843)	(1.303)	(1.614)	(1.032)
Representative Sample			4.304***			-2.288**
			(0.911)			(1.036)
Others' contr in t-1 x Representative			-0.082***			-0.033
			(0.026)			(0.022)
N.obs.	756	728	1484	756	728	1484
R-squared	0.014	0.097	0.075	0.005	0.045	0.099

Notes: OLS regression on cooperation levels with individual random effects and robust standard errors (in parentheses). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Models 4 to 6 in Table 5 replicate the same analysis for the PGG-Punishment. Both pools tend to adjust to observed contributions. However, the difference in conditional cooperation between the two samples is less pronounced in the PGG-Punishment than in the PGG-Standard: the coefficient of interaction *Others' contributions in t-1 x Representative sample* is indeed negative, although not significant (see Model 6).

We now take into account the analysis of punishment behavior. The differential impact of punishment in the two subject pools may be the result of different amounts or different types of punishment. Does the representative pool punish less than the student one? Or does the representative pool punish differently from the student pool?

Result 5 *The representative sample punishes no less than the student sample, but engages more in anti-social punishment.*

Table 6: Received punishment by treatment.

	Overall Punishment	Pro-social Punishment	Anti-social Punishment
Representative	7.21	9.98	3.91
	(6.23)	(6.33)	(4.15)
Student	5.93	11.10	2.32
	(6.79)	(6.90)	(3.66)

Notes: Average received punishment by subject pool and type of punishment. Standard deviation in parenthesis.

Support for Result 5 is presented in Figure 2 and Tables 6 and 7. The extent of punishment is similar across treatments and, if anything, it is higher in the representative than in the student sample (7.2 vs 5.9).[18] Hence, the absence of a positive effect of punishment on cooperation levels in the representative sample must stem from reasons other than the lack of punishment. We surmise that the explanation should be found in the type rather than in the extent of punishment.[19]

---

endogeneity of the variable *Others' contributions in t-1*. Results are consistent with the present estimates and are reported in Appendix A.

[18]A Wilcoxon rank-sum test does not reveal any statistically significant difference when taking each session as an independent observation (p=0.149, $N_R = N_S = 4$).

[19]Punishment should have a positive impact on contributions if free-riders, or low contributors for that matters, increase their cooperation level in order to avoid to be punished. However, there are circumstances in which peer punishment can have a detrimental effect on cooperation levels; there is indeed evidence that also high cooperators can sometimes be punished out of a desire of revenge (see, for instance, Herrmann et al., 2008).

Table 6 summarizes prosocial and anti-social punishment:[20] pro-social punishment is always more frequent than anti-social punishment but the ratio is very different in the representative and in the student sample (2.5:1 vs. 4.8:1, respectively). Figure 2 reports the average amount of punishment conditional on the deviation from the average contribution of the other group members (Fehr and Gächter, 2000). Punishment associated to negative deviations captures prosocial punishment directed toward free-riders, while punishment associated to positive deviation instead captures anti-social punishment behavior. Whereas anti-social punishment is less severe than social punishment in both samples, subjects in the *Representative* treatment are more willing to punish high cooperator than subjects in the *Student* treatment.

Figure 2: Punishment targets lower contributors in the group.
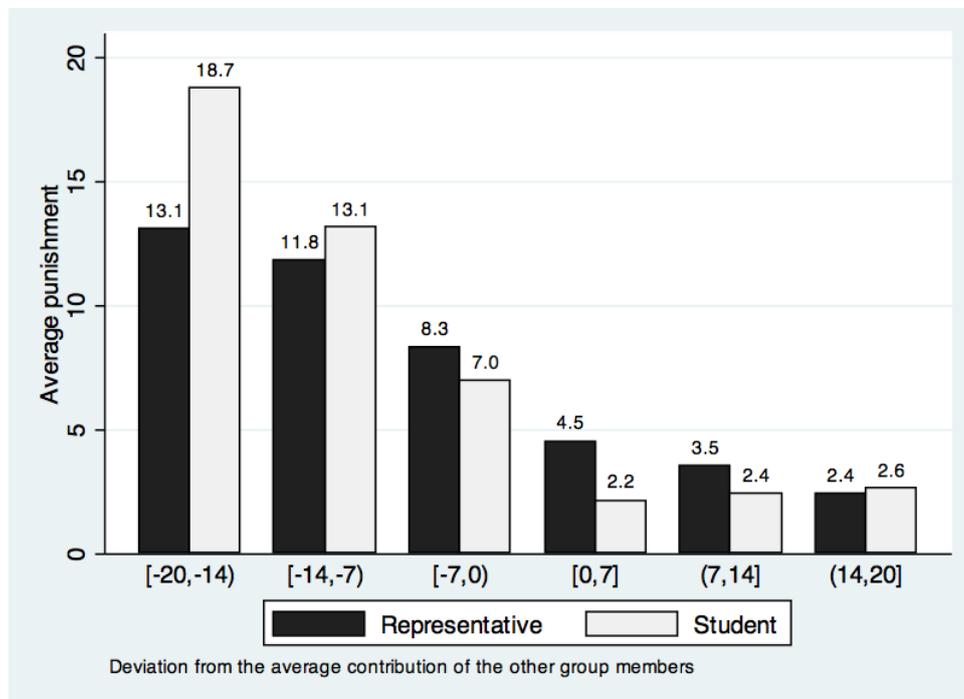
OLS regressions in Table 7 show that deviations from the average group contribution are punished, and punishment is more severe for negative as compared to positive deviations. Models 1 and 2 report results for *Representative* and *Students* treatments, respectively. The variable *Negative deviations (abs)* has a positive coefficient and is highly significant in both treatments (see Model 1 and 2) hence giving support to the idea that the more the contribution falls short of others' contributions the more severe the punishment is. Quite surprisingly, also the coefficient of the variable *Positive deviations* is positive and significant. That implies that punishment

---

[20] As defined by Herrmann et al., 2008, punishment is pro-social when the punisher has contributed more than the subject who is the target of punishment. On the contrary, punishment is anti-social when the punisher has contributed as much as the target did, or even less.

increases as the gap between other's contributions and socially-minded subjects' contributions widens. The negative and highly significant coefficient for *Other's contributions* implies that a deviation from others' contributions is punished more severely if the sum of the contributions is small; sign and magnitude of this coefficient are consistent with similar studies using a strangers-matching protocol (see, Fehr and Gächter, 2000).

When pooling all samples (Model 3), one can assess treatment-specific patterns by looking at the additional explanatory variables: *Representative sample*, *Positive deviation × Representative*, and *Negative deviation(abs) × Representative*. In the representative sample there is a (weakly) significantly higher extent of punishment, (weakly) significantly more punishment of positive deviations, and significantly less punishment of negative deviations than in the student sample. The above evidence suggests that the representative sample sanctions relatively less free-riders and relatively more contributors than the student sample. This pattern could have discouraged cooperation and could explain the weak impact of punishment in the representative sample.

4   Conclusions

When facing collective action problems, some societies benefit from the availability of opportunities for peer punishment while others do not (i.e., Herrmann et al., 2008). We show that, even within the same society, the impact of peer punishment in promoting cooperation can vary widely depending on the sample of the population considered. This study compares the cooperative behavior of two samples sharing similar geographical and cultural origins but differing along important socio-demographic dimensions: college students and a representative sample of the general adult population. We find that results from experiments on norms of cooperation and punishment among students cannot be readily generalized to the society at large.

We replicate the standard finding that students are on average less cooperative than the general population (Result 1, see for instance, Cappelen et al., 2010; Fehr et al., 2006; Belot et al., 2010). Unlike most experiments involving the general population, in our study participants interacted repeatedly. While initial contributions were indistinguishable (Result 3), differences

Table 7: Treatment effect on punishment.

Dep.var.: Deduction points received	Representative Model 1	Students Model 2	Pooled sample Model 3
Other's contributions	-0.016***	-0.016***	-0.016***
	(0.004)	(0.004)	(0.004)
Positive deviation	0.060***	0.030**	0.030**
	(0.009)	(0.013)	(0.013)
Negative deviation(abs)	0.173***	0.238***	0.237***
	(0.014)	(0.011)	(0.010)
Period	-0.023	-0.019	-0.021
	(0.022)	(0.019)	(0.015)
Low understanding	-0.040	-0.068	-0.053
	(0.091)	(0.086)	(0.063)
Representative Sample			0.439*
			(0.236)
Others' contrib. x Representative			0.000
			(0.006)
Pos. deviation x Representative			0.030*
			(0.016)
Neg. deviation x Representative			-0.064***
			(0.017)
N.obs.	864	832	1696
R-squared	0.330	0.576	0.466

Notes: OLS regression on deduction points received with individual random effects and robust standard errors (in parentheses). *Negative deviation* is the absolute value of the deviation of a subject's contribution level with respect to the average contribution of the others in her group, in the case that the contribution falls short of the average, and 0 otherwise. *Positive deviation* takes values other than 0 when a subject's contribution is larger than the average contribution of the others. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

across samples emerged over time. The diverging dynamics can be explained by the larger fraction of conditional cooperators among students than in the general population (Result 4). The general population managed to sustain contributions over time without a significant decline in cooperation levels toward the end of the game.

In the literature, there is evidence that differences in the ability to cooperate across societies are magnified when punishment opportunities are present (Herrmann et al., 2008). Our results document that punishment could reverse the ordering of societies in terms of cooperativeness even when the participant pools are sampled from the same geographical area. Punishment opportunities had a positive effect on cooperation in the student sample, whereas little or no effect was detected in the general population. In terms of cooperation levels, students contributed less than the general population in the standard task; while the introduction of peer punishment made students more cooperative than the general population (Result 2).

There was no lack of punishment actions among the general population. As a matter of fact punishment levels were higher in the general population sample than in the student sample. The main difference is the way punishment was used by the two participant pools. While punishment was mostly directed toward free-riders, in the general population a remarkable extent of punishment toward cooperators can be found (i.e., anti-social punishment), this happened with a higher frequency than in the student pool. Hence, a possible explanation for our findings is that punishment did not promote cooperation among the general population because it was frequently directed toward cooperators rather than free-riders.

This experiment was conducted with high methodological standards and ensured high levels of comparability across treatments. First, procedures and software conveyed the tasks in an intuitive and graphical way in order to facilitate understanding. Second, instructions, laboratory, and experimenters were the same for both samples. Third, every participant student and general population was born within the borders of Emilia-Romagna region in Italy. Fourth, we hired professionals recruiters to select a representative sample of the Italian population stratified according to gender, age, and employment status.

Two general conclusions follow from this study. Experiments comparing students and other

samples on the basis of one-shot games should exert caution when generalizing their results, as some differences emerge only over time. Our evidence suggests that students and the general population initially cooperate at similar levels but adopt distinct strategies in cooperation and punishment when reacting to the same situation. Finally, if we were to classify the Italian society according to the impact of peer punishment in promoting cooperation, one would draw opposite conclusions depending on whether the experiment was run with college students or with the general population.

References

Arellano, M. and S. R. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–97.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29 – 51.

Bellemare, C. and S. Kroger (2007). On representative social capital. *European Economic Review 51*(1), 183 – 202.

Bellemare, C., S. Kroger, and A. van Soest (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. *Journal of Economic Behavior & Organization 78*(3), 349 – 365.

Bellemare, C., S. Krger, and A. Van Soest (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica 76*(4), 815–839.

Belot, M., R. Duch, and L. Miller (2010). Who should be called to the lab? a comprehensive comparison of students and non-students in classic experimental games. Discussion Papers 2010001, University of Oxford, Nuffield College.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115 – 143.

Bosch-Domenech, A., J. Montalvo, R. Nagel, and A. Satorra (2002). One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review 92*(5), 1687–1701.

Branas-Garza, P., M. Bucheli, and T. Garcia-Munoz (2012). Dynamic panel data: A useful technique in experiments.

Cappelen, A. W., K. Nygaard, E. O. Sørensen, and B. Tungodden (2010). Efficiency, equality and reciprocity in social preferences: A comparison of students and a representative population. Discussion Paper Series in Economics 28/2010, Department of Economics, Norwegian School of Economics.

Carpenter, J., C. Connolly, and C. Myers (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics 11*, 282–298.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2008). Representative trust and reciprocity: prevalence and determinants. *Economic Inquiry 46*(1), 84–90.

Egas, M. and A. Riedl (2008). The economics of altruistic punishment and the mainteinance of cooperation. *Proceedings of the Royal Society B 275*, 871–878.

Ermisch, J., D. Gambetta, H. Laurie, T. Siedler, and S. C. Noah Uhrig (2009). Measuring people's trust. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 172*(4), 749–769.

Exadaktylos, F., A. Espin, and P. Branas-Garza (2013). Experimental subjects are not different.

Falk, A., S. Meier, and C. Zehnder (2012). Do lab experiments misrepresent social preferences? *Journal of the European Economic Association forthcoming*.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E., M. Naef, and K. M. Schmidt (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review 96*(5), pp. 1912–1917.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischbacher, U., S. Gchter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397 – 404.

Gächter, S., B. Hermann, and C. Thöni (2003). Norms of cooperation among urban and rural dwellers. experimental evidence from russia. *Mimeo, University of St. Gallen.*.

Gächter, S., B. Hermann, and C. Thöni (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior and Organization 55*, 505 – 531.

Greiner, B. (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.

Harrison, G. W., M. I. Lau, and M. B. Williams (2002). Estimating individual discount rates in denmark: A field experiment. *The American Economic Review 92*(5), pp. 1606–1617.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*.

Henrich, J., J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2010). Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science 327*(5972), 1480–1484.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial Punishment Across Societies. *Science 319*(5868), 1362–1367.

Hislop, D. R. (1994). State dependece, serial correlation and heterogeneity in intertemporal labor force participation of married woman. *Econometrica 6*(67), 1255–1294.

Kocher, M. G., T. Cherry, S. Kroll, R. J. Netzer, and M. Sutter (2008). Conditional cooperation on three continents. *Economics Letters 101*(3), 175 – 178.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–26.

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: self-governance is possible. *American Political Science Review 86*, 404–417.

Stewart, M. (2006, June). Maximum simulated likelihood estimation of random-effects dynamic probit models with autocorrelated errors. *Stata Journal 6*(2), 256–272.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics 126*(1), 25 – 51.

Appendix A First Period and Dynamic Panel Estimation

Result 3 suggests that contributions in the first period are indistinguishable across subject pools. In Table A-1, we regress individual contributions over the dummy *Representative sample* and provide support to Result 3.

Table A-1: Treatment effect on contributions in period 1: PGG Standard and Punishment.

	Dep.var.: Contribution in period 1	
	PGG Standard	PGG Punishment
	Model 1	Model 2
Representative Sample	0.118	-0.343
	(0.298)	(0.345)
Low understanding	-0.338	-0.291
	(0.389)	(0.360)
N.obs.	212	212
Log likelihood	-284.328	-288.298

Notes: Ordered logit regression on individual contribution levels in period 1, standard errors robust for clustering at the session level (in parentheses). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Result 4 suggests that differences emerge over time, as students condition their behaviour on previous experience more than the general population does. Apart from conditional cooperation there can be two additional factors that influence cooperation: *(i)* individual (unconditional) preferences for cooperation; and *(ii)* other unobserved individual characteristics. All these motivations are captured by the following equation:

$$x_{i,t} \;\; = \;\; v_i + \alpha x_{i,t-1} + \beta \sum_{j=1}^{3} x_{j,t-1} + u_{i,t} \qquad (1)$$

where $x_{i,t}$ indicates the contribution to the public good of subject i at time t; $\sum_j x_{j,t-1}$ indicates the sum of the contributions of the other group members in the previous period and is meant to capture conditional cooperation. Please recall that groups were formed according to a strangers-matching protocol at the beginning of each period. The variable $x_{i,t-1}$ is the contribution of subject i in period $t-1$ and measures the persistence of subjects' choice; we interpret this variable as a proxy for individual preferences toward cooperation. Finally, v_i is an individual time-invariant component capturing intrinsic characteristics of each subject that cannot be observed.

To account for the endogeneity problem arising from the introduction in the model of the variable $x_{i,t-1}$, we implement a two-step GMM system estimator.

Following the general model of Eq. 1 we estimate a dynamic panel of the form:

$$x_{i,t} = \alpha_1 x_{i,t-1} + \beta_1 \sum_j x_{j,t-1} + u_{i,t} \qquad (2)$$

$$u_{i,t} = v_i + e_{i,t} \qquad (3)$$

where v_i are unobserved individual effects, $e_{i,t}$ are the observation specific errors, which have zero mean ($E[e_{i,t}] = E[v_i e_{i,t}] = 0$), constant variance and are uncorrelated across time and individuals $E[e_{i,t}] * E[e_{i,s}] = 0$ for each i, j, t, s and $i \neq j$. We consider the variable $\sum_j x_{j,t-1}$ as exogenous thanks to the strangers matching protocol implemented in our setting and to the fact that the choice of individual i is excluded from the calculation of the aggregate group contribution in the previous period.

Endogeneity is an issue in this context because of the small number of time periods available for the estimation, or what is defined in the literature as small sample bias (Nickell, 1981). This could also be the reason for the scarce use of this methodology in the experimental literature.[1] We implement a two-step GMM system estimators.[2]. which are robust under heteroskedasticity with the Windmeijer (2005) finite-sample correction to avoid downward bias

As we are interested to introduce in our model a time invariant regressor, the *Representative Sample* dummy, difference GMM estimators are not appropriate (as the time invariant regressor would be canceled out in the procedure). Hence we use a system GMM estimator that maintains both the original and the differenced equation and uses both levels and differenced variables as instruments. For simplicity, the discrete dependent variable is approximated to continuous as in Hislop (1994).[3]

For each estimated Model we control the p-values of the Sargan and Hansen test of over-

---

[1]Notable exception is Branas-Garza et al. (2012) that compares static and dynamic panel estimation in an experimental setting.

[2]The problem of endogeneity in small samples has been originally tackled by Arellano and Bond (1991) seminal paper but other contributions have extended the applicability of the methodology in various directions in the following years. See Arellano and Bover (1995); Blundell and Bond (1998)

[3] We leave for future research the possibility to implement a dynamic discrete choice panel with endogenous regressors as in Stewart (2006) for example.

Table A-2: Instruments for system GMM estimation of PGG Standard

Instruments for first differences equation		
Type of instrument	Standard	GMM-type
Variable used	Others' contribution	Own contribution
Specification	First difference, lag 1	Level, lags from 3 to max
Instruments for levels equation		
Type of instrument	Standard	GMM-type
Variable used	Others' contribution	Own contribution
Specification	Level, lag 1	First difference, lag 2

Table A-3: Instruments for system GMM estimation of PGG with Punishment

Instruments for first differences equation		
Type of instrument	Standard	GMM-type
Variable used	Others' contribution	Own contribution
Specification	First difference, lag 2	Level, lags from 2 to max
Instruments for levels equation		
Type of instrument	Standard	GMM-type
Variable used	Others' contribution	Own contribution
Specification	Level, lag 2	First difference, lag 1

identifying restrictions and the Arellano and Bond (1991) second order autocorrelation test in first differences. The former test indicates a correctly specified set of instruments, while the Arellano Bond test evaluates the presence of residual order autocorrelation of the differenced error, which in this context is a signal of endogeneity between the lagged endogenous variable and the differenced fixed effect, condemning the related variable to be an invalid instrument. In Table A-2 and A-3 we list the instruments implemented for GMM system estimation for the Standard and Punishment treatments respectively. In Model 5 and 6 of the Punishment treatment, a correct specification was not achieved using the same instruments: this suggests that the two subject pools are substantially different and that it is necessary to explore other variables in order to find an explanation to individual behavior in the punishment treatment.

Table A-4 reports estimates for Standard (Models 1 to 3) and Punishment (Models 4 to 6) variants of the Public Goods Game. Model 1, considers data only from the *Representative* treatment in the PGG-Standard; the large and significant coefficient of the variable *Own contribution in t-1* suggests that individual contributions in the general population are very persistent. On the contrary, the coefficient of the variable *Other's contribution in t-1* is very small and not significant at any conventional level. Taken together, these two variables confirm

that in the representative sample subjects tend to stick to their choices and are less influenced by others' behavior. When considering the *Student* treatment only (Model 2), we find that coefficient on *Other's contribution in t-1* is larger and highly significant, and this supports the idea that conditional cooperation plays a key role among students even after controlling for their own contributions in $t - 1$.

When moving from the Standard to the Punishment variant, we find that none of the explanatory variables can account for observed cooperation in the *Representative* treatment (Model 4). In this case, behaviour is thus explained by a variable not included in this present model; received punishment seems to be a likely candidate (see discussion in Section **??**). For the student sample, the only marginally significant variable is *Other's contribution in t-1* hence yielding further evidence in favour of the idea that students are more conditional cooperators than the general public.

Table A-4: Dynamic Panel Estimation.

Dep.var.: Contributions

	PGG Standard			PGG Punishment		
	Representative Model 1	Students Model 2	All sample Model 3	Representative Model 4	Students Model 5	All sample Model 6
Own contribution in t-1	0.711**	0.533***	0.460***	0.059	0.143	0.109**
	(0.28)	(0.14)	(0.16)	(0.07)	(0.10)	(0.05)
Others' contribution in t-1	0.023	0.086***	0.051***	0.006	0.113*	0.051
	(0.03)	(0.02)	(0.02)	(0.08)	(0.06)	(0.06)
Representative Sample			3.134			-6.008***
			(3.11)			(1.92)
N.obs.	756	728	1484	648	624	1272

Notes: Blundell-Bond (1998) panel estimation. P-value for Sargan and Hansen test (null hypothesis that the overidentifying restrictions are valid) and Arellano-Bond test for second and third order autocorrelation in first differences: (0.686,0.518,0.077) for Model 1;(0.000,0.194,0.022) for Model 2; (0.000,0.153,0.016) for Model 3; (0.046,0.178,0.083) for Model 4; (0.002,0.047,0.403) for Model 5; (0.260,0.845,0.373) for Model 6. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Appendix B   Rectruitment

Recruitment Procedure for the Representative Sample

Participants to the *Representative* treatment were recruited from the general population of the province of Ravenna, which is part of Emilia Romagna region, located in the North of Italy. Eligible candidates for the study had to: *(i)* be at least 18 years; *(ii)* be born in the county; *(iii)* be resident in the county; *(iv)* have a good knowledge of spoken and written Italian. The experimenters, before the experimental sessions were carried out, double checked participants' ID cards so to guarantee that all subjects met the requirements (age and place of birth). At the beginning of each session, the experimenter made public that all subjects in the room were born and resident in the same province (or at least in the region) with the explicit aim to make this information common knowledge.

We wanted a representative sample of the Italian population with respect to age, sex, and employment status, since these characteristics might be relevant for the investigation of cooperation norms in a society. The sample was stratified according to three categories of age (18-39; 40-59; 60 and older), two of sex (male and female), and three for employment status (employed; housewives and retired; others, including students and unemployed). For the composition of the target sample we referred to the 2009 statistics for the Italian population.[4]

We hired two professional companies – Metis-Ricerche and Demoskopea – to recruit subjects that comply with the aforementioned requirements. We provided these companies with a message and a script to approach potential participants. Details about the study and the goal of the experiment were not disclosed to subjects during recruitment; recruiters had no prior knowledge of the purpose or the content of our study. We asked them to recruit people resident both in the town and outside the town where the experiment has been carried out; in both cases only subjects resident in the province of Ravenna could be involved.. In addition to the aforementioned requirements, special categories of people were ex-ante barred from participation, such as: employees of the research sector; people that participated to market researches in the preceding three months; recruiters family members; employees of marketing companies

---

[4]We referred to the number of inhabitants registered on 1st of January 2009. Age range: 18-39 years, 34.8%; 40-59 years, 34.6%; 60 and more, 31.6%. Sex: male, 48%; female 52%. Employment status: employed, 42%; housewives and retired, 37%; others 21%. Source: http://demo.istat.it/pop2009/index1.html

and of the press sector in general. Moreover, no more than two people per session needed to be acquainted with each other.

One company (Metis-Ricerche) recruited subjects for the first three sessions. Potential subjects were identified with the use of telephone book entries and approached by telephone calls. All phases of the recruitment process were performed from the companys headquarters, and, in case of acceptance, the company provided the participant with a confirmation letter. This letter contained the same information that was provided on the phone by the company operators and that we had previously agreed upon with the company. The former company Metis-Ricerche decided not to renew their contract for recruitment of people in other locations, as the recruitment procedures turned out to be more expensive than what they expected. The latter company (Demoskopea) was in charge of the recruitment of subjects for the last session. Local representatives of Demoskopea contacted directly subjects in each province. The local recruiters proceeded with the choice/random extraction of names from telephone books and with random contacts obtained through personal interactions as instructed by the headquarters. In the following part of our study, we report the message used by both companies to recruit subjects for our study.

Message for recruiters with instructions

We would like to invite you to participate to a meeting organized by the Universities of Bologna and Oxford. We are looking for people born in the city and within the province of Faenza. The aim of this study is strictly scientific. There are no commercial purposes and the identity of all participants will be always kept anonymous. Our interest is to understand how Italians take decisions in situations dealing with money. During the meeting you will be given several different situations and you will be kindly asked to take decisions. Taking decisions is an easy task. No particular skills are required.

We offer you a payment of 30 euros in petrol tokens, plus a sum in cash that, according to your choices and those of other participants, will amount up to 25 euros. You will be paid at the end of the meeting, which we expect to last no more than 2 hours and a half.

If you wish to verify the accuracy of these information, please contact *Name of the secretary in charge*, from University of Bologna, or visit the website http://www.unibo.it/Portale/Ricerca. If you accept our offer, you may show up *location* at *time*, which is *description of how to get to the location*

By participating to this meeting, you make a contribution to one of the few scientific research projects supported by the European commission in Italy.

F.A.Q. (In case somebody asks, recruiters are allowed to provide the following extra informa-

tion)

- How do we make our choices?
  - Choices are made very easily, touching the screen with a finger. It is just like an ATM or a cell phone with touch screen.

- In a nutshell, how does this activity work?
  - You will be given several different situations and you will be kindly asked to choose among alternatives. There is no right answer, we just want to know your opinion.

Recruitment Procedure for the Students sample Subjects that belonged to the students sample were recruited according to the standard procedure implemented in a regular laboratory experiment. Announcements were sent to potential participants in the ORSEE database: this database is the one commonly used in the Bologna Laboratory for Experiments in Social Sciences (BLESS). We slightly changed the standard announcement to include the requirement that subjects to this study should be born in Emilia-Romagna. Subjects were asked to reach the laboratory on the agreed day for the carrying out of the session with a valid ID card, which was checked by experimenters to verify the birth and residence requirements. In the following part of our study, we will provide the announcements sent via the ORSEE platform to recruit the participants for the students sessions.

Message for the students sessions
–— PLEASE, IGNORE THIS MESSAGE IF YOU WERE NOT BORN IN EMILIA ROMAGNA

Hello *(first name last name)*
you are kindly invited to participate to a research in our Laboratory of Experimental Economics.
Only people that were born in Emilia Romagna can take part to this study.
Please, do not sign in if you were born in another region. People born outside of Emilia Romagna, even if they sign in, will not be granted the chance to participate to the study.
Sessions wil ltake place in the following dates and time slots:
(session list)
To choose the session, please click on the following link:
(link)

(If you cannot click on the link, please select it, copy it bly cliking on the right button of the mouse and paste it in the address line by clicking on right button once again)

Appendix C   Sample screens

Figure C-1: Sample screenshot for Lottery task.

Figure C-2: Sample screenshot for PG task

(a) Decision screen   (b) Feedback screen

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it