

# Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Exadaktylos, Filippos, Espín, Antonio M. and Branas-Garza, Pablo (2013) Experimental subjects are not different. *Scientific Reports*, 3 (1213). pp. 1-6. ISSN 2045-2322 (doi:10.1038/srep01213)

UNSPECIFIED

This version is available at: <http://eprints.mdx.ac.uk/10510/>

## Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

[eprints@mdx.ac.uk](mailto:eprints@mdx.ac.uk)

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>



# Experimental subjects are not different

Filippos Exadaktylos<sup>1</sup>, Antonio M. Espín<sup>2</sup> & Pablo Brañas-Garza<sup>3</sup>

## SUBJECT AREAS:

PSYCHOLOGY

PSYCHOLOGY AND BEHAVIOUR

APPLIED PHYSICS

RANDOMIZED CONTROLLED  
TRIALS

Received

10 October 2012

Accepted

3 January 2013

Published

14 February 2013

Correspondence and  
requests for materials  
should be addressed to  
P.B.-G.  
(branasgarza@gmail.  
com)

<sup>1</sup>BELIS, Murat Sertel Center for Advanced Economic Studies, İstanbul Bilgi University, Santral Campus, İstanbul, 34060, Turkey, <sup>2</sup>GLoBE, Universidad de Granada; Departamento de Teoría e Historia Económica, Campus de Cartuja s/n, 18071, Granada, SPAIN, <sup>3</sup>Middlesex University Business School, The Burroughs, NW4 4BT, London, UK.

Experiments using economic games are becoming a major source for the study of human social behavior. These experiments are usually conducted with university students who voluntarily choose to participate. Across the natural and social sciences, there is some concern about how this “particular” subject pool may systematically produce biased results. Focusing on social preferences, this study employs data from a survey-experiment conducted with a representative sample of a city’s population ( $N = 765$ ). We report behavioral data from five experimental decisions in three canonical games: dictator, ultimatum and trust games. The dataset includes students and non-students as well as volunteers and non-volunteers. We separately examine the effects of being a student and being a volunteer on behavior, which allows a *ceteris paribus* comparison between self-selected students (students\*volunteers) and the representative population. Our results suggest that self-selected students are an appropriate subject pool for the study of social behavior.

An introduction on the importance of experimental research using economic games is no longer necessary. Economic experiments are well established as a useful tool for studying human behavior within social scientists. Over the last years however, human experimentation has also found a central place in the research agendas of evolutionary biologists<sup>1–6</sup>, physiologists<sup>7–12</sup>, neuroscientists<sup>13–18</sup> and physicists<sup>19–24</sup>. The increasing number of well-published experimental studies and the impact they have on various fields across a number of disciplines has touched off a lively debate over the degree to which these data can indeed be used to refine, falsify and develop new theories, to build institutions and legal systems, to inform policy and to even make general inferences about the human nature<sup>25–29</sup>. In other words, the central issue is now about the external validity of the experimental data.

The main concern about external validity is related to certain features of experimental practices on the one hand (high levels of scrutiny, low monetary stakes and the abstract nature of the tasks), and a very particular subject pool on the other.

The latter has two dimensions. First, the subject pool in behavioral experiments is almost exclusively comprised of university students. More than the narrow socio-demographic array of characteristics that this group offers, what really threatens external validity is the existence of different behavioral patterns *once* such characteristics have been controlled for. That is, the under-representation of certain strata of the population is obviously true but not the real issue: once the distribution of these characteristics is known for the general population, researchers can account for such differences by adjusting the right weights to their statistical models. The real question in extrapolating students’ behavior to general populations is whether the coefficient estimates differ across the groups due to non-controllable variables. We should say that there is *student bias* if, after controlling for socio-demographics, students behave differently than the general population. The second dimension is that participants are *volunteers*. Naturally, the behavior of non-volunteers is not observed. There is a *self-selection bias* if volunteers share some attributes that make their behavior systematically diverge from that of non-volunteers.

The concern of the researchers of such biases is echoed by the increasing number of studies recruiting other, more general samples. A pronounced example is the use of the web in order to recruit subjects using platforms such as the Amazon Mechanical Turk<sup>30,31</sup>. Such attempts are very valuable since alternative samples are the best way of testing the robustness and generality of the results. However without specific information on how the alternative subject pool affects the results, leaving the physical laboratory and the control that this offers can be time-, energy- and money-consuming without necessarily positive returns in terms of generalizability.

So far insights as to whether student and self-selection biases systematically affect behavior can be found mainly in the economics’ literature. Regarding student bias there are two main sources. The first comes from experiments using both students and individuals pooled from a target population<sup>32–36</sup>. These belong to the family of the so-called artefactual field experiments<sup>37</sup>. The second comes from databases containing behavioral data drawn from



more general populations. This allows researchers to test whether different sub-samples (e.g., students) exhibit different behavioral patterns<sup>38–43</sup>. In the realm of social preferences, both practices have been extensively used over the last years, giving rise to a large number of field experiments. There is now plenty of evidence demonstrating that students are slightly less “pro-social” than other groups in a variety of designs and settings. For example students have been shown to behave less generously<sup>44,45</sup>, less cooperatively<sup>40,42,46,47</sup> and less trustfully<sup>48,49</sup>.

However, the bulk of this evidence comes from comparing students who self-select to experiments with other non-student samples who again self-select. So, what this literature gives evidence for is a small student bias but only *within volunteers*. Whether self-selected students’ behavior is representative for individuals who are not students and do not volunteer in scientific studies (presumably the “median” individual) we cannot know. Nor can we know whether self-selected students behave differently than non-self-selected students (the majority of the student population); ultimately we cannot know whether students in general are less pro-social than non-students (either self-selected or not). Thus, responding to concerns about student bias requires the *simultaneous* study of self-selection bias, which ultimately implies looking also within non-student populations.

Concerning self-selection bias, research has been relatively limited since this involves obtaining behavioral data of individuals *not willing* to participate to experiments. For student populations, researchers get hold of such datasets by making participation semi-obligatory during a class<sup>50,51</sup>. However, there are good reasons to assume that the behavior of these pseudo-volunteers will be quite distinct of the non-volunteers’ due to prominent demand effects<sup>52</sup>. Indeed both Eckel and Grossman (2000)<sup>50</sup> in a Dictator Game where the recipient was a charity and Cleave et al. (in press)<sup>51</sup> in a Trust Game found pseudo-volunteers to behave more “pro-socially”, which is in accordance of such hypothesis. Such effects could be even more pronounced when the experimenter is a professor of that specific class or course. The most recent evidence concerning self-selection<sup>49</sup> compares the frequency of a non-experimental decision (i.e., donation to a fund) between students who self-select to experiments and students who do not and finds no difference. Focusing on non-student populations, an appropriate dataset is even more difficult to obtain. We are aware of only two studies Anderson et al. (in press)<sup>47</sup> compares truck drivers (a kind of pseudo-volunteers) with volunteers sampled from a non-student population in a social dilemma game; Bellemare and Kröger (2007)<sup>48</sup> compares the distribution of *attributes* between participants of a survey who decide to participate in an experiment and those who decide not to. Both studies report non-significant differences.

Summarizing, the literature is not conclusive on whether self-selection is an issue in extrapolating experimental subjects’ behavior into other groups. Even less on whether self-selection affects students

and non-students in the same way since differences in methodologies (regarding whether the comparison is about attributes or decisions, whether the latter are experimental or non-experimental and more importantly whether the same design and recruitment procedures were followed) do not allow comparisons.

So, studies on student and self-selection bias, taken together suggest that studying the representativeness of subjects’ social behavior requires the *simultaneous* examination of student bias within both volunteers and non-volunteers *and* self-selection bias within both students and non-students.

Using the  $2 \times 2$  factorial design depicted in *Figure 1a*, we report data from a large-scale survey-experiment that allows such a *ceteris paribus* investigation of student and self-selection bias.

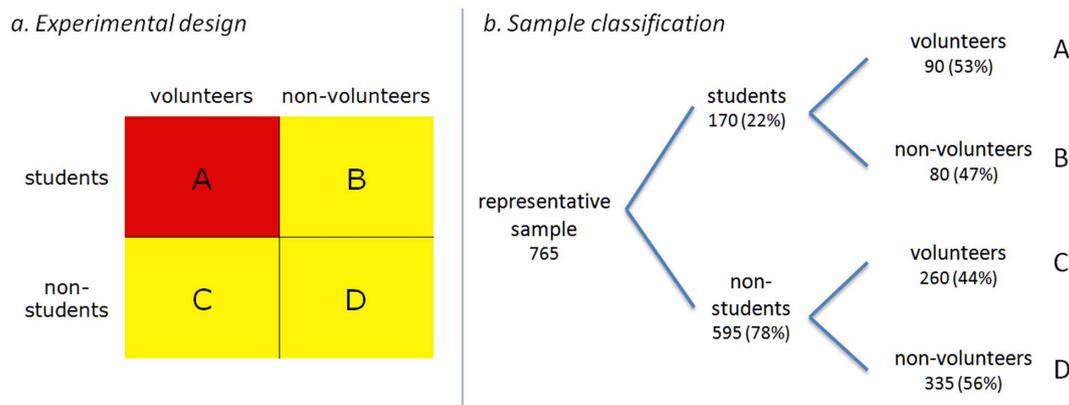
A representative sample of a city’s adult population participated in three experimental games (Dictator Game (DG), Ultimatum Game (UG), and Trust Game (TG)) involving five decisions (see *Figure 2*). In addition, a rich socio-demographic set of information was gathered in order to serve as controls, which are necessary in order to isolate student and self-selection effects. Lastly, each individual was classified as a volunteer or non-volunteer based on their willingness to participate in future experiments in the laboratory (see Methods). Our final sample ( $N = 765$  after excluding incomplete observations) therefore consists of both students and non-students as well as both volunteers and non-volunteers (see *Figure 1b*).

## Results

As *Figure 1b* illustrates, our final sample consists of:

- 22% students ( $n = 170$ ).
- 46% volunteers ( $n = 350$ ).
- 12% “standard” subject pool (*students x volunteers*) ( $n = 90$ ).

The first models (left-hand side) in each column of *Table 1* report the estimated main effects of being a *student* and a *volunteer* on behavior. The second models explore the interaction effects of the two (*student*  $\times$  *volunteer*). These models allow student bias to be studied separately within volunteers and non-volunteers and in the same manner, self-selection bias within students and non-students. The regressions in columns *i*, *ii*, and *iii* model participants’ offers in the DG, the UG and the difference between the two (thus capturing strategic behavior) respectively. Columns *iv*, *v*, and *vi* repeat the same exercise for the minimum acceptable offer (MAO) as a second mover in the UG, the decision to pass money or not in the binary TG, and the decision to return money or not as a second mover in the same game, respectively. Note that in all regressions we control for basic socio-demographics (age, sex, income and educational level) as well as for risk and time preferences, cognitive abilities and social capital as possible confounding factors.



**Figure 1** | Experimental design and sample classification.

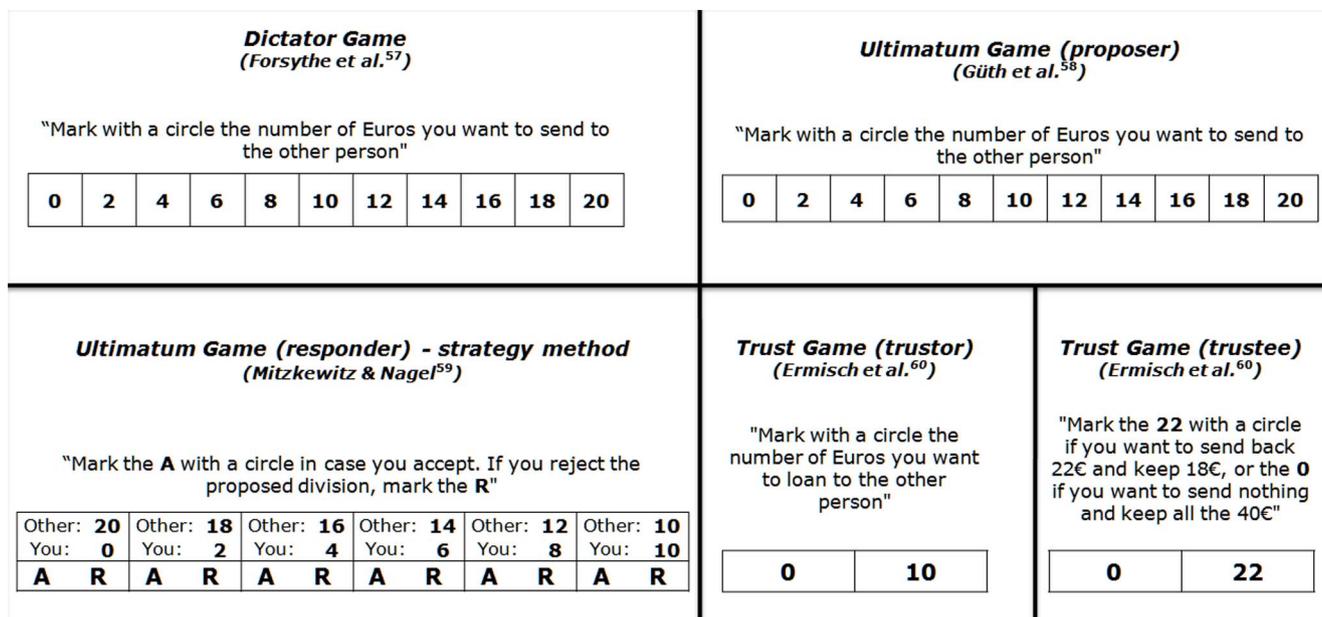


Figure 2 | Experimental decisions<sup>57-60</sup>.

Table 2 reports the coefficient estimates from the between-group comparisons obtained by the corresponding Wald tests on Table 1 models.

**Student bias:** Students are more strategic players ( $p = 0.012$ ) mostly because they make less generous DG offers ( $p = 0.060$ ). However, these differences are never larger than 6% of the pie. Through Wald tests, we identify the student bias to be mainly manifested among volunteers (A vs. C,  $p = 0.028$ ; see Table 2).

**Self-selection bias:** Volunteers are more likely to both trust (6.6%, marginal effects corresponding to the probit estimates reported in Tables 1 and 2) and reciprocate the trust (7.7%) than non-volunteers in the TG ( $p = 0.051$  and  $p = 0.011$ , respectively). However, the first difference vanishes when making pairwise comparisons within groups. That is, the aggregate effect is not specifically attributable to either students (A vs. B) or non-students (C vs. D) ( $p > 0.12$  in both cases). The second difference can be essentially traced back to non-students ( $p = 0.023$ ) since it is largely insignificant for students ( $p = 0.440$ ). Nonetheless, self-selection bias slightly affects students as well: self-selected students make (marginally) significantly higher offers than the rest of students in the UG ( $p = 0.084$ ).

As a final exercise we compare self-selected students with both the rest of the sample (A vs. B + C + D) and group D, which comprises non-students, non-volunteers as an estimation of the *subject-pool*

*bias*. We find the behavior of group A to be different from the rest of the sample only regarding UG offers, and at marginally significant levels ( $p = 0.092$ ), as they offer €0.66 more (3.3% of the pie). As can be inferred from Table 2, this effect must be emanating from the self-selection bias revealed in this decision among students. The comparison between groups A and D yields only one (marginally) significant result as well. Self-selected students increase their offers between DG and UG by €0.94 more than non-self-selected, non-students ( $p = 0.094$ ). This effect makes sense as well since students have been reported previously to be more strategic players than non-students (A + B vs. C + D). Finally, since self-selection was revealed to be an issue only among non-students (C vs. D), the absence of significant differences in TG behavior ( $ps > 0.49$ ) is not surprising.

Due to the complex interpretation of non-linear interaction effects<sup>53</sup>, we replicate the regressions of columns *iv*, *v*, and *vi* using one dummy for each group (A, B, C, and D). The results remain exactly the same. Additionally, replication of the regressions using alternative classification of students does not alter the general picture (see Methods and Tables S2 - S4 in the supplementary materials).

## Discussion

This paper presents data that allows disentangling the separate effects of student and self-selection bias. Evidence for both is found.

Table 1 | Student and self-selection biases on behavior

	DG <i>i</i>		UG <i>ii</i>		UG-DG <i>iii</i>		MAO <i>iv</i>		TG trustor <i>v</i>		TG trustee <i>vi</i>	
<i>students</i>	-0.060*	-0.067	0.007	-0.006	0.054**	0.047	-0.039	-0.079	-0.167	-0.242	-0.083	-0.034
	(0.032)	(0.044)	(0.015)	(0.021)	(0.021)	(0.030)	(0.105)	(0.165)	(0.152)	(0.198)	(0.143)	(0.191)
<i>volunteers</i>	0.039	0.036	0.023	0.016	-0.010	-0.013	0.019	0.000	0.196*	0.159	0.239**	0.266**
	(0.026)	(0.024)	(0.015)	(0.016)	(0.019)	(0.019)	(0.092)	(0.112)	(0.101)	(0.103)	(0.094)	(0.117)
<i>students × volunteers</i>		0.013		0.027		0.013		0.077		0.149		-0.096
		(0.052)		(0.027)		(0.039)		(0.201)		(0.259)		(0.268)
<i>R</i> <sup>2</sup>					0.0941	0.0943	0.0223	0.0224	0.0600	0.0604	0.1012	0.1013
<i>LR</i>	3.80***	3.79***	1.46**	1.46**	5.81***	5.68***	56.02***	56.60***	78.49***	81.52***	98.87***	98.20***

Notes: The dependent variables are (i) the fraction offered in DG; (ii) the fraction offered in UG; (iii) the fraction offered in UG - the fraction offered in DG; (iv) the minimum acceptable offer as a fraction of the pie in UG; (v) TG decision as a trustor - 1 if [s]he makes the loan, zero otherwise; and (vi) TG decision as a trustee - 1 if [s]he returns part of the loan, zero otherwise. Models *i* and *ii* are Tobit regressions, model *iii* is an OLS regression; model *iv* is an ordered probit regression, while the last two models are Probit regressions.  $N = 765$  in all regressions. Controls are: age, gender, education, household income, social capital, risk preferences, time preferences, and cognitive abilities. The variables are explained in depth in the supplementary materials. All models are also controlling for order effects. All the likelihood ratios (LR) shown correspond to  $Chi^2$  statistics, except for column *iii*, where they are based on  $F$ . Robust SE clustered by interviewer (108 groups) and presented in brackets. \*, \*\*, \*\*\* indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.



Table 2 | Between-group comparisons

			DG	UG	UG-DG	MAO	TG trustor	TG trustee
Student bias								
(A + B)	vs	(C + D)	-0.060*	0.008	0.054**	-0.039	-0.168	-0.083
A	vs	C	-0.031	0.021	0.061**	-0.002	-0.093	-0.130
B	vs	D	-0.068	-0.007	0.047	-0.079	-0.242	-0.034
Self-selection bias								
(A + C)	vs	(B + D)	0.040	0.023	-0.010	0.020	0.197*	0.240**
A	vs	B	0.051	0.044*	0.000	0.078	0.309	0.170
C	vs	D	0.037	0.017	-0.013	0.001	0.159	0.266**
Subject-pool bias								
A	vs	(B + C + D)	-0.012	0.033*	0.039	0.021	0.080	0.049
A	vs	D	-0.017	0.038	0.047*	-0.002	0.067	0.136

Notes: Letters A, B, C and D refer to the groups depicted in Figure 1a. Group A denotes students, volunteers; B students, non-volunteers; C non-students, volunteers; D non-students, non-volunteers. (A + B) correspond to all students (volunteers and non-volunteers); (C + D) to all non-students (volunteers and non-volunteers); (A + C) to all volunteers (students and non-students); (B + D) to all non-volunteers (students and non-students). Lastly (B + C + D) correspond to the sum of the subject pool except students volunteers. \*, \*\* indicate significance at the 0.10, and 0.05 levels, respectively. Comparisons based on Wald tests from models of Table 1.

However, the results also tell another parallel story: in five experimental decisions and following the exact same procedures for all subjects, self-selected students have been proven to behave in a very similar manner with every other group separately and in combination. Indeed, at the conventional 5% level only one significant effect concerning self-selected students is observed and, in addition, the difference is economically small. That said, we suggest that the findings do not discredit the use of self-selected students in experiments measuring social preferences. Rather the opposite: the convenient sample of self-selected college students that allowed a boom in human experimentation in both social and natural sciences produces qualitatively and quantitatively accurate results. Models on human social behavior, evolutionary dynamics and social networks together with the implications that they bare are not in danger from this particular subject pool. The results caution, however, on the use of alternative samples such as self-selected non-students that typically participate in artefactual field and internet experiments, aimed at better representativeness, since the effect of self-selection can be even more pronounced outside the student community (self-selection bias is proved to be an issue mainly among non-students in the Trust Game).

## Methods

The experiment took place from November 23<sup>rd</sup> to December 15<sup>th</sup> 2010. A total of 835 individuals aged between 16 and 91 years old participated in the experiment. One out of ten participants was randomly selected to be paid. The average earnings among winners, including those winning nothing (18.75%), were €9.60.

**Sampling.** A stratified random method was used to obtain the sample. In particular, the city of Granada (Spain) is divided into nine geographical districts, which served as sampling strata. Within each stratum we applied a proportional random method to minimize sampling errors. In particular, the sample was constructed in four sequential steps: 1. We randomly selected a number of sections proportional to the number of sections within each district; 2. We randomly selected a number of streets proportional to the number of streets within each section; 3. We randomly selected a number of buildings proportional to the number of buildings on each street; 4. Finally, we randomly selected a number of apartments proportional to the number of apartments within each building. This method ensures a *geographically representative sample*. Detailed information can be found in supplementary materials.

Our sample consists of individuals who agreed to complete the survey at the moment the interviewers asked them to participate. Being interviewed in their own apartments decreased opportunity cost (thus increasing the participation rate). In order to control for selection bias within households, only the individual who opened the door was allowed to participate. Lastly, the data collection process was well distributed across both daytime and weekday. Our sampling procedure resulted in a representative sample in terms of age and sex (see Table S7 in the supplementary materials).

**Interviewers.** The data were collected by 216 university students (grouped in 108 pairs) enrolled in a course on field experiments in the fall of 2010. The students underwent ten hours of training in the methodology of economic field experiments, conducting surveys, and sampling procedures. Their performance was carefully monitored through a web-based system (details in the supplementary materials).

**Protocol.** The interviewers introduced themselves to the prospective participants and explained that they were carrying out a study for the University of Granada. Upon agreement to participate, the participants were informed that the data would be used for scientific purposes only and under conditions of anonymity according to the Spanish law on data protection. One interviewer always read the questions aloud, while the other noted down the answers (with the exception of the experimental decisions). The survey lasted on average 40 minutes and consisted of three parts. In the first part, extensive socioeconomic information of the participants was collected including, among others, risk and time preferences, and social capital. In the second part, participants played three paradigmatic games of research on social preferences, namely the Dictator Game, the Ultimatum Game and the Trust Game (see Figure 2). In the last part, they had to state their willingness to participate in future monetary-incentivized experiments (which would take place in the laboratory at the School of Economics).

**Experimental games.** At the beginning of the second part, and before any details were given about each decision in particular, the participants received some general information about the nature of the experimental economic games according to standard procedures. In particular, participants were informed that:

- The five decisions involved real monetary payoffs coming from a national research project endowed with a specific budget for this purpose.
- The monetary outcome would depend only on the participant's decision or on both his/her own and another randomly matched participant's decision, whose identity would forever remain anonymous.
- One of every ten participants would be randomly selected to be paid, and the exact payoff would be determined by a randomly selected role. In deciding 1/10 instead of higher probabilities (for instance 1/5), we took into account two issues: the cognitive effects of using other probabilities and the (commuting) costs of paying people given the dispersion of participants throughout the city. Interestingly, 297 subjects (39% of the sample) believed that they would be selected to be paid (last item of the second part).
- Matching and payment would be implemented within the next few days.
- The procedures ensured absolute double-blinded anonymity by using a decision sheet, which they would place in the envelope provided and then seal. Thus, participants' decisions would remain *forever* blind in the eyes of the interviewers, the researchers, and the randomly matched participant.

Once the general instructions had been given, the interviewer read the details for each experimental decision separately. After every instruction set, participants were asked to write down their decisions privately and proceed to the next task. To control for possible order effects on decisions, the order both between and within games was randomized across participants, resulting in 24 different orders (always setting aside the two decisions of the same game).

In the Dictator and Ultimatum Game (proposer) participants had to split a pie of €20 between themselves and another anonymous participant. Subjects decided which share of the €20 they wanted to transfer to the other participant. In the case of the Ultimatum Game, implementation was upon acceptance of the offer by the randomly matched responder; in case of rejection neither participant earned anything. For the role of the responder in the Ultimatum Game we used the strategy method in which subjects had to state their willingness to accept or reject each of the proposals depicted in Figure 2. In the Trust Game, the trustor (1<sup>st</sup> pl.) had to decide whether to pass €10 or €0 to the trustee (2<sup>nd</sup> pl.). In case of passing €0, the trustor earned €10 and the trustee nothing. If she passed €10, the trustee would receive €40 instead of €10 (money was being quadrupled). The trustee, conditional on the trustor having passed the money had to decide whether to send back €22 and keep €18 for himself or keep all €40 without sending anything back, in which case the trustor did not earn anything (see the supplementary materials).



**Classifying students.** Individuals between 18 and 26 years old who reported to be studying at the moment were classified as students. The upper age bound (26 years old) was selected taking into account the mean maximum age of the lab experiments taken place in the University of Granada and a large drop in the age histogram of our sample. In order to address potential concerns regarding this classification, alternative ways of classifying students were used. In particular we replicated the analysis setting the upper bounds at 24 and 28 years old. Moreover, we did the same classifying as “students” all individuals who have ever been in the university, without posing any age limit whatsoever. Results in the three cases remained the same in essence. The regressions can be found in the supplementary materials.

**Classifying volunteers.** Following Van Lange et al.<sup>54</sup> in their application of the measure developed by McClintock and Allison<sup>55</sup>, we classified participants according to the response to the following question:

“At the School of Economics we invite people to come to make decisions with real money like the ones you made earlier (the decisions in the envelope). If we invite you, would you be willing to participate?”

Note, however that we have intentionally removed any helping framing. Van Lange et al. (54, pg. 281) for example first stated: “the quality of scientific research of psychology at the Free University depends to a large extent on the willingness of students to participate in these studies” and then proceeded in asking them their willingness to participate in future studies. It is also important to mention that the willingness to participate in future experiments was stated before matching between participants and payments were done. So, by design, the variable of interest could not have been affected by the outcome of the games.

Furthermore, in order to differentiate self-selection in economic experiments from the general propensity to help research studies and the need for social approval (see 25), we also asked individuals about their willingness to participate in future surveys. A total of 478 stated that they would be willing to participate in future surveys, while only 350 said they would participate in experiments. Of these, 49 stated that they would not participate in a survey. In addition, two months after the experiment, we hired an assistant to call all the individuals classified as volunteers in order to confirm their interest. In particular, we requested participants’ authorization to include their data in the experimental dataset of the Economics Department (ORSEE)<sup>56</sup>. Of those who we were able to contact after two attempts on two consecutive days (60%), 97% of students and 83% of non-students confirmed their interest. Not answering the phone makes sense if we consider the enormous amount of telemarketing calls people receive in Spain and even more so given that the assistant made calls from a university phone number which is comprised of 13 digits like those of telemarketing companies. Note that regular private numbers in Spain have 9 digits.

This method of classifying volunteers raises some concerns. In particular, the stated preference regarding the willingness to participate in future experiments is never realized. Despite our attempts to ensure that this was not just cheap talk (by being granted permission to add individuals’ personal details in ORSEE) the matter of the fact is that we do not know with certainty whether those classified as volunteers are indeed volunteers. Actually, completely separating volunteers and non-volunteers is a virtually impossible task. The very idea of volunteering is a continuous quality instead. However, by definition, classification requires a line to be drawn. We believe that this classification method provides a rather clean way to separate ‘more’ self-selected from ‘less’ self-selected individuals.

A second concern is related to the fact that our sample consists of only individuals who had accepted to fill in a survey. In other words it seems that we study self-selection using an already self-selected sample. Note however that individuals have been self-selected into filling in a survey and *not* into participating in a lab experiment. In addition our procedures decreased opportunity costs for participants minimizing this type of self-selection. So, individuals had to fill in the questionnaire in the comfort of their houses and without any ex-ante commitment for the future, in contrast to most nation-wide surveys (CentER, SOEP, BHPS, etc.). Actually, 38% of the participants were unwilling to participate in a future survey while 54% were not willing to participate in a lab experiment. This allowed us to observe *experimental behavior of people not willing to participate in lab experiments*, playing with real money and what is more doing so voluntarily.

Of course it can still be true that we are missing one “extreme” category; those who had refused participation in the survey in the first place. Even in this case however, if self-selection does indeed affect behavior, it should do so even in the absence of this extreme category.

**Ethics statement.** All participants in the experiments reported in the manuscript were informed about the content of the experiment before to participate (see Protocol). Besides, their anonymity was always preserved (in agreement with the Spanish Law 15/1999 for Personal Data Protection) by assigning them randomly a numerical code, which would identify them in the system. No association was ever made between their real names and the results. As it is standard in socio-economic experiments, no ethic concerns are involved other than preserving the anonymity of participants.

This procedure was checked and approved by the Vicedean of Research of the School of Economics of the University of Granada, the institution hosting the experiment.

1. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **289**, 850–852 (2000).

2. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the ‘tragedy of the commons’. *Nature* **415**, 424–426 (2002).
3. Semmann, D., Krambeck, H. J. & Milinski, M. Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature* **425**, 390–393 (2003).
4. Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. Winners don’t punish. *Nature* **452**, 348–351 (2008).
5. Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H. J. & Milinski, M. Strategy updating in evolutionary games. *Proc. Natl. Acad. Sci.* **107**, 2962–2966 (2010).
6. Rand, D. G. & Nowak, M. A. The evolution of antisocial punishment in optional public goods games. *Nature Commun.* **2**, 434 (2011).
7. Crone, E. A., Somsen, R. J. M., Beek, B. V. & Van Der Molen, M. W. Heart rate and skin conductance analysis of antecedents and consequences of decision making. *Psychophysiology* **41**, 531–540 (2004).
8. Li, J., McClure, S. M., King-Casas, B. & Montague, P. R. Policy adjustment in a dynamic economic game. *PLoS ONE* **1**, e103 (2006).
9. Van den Bergh, B. & Dewitte, S. Digit ratio (2D : 4D) moderates the impact of sexual cues on men’s decisions in ultimatum games. *P. Roy. Soc. Lond. B. Bio.* **273**, 2091–2095 (2006).
10. van’t Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* **169**, 564–568 (2006).
11. Burnham, C. T. High-testosterone men reject low ultimatum game offers. *P. Roy. Soc. Lond. B. Bio.* **274**, 2327–2330 (2007).
12. Chapman, H. A., Kim, D. A., Susskind, J. M. & Anderson, A. K. In bad taste: evidence for the oral origins of moral disgust. *Science* **323**, 1222–1226 (2009).
13. Elliott, R., Friston, K. J. & Dolan, R. J. Dissociable neural responses in human reward systems. *J. Neurosci.* **20**, 6159–6165 (2000).
14. Breiter, H. C., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30**, 619–639 (2001).
15. O’Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornack, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4**, 95–102 (2001).
16. Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S. & Kilts, C. D. A neural basis for social cooperation. *Neuron* **35**, 395–405 (2002).
17. Sanfey, G. A. Social decision-making: insights from game theory and neuroscience. *Science* **318**, 598–602 (2007).
18. Lee, D. D. Game theory and neural basis of social decision making. *Nat. Neurosci.* **11**, 404–409 (2008).
19. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
20. Roca, J., Cuesta, A. & Sánchez, A. Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics. *Physics of Life Reviews* **6**, 208–249 (2009).
21. Gruijic, J., Fosco, C., Araujo, L., Cuesta, J. A. & Sánchez, A. Social experiments in the mesoscale: humans playing a spatial Prisoner’s Dilemma. *PLoS ONE* **5**, e13749 (2010).
22. Perc, M. & Szolnoki, A. Coevolutionary games - a mini review. *BioSystems* **99**, 109–125 (2010).
23. Suri, S. & Watts, D. J. Cooperation and contagion in Web-based, networked public goods experiments. *PLoS ONE* **6**, e16836 (2011).
24. Garcia-Lázaro, C. et al. Heterogeneous networks do not promote cooperation when humans play a Prisoner’s Dilemma. *Proc. Natl. Acad. Sci.* **109**, 12922–12926 (2012).
25. Levitt, S. D. & List, J. A. What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **21**, 153–174 (2007).
26. Levitt, S. D. & List, J. A. Homo economicus evolves. *Science* **319**, 909–910 (2008).
27. Falk, A. & Heckman, J. Lab experiments are a major source of knowledge in the social sciences. *Science* **326**, 535–38 (2009).
28. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–135 (2010).
29. Janssen, M. A., Holahan, R., Lee, A. & Ostrom, E. Lab experiments for the study of social-ecological systems. *Science* **328**, 613–617 (2010).
30. Paolacci, G., Chandler, J. & Ipeirotis, P. G. Running experiments on amazon mechanical turk. *Judgment and Decision Making* **5**, 411–419 (2010).
31. Rand, D. G. The promise of mechanical turk: how online labor markets can help theorists run behavioral experiments. *J. Theor. Biol.* **299**, 172–179 (2011).
32. Cooper, D., Kagel, J. H., Lo, W. & Gu, Q. L. Gaming against managers in incentive systems: experiments with Chinese managers and Chinese students. *Amer. Econ. Rev.* **89**, 781–804 (1999).
33. Fehr, E. & List, J. A. The hidden costs and returns of incentives—trust and trustworthiness among CEOs. *Journal of the European Economic Association* **2**, 743–771 (2004).
34. Haigh, M. S. & List, J. A. Do professional traders exhibit myopic loss aversion? *J. Finance* **60**, 523–534 (2005).
35. Cárdenas, J. C. Groups, commons and regulations: experiments with villagers and students in Colombia. In *Psychology, Rationality and Economic Behavior: Challenging Standard Assumptions*, eds. Agarwal, B. & Vercelli, A. pp. 242–270. Palgrave, London (2005).
36. Palacios-Huerta, I. & Volij, O. Field centipedes. *Amer. Econ. Rev.* **99**, 1619–1635 (2009).
37. Harrison, G. W. & List, J. A. Field experiments. *J. Econ. Lit.* **92**, 1009–1055 (2004).



38. Harrison, G. W., Lau, M. I. & Williams, M. B. Estimating individual discount rates in Denmark: a field experiment. *Amer. Econ. Rev.* **92**, 1606–1617 (2002).
39. Fehr, E., Fischbacher, U., von Rosenblatt, B., Schupp, J. & Wagner, G. A nationwide laboratory examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *Schmollers Jahrbuch* **122**, 519–542 (2003).
40. Gächter, S., Herrmann, B. & Thöni, C. Trust, voluntary cooperation and socio-economic background: survey and experimental evidence. *J. Econ. Beh. Organ.* **55**, 505–531 (2004).
41. Bellemare, C., Kröger, S. & van Soest, A. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* **76**, 815–839 (2008).
42. Egas, M. & Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *P. Roy. Soc. Lond. B. Bio.* **275**, 871–878 (2008).
43. Dohmen, T., Falk, A., Huffman, D. & Sunde, U. Are risk aversion and impatience related to cognitive ability? *Amer. Econ. Rev.* **100**, 1238–1260 (2010).
44. Carpenter, J. P., Burks, S. & Verhoogen, E. Comparing students to workers: the effects of stakes, social framing, and demographics on bargaining outcomes. In *Field Experiments in Economics*, eds. Carpenter, J., Harrison, G. and List, J. A. pp. 261–290, JAI Press, Stamford, CT (2005).
45. Carpenter, J. P., Connolly, C. & Myers, C. Altruistic behavior in a representative dictator experiment. *Exper. Econ.* **11**, 282–298 (2008).
46. Burks, S., Carpenter, J. P. & Goette, L. Performance pay and worker cooperation: evidence from an artefactual field experiment. *J. Econ. Beh. Organ.* **70**, 458–469 (2009).
47. Anderson, J. *et al.* Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. *Exper. Econ.* (in press).
48. Bellemare, C. & Kröger, S. On representative social capital. *Europ. Econ. Rev.* **51**, 183–202 (2007).
49. Falk, A., Meier, S. & Zehnder, C. Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of European Economic Association* (in press).
50. Eckel, C. C. & Grossman, P. J. Volunteers and pseudo-volunteers: the effect of recruitment method in dictator experiments. *Exper. Econ.* **3**, 107–120 (2000).
51. Cleave, B. L., Nikiforakis, N. & Slonim, R. Is there selection bias in laboratory experiments? The case of social and risk preferences. *Exper. Econ.* (in press).
52. Zizzo, J. D. Experimenter Demand Effects in Economic Experiments. *Exper. Econ.* **13**, 75–98 (2010).
53. Ai, C. & Norton, E. Interaction terms in logit and probit models. *Econ. Letters* **80**, 123–129 (2003).
54. Van Lange, P. A. M., Schippers, M. & Balliet, D. Who volunteers in psychology experiments? An empirical review of prosocial motivation in volunteering. *Pers. Indiv. Differ.* **51**, 279–284 (2011).
55. McClintock, C. G. & Allison, S. T. Social value orientation and helping behavior. *J. Appl. Soc. Psychol.* **19**, 353–62 (1989).
56. Greiner, B. An online recruitment system for economic experiments. In *Forschung und wissenschaftliches Rechnen 2003*, eds. Kremer, K. & Macho, V., pp. 79–93, GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen (2004).
57. Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. Fairness in simple bargaining experiments. *Game Econ. Behav.* **6**, 347–69 (1994).
58. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econ. Beh. Organ.* **3**, 367–88 (1982).
59. Mitzkewitz, M. & Nagel, R. Experimental results on ultimatum games with incomplete information. *Int. J. Game Theory* **22**, 171–98 (1993).
60. Ermisch, J., Gambetta, D., Laurie, H., Siedler, T. & Noah Uhrig, S. C. Measuring people's trust. *J. R. Stat. Soc. Ser. A (Statistics in Society)* **172**, 749–769 (2009).

## Acknowledgments

This paper has benefitted from the comments and suggestions of Jordi Brandts, Juan Carrillo, Coralio Ballester, Juan Camilo Cárdenas, Jernej Copic, Ramón Cobo-Reyes, Nikolaos Georgantzis, Ayça Ebru Giritligil, Roberto Hernán, Benedikt Herrmann, Praveen Kujal, Matteo Migheli, Rosi Nagel and participants at seminars at ESI/Chapman, the University of Southern California, and the University of los Andes, the 2<sup>nd</sup> Southern Europe Experimentalists Meeting (SEET 2011), the VI Alhambra Experimental Workshop and the Society for the Advancement of Behavioral Economics (SABE 2012). Juan F. Muñoz designed the sampling procedure. We thank him for his professional advice. Research assistance by Ana Trigueros is also appreciated. FE acknowledges the post-doctorate fellowship granted by The Scientific and Technological Research Council of Turkey (TUBITAK). Financial support from the Spanish Ministry of Science and Innovation (ECO2010-17049), the Government of Andalusia Project for Excellence in Research (P07.SEJ.02547) and the Fundacion Ramón Areces R + D 2011 is gratefully acknowledged.

## Author contributions

All authors contributed equally to all parts of the research.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

**How to cite this article:** Exadaktylos, F., Espín, A.M. & Brañas-Garza, P. Experimental subjects are not different. *Sci. Rep.* **3**, 1213; DOI:10.1038/srep01213 (2013).