



Munich Personal RePEc Archive

# **Recruiting experimental subjects using WhatsApp**

Jorrat, Diego

Universidad Loyola Andalucía

1 July 2020

Online at <https://mpra.ub.uni-muenchen.de/101467/>  
MPRA Paper No. 101467, posted 07 Jul 2020 07:19 UTC

# Recruiting experimental subjects using WhatsApp\*

Diego Jorrat<sup>†</sup>

July 1, 2020

## Abstract

The aim of many experiments is to estimate the effect of different interventions on subjects' decision making. However, obtaining large samples and internal validity is challenging. This paper presents an alternative device at almost no cost that can easily provide a very large number of participants (700 in 5 hours). We asked 14 students to invite their WhatsApp contacts to participate in an online experiment. The students created a total of 80 diffusion groups with 25 contacts each. Using the diffusion groups as clusters, we ran a cluster randomization procedure in order to assign subjects to a framing experiment (treatment + control). We obtained the same level of attrition, duplicates and uninvited subjects across the treatment and control groups. Moreover, the experiment yielded consistent results in line with the framing literature.

**Keywords:** Recruiting, Online Experiments, Prisoner's Dilemma, Randomization.

**JEL codes:** C8, C99, D70.

---

\*I would like to thank Pablo Brañas for his unconditional guidance, comments and all his support over these years. Alfonso Robles provided excellent research assistance. Our students Carmen Hidalgo, Esteban Marquez, Guillermo Molina, Humberto Franco, Ignacio Perez, Jaime Gómez, Javier Pino, Javier Sánchez, Jimena Moreno, Luis Alcala, María Alonso, María Neira, Marta París, Miguel Villar, Miguel Fernández, Roberto Mengibar, Valeria Herrera y Vicente Moreno made this experiment possible. Funding provided by the Ministry of Spain PGC2018-093506-B-100 and the Regional Government of Andalusia PY18-FR-0007 is also gratefully acknowledged.

<sup>†</sup>*LoyolaBehLab*, Universidad Loyola Andalucía

# 1 Introduction

The cornerstone of most experiments in social sciences is the random assignment of subjects to different treatments. Depending on the size effect and the number of treatments, it is not only necessary to recruit a sufficiently large number of subjects to detect a moderate minimum detectable effect, but also to obtain comparable groups after the randomization process (balance). Inviting a sufficiently large number of subjects to participate - the recruiting process - is time consuming and costly.

Although experiments were initially conducted with paper and pencil, researchers have now developed tools to perform experiments over local computer networks where subjects receive instructions and make their decisions in a computer-based environment (Fischbacher, 2007). Using these tools, experimenters also gain control over the flow of information and thereby reduce potential confounding factors (Horton et al., 2011). To conduct lab experiments, however, subjects are typically brought into physical laboratories despite the fact that many of these experiments can be conducted remotely (online).

Online participation in experiments has several advantages. Firstly, costs can be notably reduced since experimental subjects do not need to be paid travel expenses. Participants also have more freedom to choose a schedule that best suits them.<sup>1</sup> Additionally, online samples allow for larger and longer studies with potentially more diverse subjects (Horton et al., 2011). This is important because larger samples are critical to have sufficient statistical power to measure treatment effects. Recently, online labor markets such as Amazon Mechanical Turk (MTurk) or Prolific Economics enable researchers to recruit a sufficiently large number of subjects and randomly assign participants to different treatments. These platforms connect researchers to experimental subjects, who are paid to complete the tasks on the computer (Rand, 2012).

In this paper, we present an alternative and practically cost-free device that can be used to recruit a huge number of participants and provide a larger number

---

<sup>1</sup>In the case of simultaneous games, coordination between participants is necessary.

of responses (1000 in 5 hours). To do so, we asked 14 students to divide their WhatsApp contacts into different diffusion groups with 25 members each and subsequently invite their contacts to participate in an online experiment.

To ensure a perfect randomization, we assigned 80 diffusion groups to the treatment and the control following a cluster randomization procedure. A total of 1979 messages were sent to the subjects, 1005 to the control group and 974 to the treatment group. Only 37% of the invited subjects participated. The final sample comprised a total of 737 subjects and was divided in the following way: 387 in the treatment and 350 in the control group. The attrition level was similar in both groups (64% in the control group and 61% in the treatment group). Both samples were similar in terms of their observable characteristics. On average, each student sent 141 invitations which produced 53 responses (37%).

Our procedure to recruit subjects achieved four major goals in terms of internal validity. First, we obtained two comparable groups with different observable characteristics. Second, our procedure led to the same level of attrition and the same percentage of duplicates (subjects who performed the experiment more than once). Third, our procedure allowed us to recruit subjects who were not invited to participate in the experiment. These subjects were evenly distributed across the groups and did not affect the between-group comparability. Lastly, our procedure allowed us to recruit nearly 800 subjects in less than 12 hours. This is a very fast procedure that reduces the probability of subjects interacting with each other and thus contaminating the results.

To check the external validity of the procedure, we ran a framed Prisoner's Dilemma experiment. Our results are in line with the framing literature on the Prisoner's Dilemma Game (PDG).

Overall, the results presented in this paper suggest that our procedure worked properly. A device like this is of particular interest for researchers with limited funding and who do not have access to proper labs and funding.

The paper is structured as follows. The recruitment and randomization procedure is presented in Section 2 and the framed PDG experiment is presented

in Section 3. The results of the randomization and the degree of comparability of the treatment and control groups are discussed in Section 4. The presence of different threats to the internal validity are analyzed in Section 5. The results of the experiment are discussed in Section 6 in line with the framing literature. Finally, section 7 concludes.

## 2 Recruiting procedure

The goal of experimental research is to estimate whether subjects are sensitive to treatments. To obtain good estimators of these effects it is necessary to design a procedure that allows collecting sufficiently large samples. Additionally, subjects should be assigned to treatments that guarantee orthogonality with potential confounders. Sufficiently large samples ensure that enough statistical power is available to find moderate treatment effects, while orthogonality ensures internal validity. In this section we present the randomization procedure in order to solve both problems at once.

In order to reach a large number of subjects, we asked students to invite their WhatsApp contacts to participate in an online experiment. Subjects under the age of 16 were not allowed to participate.<sup>2</sup> Because assigning each WhatsApp contact to the treatment might be time consuming and, more importantly, be subject to human error due to inattention, we used a cluster randomization design.

We asked the students to generate different diffusion groups (with  $k=25$ ) following the alphabetical order of their contacts. Alphabetical ordering (as a quasi-natural or natural experiment) might be a reasonable solution when individual randomization is impossible. Indeed, there is evidence in the literature that alphabetical ordering is adequate.<sup>3</sup>

---

<sup>2</sup>Participants aged 16 and 17 years old (a total of 16 in our sample) can give their consent without the need for parental authorization (Article 8 and Recitals 38 and 58 of Directive 95/46/EC).

<sup>3</sup>In Miguel and Kremer (2004), school children were assigned to a deworming treatment by alphabetical order. This was also the case of the experiment of Glewwe et al. (2004) who used flip charts in schools. Both papers found that the control and treatment groups were

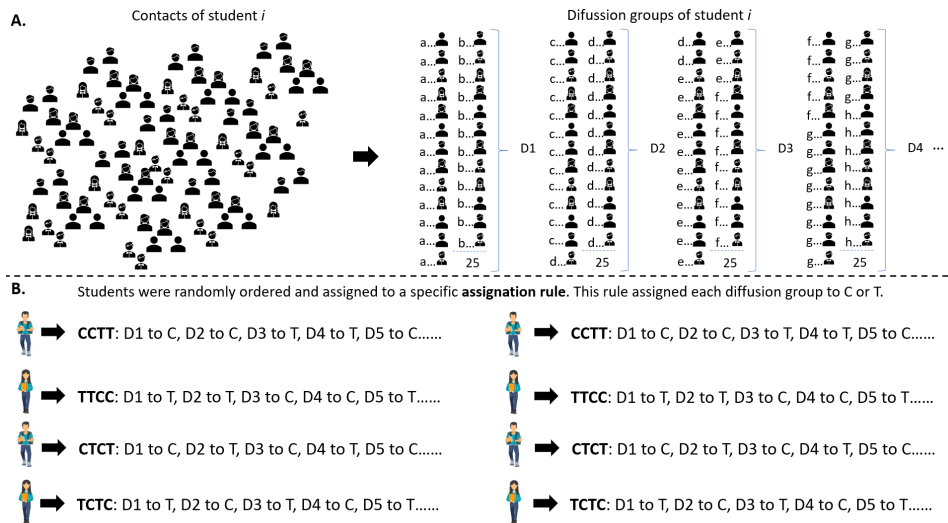


Figure 1: Randomization process

Each of the students created four diffusion groups: D1, D2, D3 and D4. Each of them then assigned their first 25 contacts to D1, the next 25 contacts to D2, and so on. After contact 100 they repeated the process again: 101-125 to D1, 126-150 to D2, and so on. The procedure is represented graphically in Panel A of Figure 1. As can be observed, each diffusion group is represented by a cluster. It is important to note that the cut-off of 25 contacts was selected by the researchers (and not by the students) and was independent of each student's number of contacts. All the students were asked to report the number of contacts as well as the number of diffusion groups they had created.

The students were then provided an assignment rule (where C refers to the control group and T to the treatment group): CCTT, TTCC, CTCT or TCTC. The students who received CTCT sent invitations to all the contacts included in D1 and D3 to participate in the control, as well as invitations to the members of

balanced before the interventions. Moreover, in Jorrat et al. (2018), the participation rule for police training was the alphabetical order of the officers' last names. The authors found no significant differences between the control and treatment groups regarding their socioeconomic characteristics and participation in violent events. However, this assignment rule is not perfect because it does not guarantee a balance in unobservable characteristics (for further discussion, see Deaton, 2010)

D2 and D4 to participate in the treatment group. Each student was randomly assigned a different profile (CCTT, TTCC, CTCT or TCTC) to ensure that the assignation was independent of the student and the contacts' characteristics. The assignment procedure is illustrated in Panel B of Figure 1. It is important to mention that we did not send reminders and we did not use WhatsApp groups.<sup>4</sup>

### 3 The experiment

We ran an experiment to test the internal and external validity of the randomization procedure. The experiment resembles that of the original paper of Hoffman et al. (1996) on the use of certain words which may trigger pro-social behavior (see also Brañas-Garza, 2007).

The aim of the experiment was to study framing effects in a one-shot PDG. To do so, we ran one treatment: subjects in the treatment group faced the PDG with the strategies "Cooperate" or "Not cooperate", while the students in the control group faced neutral-framed options "A" or "B", where A and B represent the cooperative and the non-cooperative strategy, respectively.<sup>5</sup> These labels may create an experimenter demand effect by providing cues about what constitutes appropriate behavior (Capraro and Rand, 2018; Zizzo, 2010). We also examined whether framing has a differential impact on participants endowed with certain characteristics. In particular, we studied the interaction between framing and preferences for inequity.

Figure 2 presents the PDG faced by subjects in the control (panel A) and the treatment (panel B) group. It also shows the payoffs in euros that the individual would obtain depending on another player's decision. The experiment was conducted in two Google Forms: one for the control group and another for the treatment.

The experimental setup also included: a) informed consent; b) contact num-

---

<sup>4</sup>Sending invitations through groups may jeopardize the randomization procedure.

<sup>5</sup>For more on the instructions, see Section 2 of the Supplementary Information.

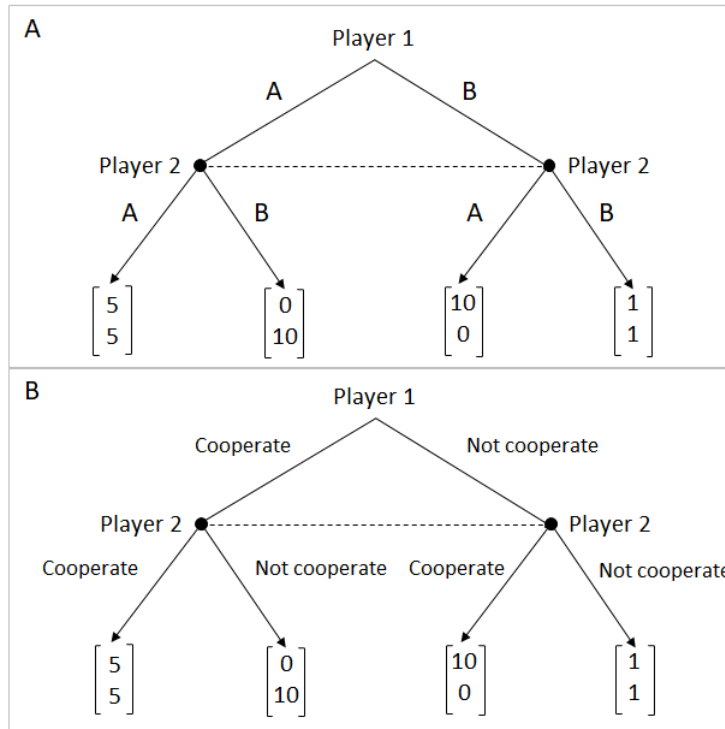


Figure 2: Prisoner's dilemmas in the decision tree scheme for the control (panel A) and treatment group (panel B).

ber (to pay if the participant was randomly selected to be paid); c) name of the student who sent the invitation; d) gender and e) self-reported inequity aversion. The instructions explicitly stated that 10 participants would be randomly chosen to be paid and that the potential earnings ranged from €0, €1, €5 to €10.

As mentioned above, we also elicited self-reported solidarity and envy. These variables refer to people's aversion to advantageous or disadvantageous inequality as in the model of Fehr and Schmidt (1999) (for more details, see Section B of the Supplementary Information). Using a 10-item Likert scale, we asked participants if they agreed with the statement *"I do not care about how much money I have; what concerns me is that there are people who have less (more, for envy) money than I have"*. These variables are associated with donations in



dictator games and cooperation in PDG (Espín et al., 2018, 2019).

In the following section, we test if the entire set of collected variables were balanced between the control and treatment groups.

## 4 Results of the randomization procedure

Figure 3 summarizes the results of the recruiting procedure. We invited 1979 contacts using WhatsApp. A total of 1005 contacts received the invitation to the control group (C) and 974 contacts to the treatment group (T). Although the distributions between the treatment and the control group were well balanced (51% in C and 49% in T), not all the invited contacts responded. The response rate was 37%. In the literature on survey research there is no agreed-upon standard for a minimum acceptable response rate (Fowler Jr, 2013). However, our response rate (about 30%) was higher than the average response rate for online surveys reported in the literature (Millar and Dillman, 2011; Kaplowitz et al., 2004; Dillman, 2011). Given that the invitation was sent only once (with no reminders), the procedure proved to work well. This left us with a total of 737 observations, of which 387 were distributed in C (53%) and 350 in T (47%).

The overall attrition rate was 63%. Because we had no information on the non-respondents' characteristics, we could not check if the respondents differed from the non-respondents in each group. However, we checked if the attrition rates differed between the control and the treatment, and found no evidence of differential attrition (two-sample test of proportions,  $p = 0.116$ ).

We also tested whether the control and the treatment were balanced in terms of observable characteristics. Table 1 provides the results of the mean comparison tests. We only found that both groups are marginally different in solidarity and disposition to participate in future experiments. However, these differences are weakly significant ( $p > 0.05$  in both cases). Therefore, we conclude that the randomization process worked properly.

To sum up, this procedure allowed us to solve three potential problems in experiments. First, we obtained nearly 750 responses in just half a day. Second,

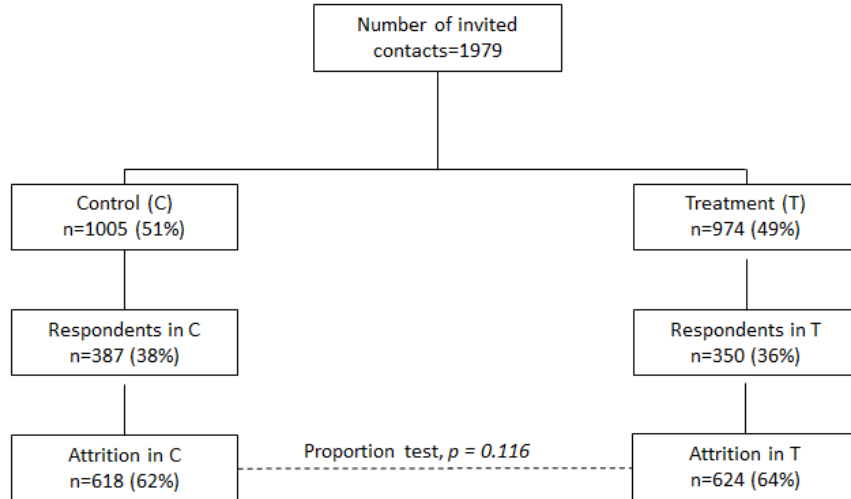


Figure 3: Randomization results

Table 1: Balance between control and treatment.

	Control	Treatment	$T - C$	$p - value$
Female	0.447	0.440	-0.007	0.848
Student	0.832	0.840	0.008	0.771
Solidarity	4.778	5.126	0.348*	0.070
Envy	2.948	3.045	0.097	0.558
Future participation	0.553	0.614	0.061*	0.092

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

we ensured independence between the students and the assigned treatments because they could not change the treatment status and the cut-off in the diffusion groups was exogenous for them. Third, random assignment by diffusion groups guaranteed a balance in the observable and unobservable factors between the control and the treatment. Taking all this together, our procedure ensures internal validity.

## 5 Threats to the randomization procedure

In this section, we present the different threats that arose during the implementation of the experiment.

### Duplicates and contamination

One potential problem of this randomization procedure is the presence of duplicates, that is, subjects who completed the experiment more than once.<sup>6</sup> In fact, playing twice is very sensible since it doubles the probability of getting paid. The same subject may also receive an invitation from different students; a problem that is even more concerning if invitations come from both the control and the treatment group.

We were able to identify duplicates since we asked subjects to indicate their telephone number in order to be notified if they were randomly selected to be paid. Using phone numbers, we identified two types of duplicates: a) subjects who did the experiment twice but did not change the treatment; and b) those who did the experiment twice but changed their treatment. This information is summarized in Table 2. A total of 15 duplicates were identified in each group. The percentage was 3.9% and 4.3% in C and T, respectively, and the proportions were not statistically different (two-sample test of proportions,  $p > 0.3$ ). Regarding the second type, only 2 subjects (0.5%) in the control and 3 in the treatment group (0.9%) changed their treatment status. These subjects accounted for 0.5% and 0.9% of the total, respectively. Again the proportions were not statistically different (two-sample test of proportions,  $p > 0.25$ ).

Table 2: Duplicates in the sample.

	Control	Treatment
a) Not changing (%)	15 (3.9%)	15 (4.3%)
b) Changing treatment status (%)	2 (0.5%)	3 (0.9%)

<sup>6</sup>No one completed the experiment more than twice.

We decided to keep only the first participation of these subjects. Google Forms does not provide information about IP, but records the time of the experiment, so it is easy to identify the first response.<sup>7</sup>

## Uninvited subjects

Another potential concern was that subjects could resend the invitations to their contacts. This problem can lead to an imbalance in the number of observations in each group, and can also reduce the comparability between the treatment and the control group. Because participants were also asked to provide the name of the student who invited them, we were able to identify those subjects who were invited by another person whose name was not on our list of students.

We had 99 uninvited subjects: 49 in C and 50 in T. This represented an increase of 14% and 13%, respectively. The proportions in each group are not statically different (two-sample test of proportions,  $p > 0.5$ ). Regarding comparability, Table 3 shows that if we add these observations, the differences between the control and the treatment remain the same. Indeed, the difference for future participation vanishes. This resulted in a total of 836 different subjects and also suggests that our procedure is robust to this potential threat.

Table 3: Balance adding subjects invited by an unknown person.

	Control	Treatment	$T - C$	$p - value$
Female	0.45	0.445	-0.005	0.895
Student	0.839	0.845	0.006	0.826
Solidarity	4.775	5.107	0.332*	0.065
Envy	2.984	3.13	0.146	0.36
Future participation	0.557	0.61	0.053	0.123

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

<sup>7</sup>In some survey platforms, you can allow only one entry per IP address, that is, two subjects cannot fill out a survey using the same PC. However, this solution is not available in Google Forms. Using this tool, you can set a limit of one response if you only allow respondents to log in to Google, but this excludes subjects without a Google account.

## 6 Results

In this section we present the results of our framing experiment. In order to check the external validity of the procedure, we compare our results with those found in the literature on framing effects in cooperation. Specifically, we compare the results with those of Capraro and Rand (2018)

It is well known in experimental economics that small changes in the framing of games (i.e., the way in which the game is described to participants or how options are presented) can have large effects on players' choices. The literature of framing effects on giving in dictator games and PDGs is extensive. When information regarding the recipient or any other social context is provided, there is a significant increase in generosity (Hoffman et al., 1996; Brañas-Garza, 2007; Engel, 2011). For cooperation games, the literature is more concise but also suggests that framing effects are important. Changing the title of the game (e.g., "Community Game" vs. "Wall Street Game") increases the cooperation rate significantly (Lieberman et al., 2004). Engel and Rand (2014) found that the cooperation rate in PDG increases with a cooperative frame compared to a competitive frame, but the behavior in a neutral frame was not different than in a cooperative frame. Goerg et al. (2019) also found that "give versus take" frames influence subjects' behavior and beliefs in PDG but not in dictator games. Lastly, Capraro and Rand (2018) - the experiment that is most similar to ours - found that prosocial people tend to choose the option that is presented as being morally right in the given situation, be it equitable or efficient. Their results support the idea that general morality preferences rather than social preferences explain cooperation for minimizing social inequities.

In our experiment, we tested whether strategies labeled as "Cooperate" and "No cooperate" had an impact on cooperation compared to neutral instructions (A vs. B). We also tested whether this effect might be explained by individual social preferences (measured by self-reported solidarity).

Panel A of Figure 4 shows a pronounced treatment effect on choices. Cooperation becomes more likely when the labels "Cooperate" or "Not coop-

erate” are used compared to neutrally framed strategies (two-sample t test,  $p = 0.024$ ).<sup>8</sup> Panel B shows that the treatment has heterogeneous effects by solidarity groups.<sup>9</sup> For subjects with low solidarity, the effect was not significant (t test,  $p > 0.7$ ). However, for those with medium or high solidarity the use of non-neutral instructions increased cooperation (t test: Medium,  $p = 0.093$ ; High,  $p = 0.017$ ).

Model 1 in Table 4 shows that the use of framing increases cooperative behavior by 0.076 points, which represents a 13% increase in cooperation compared to the control group. Models 2 and 3 show that the estimations are robust to the inclusion of different controls. Model 4 shows that the interaction between *solidarity* and *treatment* is positive and weakly significant ( $p = 0.092$ ) and that the interaction between *envy* and *treatment* is not significant.

In sum, we found that framed instructions increase cooperation and that the treatment effect is higher for subjects with medium and high self-reported solidarity. This evidence is in line with the results of Andreoni (1995). He found that people are significantly more willing to cooperate in games when the problem is posed as a positive externality (i.e., subjects’ choices will benefit the other subjects). Indeed, our results extend those found by Capraro and Rand (2018), since part of the framing effect is explained by an individual taste for solidarity. These results are also in line with those found in Capraro et al. (2020) for the Stag Hunt Game.

## 7 Discussion

In lab and online experiments, internal validity requires that: i) the treatment assignment rule is exogenous; ii) there is no differential attrition; and iii) subjects are unable either to interact with or influence another participant’s decisions (stable unit treatment value assumption or SUTVA). This paper presents

---

<sup>8</sup>Note that we use the 737 subjects plus the 99 special cases described in section 5. The final distribution was 52% in the treatment and 48% in the control group.

<sup>9</sup>We divided solidarity in 3 categories: low for solidarity lower than 5, medium for solidarity equal to 5 or 6 and high for solidarity higher than 6.

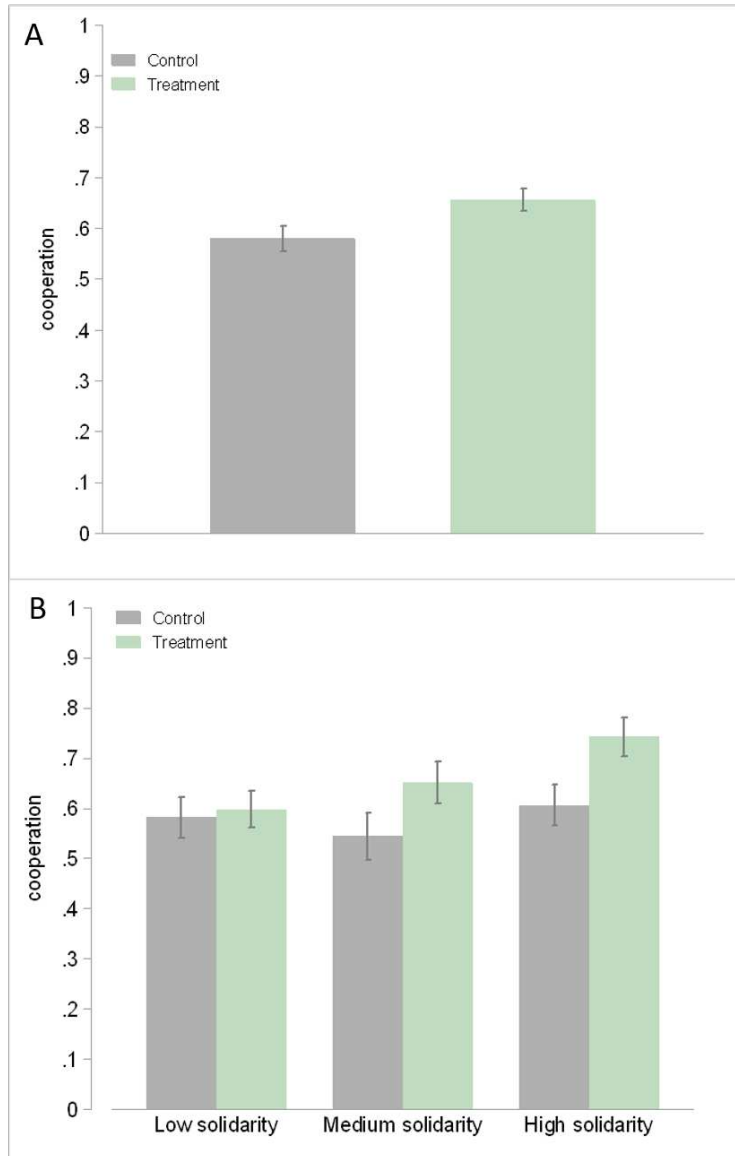


Figure 4: Cooperation in control and treatment group. Panel (A) displays average cooperation by treatment. Panel (B) displays average cooperation by treatment and solidarity groups. Error bars represent SEM.

Table 4: OLS estimation of the treatment effects.

	(1)	(2)	(3)	(4)
<i>Treatment</i>	0.076** (0.034)	0.075** (0.034)	0.077** (0.033)	-0.035 (0.084)
<i>Student</i>		-0.023 (0.045)	-0.021 (0.045)	-0.018 (0.039)
<i>Female</i>		0.078** (0.034)	0.070** (0.034)	0.073** (0.034)
<i>Solidarity</i>			0.012* (0.006)	0.001 (0.010)
<i>Envy</i>			-0.015** (0.007)	-0.017 (0.011)
<i>Treatment * Solidarity</i>				0.022* (0.013)
<i>Treatment * Envy</i>				0.002 (0.015)
<i>Constant</i>	0.580*** (0.025)	0.565*** (0.049)	0.553*** (0.062)	0.611*** (0.073)
Observations	836	836	836	836
R-squared	0.006	0.013	0.022	0.025

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .



an alternative device at almost no cost that can provide a huge number of participants and fulfill all the critical requirements regarding power, balance and attrition. Using WhatsApp, we recruited a large number of subjects in only 12 hours at a negligible cost.<sup>10</sup>

The invitation and randomization procedure described in Section 2 shows that we were able to obtain two comparable groups. Both groups were balanced in two self-reported variables that are related to the alpha and beta parameters of Fehr and Schmidt (1999) and correlated with giving and cooperation (Espín et al., 2018).

Moreover, we found identical attrition levels in the control and the treatment. Differential attrition leads to a selection bias if there is unobserved sorting that could be driving the results. This is an untestable assumption that we need make in all experiments. Since we obtained the same level of attrition in both groups, this problem can be considered less important.

However, we found certain threats to internal validity that are magnified in any online setting. First, we found that some subjects performed the experiment twice. This potential problem could be greater in platforms like Google Forms because it is not possible to restrict more than one access per IP address. However, only 4% of the subjects did the experiment twice, and this percentage was distributed equally between the control and the treatment group. This also suggests that the randomization procedure worked properly and assured between-group comparability.

Second, we found uninvited people who completed the experiment. However, they were distributed evenly between the treatment and the control and did not affect the comparability of the groups.

Lastly, the SUTVA assumption requires that any individual's outcome depends only upon his or her treatment assignment and not upon the treatment assignment or outcome of any other subject (Rubin, 1974). This potential problem is a threat in both online and physical laboratories unless the full experiment

---

<sup>10</sup>Recruiting from labor markets such as MTurk or Prolific Economics also permits reaching a large number of subjects.

is run in one large session. We tried to reduce this problem by running the experiment in a very short time (11 hours) as suggested by Horton et al. (2011). A remarkable advantage of our experimental procedure was that we reached 300 participants in the first 25 minutes, 400 in 40 minutes and 700 in just one hour.

Regarding external validity, in both online and physical experiments some degree of self-selection of participation is inevitable. Our results are in line with the framing effect literature, suggesting that our experiment is informative.

Finally, we should mention that the majority of participants (84%) were students, which indicates that our procedure is more likely to attract students than participants who are not students. However, Exadaktylos et al. (2013) found that self-selected students are an appropriate subject pool for the study of social behavior.

We therefore consider that our recruiting procedure for online experiments is a proper avenue for experimental and social research.

## References

- Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 110(1):1–21.
- Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4):477 – 486.
- Capraro, V. and Rand, D. G. (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*, 13(1):99–111.
- Capraro, V., Rodriguez-Lara, I., and Ruiz-Martos, M. J. (2020). Preferences for efficiency, rather than preferences for morality, drive cooperation in the one-shot stag-hunt game. *Journal of Behavioral and Experimental Economics*, 86:101535.

- Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2):424–55.
- Dillman, D. A. (2011). *Mail and Internet surveys: The tailored design method—2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.
- Engel, C. and Rand, D. G. (2014). What does “clean” really mean? the implicit framing of decontextualized experiments. *Economics Letters*, 122(3):386–389.
- Espín, A. M., Correa, M., and Ruiz-Villaverde, A. (2019). Patience predicts cooperative synergy: The roles of ingroup bias and reciprocity. *Journal of Behavioral and Experimental Economics*, 83:101465.
- Espín, A. M., Moreno-Herrero, D., Sánchez-Campillo, J., and Martín, J. A. R. (2018). Do envy and compassion pave the way to unhappiness? social preferences and life satisfaction in a spanish city. *Journal of Happiness Studies*, 19(2):443–469.
- Exadaktylos, F., Espín, A. M., and Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific Reports*, 3(1):1–6.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fowler Jr, F. J. (2013). *Survey research methods*. Sage Publications.
- Glewwe, P., Kremer, M., Moulin, S., and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya. *Journal of Development Economics*, 74(1):251–268.

- Goerg, S. J., Rand, D., and Walkowitz, G. (2019). Framing effects in the prisoner’s dilemma but not in the dictator game. *Journal of the Economic Science Association*.
- Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, 86(3):653–660.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.
- Jorrat, D., Ortega, D., and Ronconi, L. (2018). No al gatillo fácil: Experimental evidence from a rational use of force police training program in argentina. *Mimeo*.
- Kaplowitz, M. D., Hadlock, T. D., and Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, 68(1):94–101.
- Liberman, V., Samuels, S. M., and Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner’s dilemma game moves. *Personality and Social Psychology Bulletin*, 30(9):1175–1185.
- Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.
- Millar, M. M. and Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2):249–269.
- Rand, D. G. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299:172–179.

- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

# Supplementary Information

## Recruiting experimental subjects using WhatsApp

Diego Jorrat

### A Design

The front page welcomes the experimental subjects and starts with an introduction to the prisoner's dilemma task (see Panel A in Figures). The subjects then had to click on the acceptance of the informed consent, provide their contact information and indicate the name of the person who invited them. On the third page, a decision tree explaining the PD was displayed again and they had to make the choice (see Panel B in Figures).

#### Control group

Figure S1 shows the screenshots for the instructions and the decision tree to answer the PD for the control group. The translated instructions are as follows:

##### *Panel A*

Hello

This is an experiment done by students at Loyola University. You are going to make a decision that can allow you to earn money. At the end of the day we will randomly choose 5 pairs (10 participants) and pay them the money they have won. The money is real and is funded by a research project.

The task is very simple: We are going to match you with another person (neither of you will know who your partner is) and you are going to make a single decision. The task is to choose "A" or "B".

- If you choose "A" and your partner chooses "A", both will win €5.
- If you choose "A" and your partner chooses "B", you will win €0 and your partner will win €10.
- If you choose "B" and your partner chooses "A", you will win €10 and your partner will win €0.

- If you choose "B" and your partner chooses "B", both will win €1.

You will be person 1. Person 2 will be another participant chosen at random. You have until 23:59 today to participate. At that time we will close the page. You will see the answers when we're done (we will prepare a website with the answers). Thank you for participating.

Remember: You will not know who your partner is. At the end of the day there will be a lottery and we will select the 10 winners. We will then contact the winners to pay them the money (by Bizum or bank transfer).

If for any reason you regret your decision and want to delete your data, send an email to ULOYOLADE@gmail.com and we will delete your data. If you have any questions, please write to this email.

*Panel B*

Your decision

You can choose "A" or "B". If you choose road "A" and your partner chooses "A", both will win €5; but if your partner chooses "B", you will win €0 and your partner will win €10. If you choose road "B" and your partner chooses "B", both will win €1; but if your partner chooses "A", you will win €10 and your partner will get €0.

### **Treatment group**

Figure S2 shows the screenshots of the instructions and the decision tree to answer the PD for the control group. The translated instructions are as follows:

*Panel A*

Hello

This is an experiment done by students at Loyola University. You are going to make a decision that can allow you to earn money. At the end of the day we will randomly choose 5 pairs (10 participants) and pay them the money they have won. The money is real and is funded by a research project.

The task is very simple: We are going to match you with another person (neither of you will know who your partner is) and you are going to make a single decision. The task is to choose "Cooperate" or "Not cooperate".

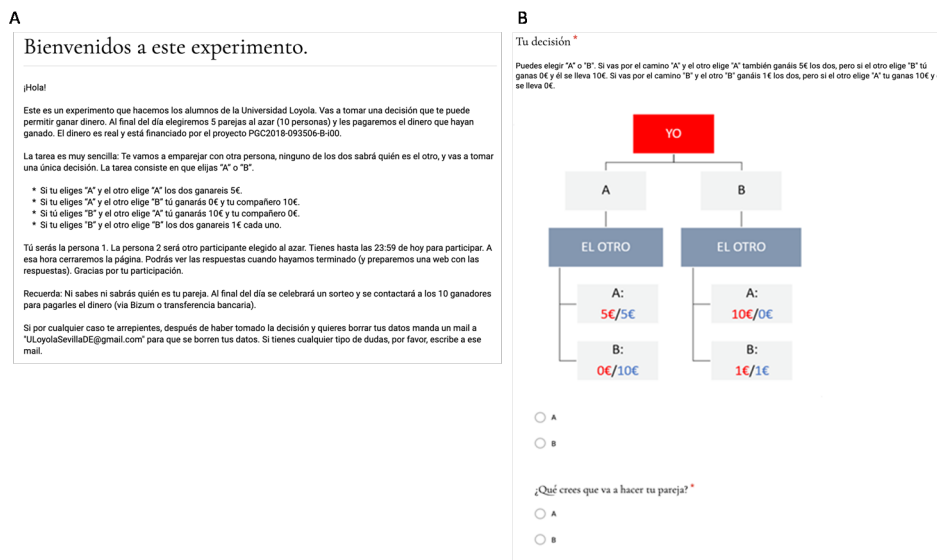


Figure S1: Screenshots of the experimental instructions for the control group. Panel A shows the instructions and Panel B the decision tree.

- If you choose "Cooperate" and your partner chooses "Cooperate", both will win €5.
- If you choose "Cooperate" and your partner chooses "Not cooperate", you will win €0 and your partner will win €10.
- If you choose "Not cooperate" and your partner chooses "Cooperate", you will win €10 and your partner will win €0.
- If you choose "Not cooperate" and your partner chooses "Not cooperate", both will win €1.

You will be person 1. Person 2 will be another participant chosen at random. You have until 23:59 today to participate. At that time we will close the page. You will see the answers when we're done (we will prepare a website with the answers). Thank you for participating.

Remember: You will not know who your partner is. At the end of the day there will be a lottery and we will select the 10 winners. We will then contact them to pay them the money (by Bizum or bank transfer).



If for any reason you regret your decision and want to delete your data, send an email to ULOYOLADE@gmail.com and we will delete your data. If you have any questions, please write to this email.

*Panel B*

Your decision

You can choose "Cooperate" or "Not cooperate". If you choose the road "Cooperate" and your partner chooses "Cooperate", both will win €5; but if your partner chooses "Not cooperate", you will win €0 and your partner will win €10. If you choose the road "Not cooperate" and your partner chooses "Not cooperate", both will win €1; but if your partner chooses "Cooperate", you will win €10 and your partner will get €0.

**A**

Bienvenidos a este experimento!

¡Hola!

Este es un experimento que hacemos los alumnos de la Universidad Loyola. Vas a tomar una decisión que te puede permitir ganar dinero. Al final del día elegiremos 5 parejas al azar (10 personas) y les pagaremos el dinero que hayan ganado. El dinero es real y está financiado por el proyecto PGC2018-093356-B-I00.

La tarea es muy sencilla. Te vamos a emparejar con otra persona, ninguno de los dos sabrá quién es el otro, y vas a tomar una única decisión. La tarea consiste en que elijas "Cooperar" o "No cooperar".

- Si tu eliges "COOPERAR" y el otro elige "COOPERAR" los dos ganaréis 5€.
- Si tu eliges "COOPERAR" y el otro elige "NO COOPERAR" tú ganarás 0€ y tu compañero 10€.
- Si tu eliges "NO COOPERAR" y el otro elige "COOPERAR" tú ganarás 10€ y tu compañero 0€.
- Si tu eliges "NO COOPERAR" y el otro elige "NO COOPERAR" los dos ganaréis 1€ cada uno.

Tú serás la persona 1. La persona 2 será otro participante elegido al azar. Tienes hasta las 23:59 de hoy para participar. A esa hora cerraremos la página. Podrás ver las respuestas cuando hayamos terminado (y preparemos una web con las respuestas). Gracias por tu participación.

Recuerda: Ni sabes ni sabrás quién es tu pareja. Al final del día se celebrará un sorteo y se contactará a los 10 ganadores para pagarles el dinero (via Bizum o transferencia bancaria).

Si por cualquier caso te arrepientes, después de haber tomado la decisión y quieres borrar tus datos manda un mail a "ULoyolaSevillaDE@gmail.com" para que se borren tus datos. Si tienes cualquier tipo de dudas, por favor, escribe a ese mail.

**B**

Tu decisión \*

Puedes elegir "Cooperar" o "No cooperar". Si vas por el camino "Cooperar" y el otro coopera también ganarás 5€ los dos, pero si el otro elige no cooperar tu ganas 0€ y él se lleva 10€. Si vas por el camino "No cooperar" y el otro no coopera ganarás 1€ los dos, pero si el otro elige cooperar tu ganas 10€ y él se lleva 0€.

```

graph TD
    YO[YO] --> C1[Cooperar]
    YO --> N1[No cooperar]
    C1 --> C2[Cooperar: 5€/5€]
    C1 --> N2[No cooperar: 0€/10€]
    N1 --> C3[Cooperar: 10€/0€]
    N1 --> N3[No cooperar: 1€/1€]
            
```

Cooperar

No cooperar

¿Qué crees que va a hacer tu pareja? \*

Cooperar

No cooperar

Figure S2: Screenshots of the experimental instructions for the treatment group. Panel A shows the instructions and Panel B the decision tree.

## B Self-reported solidarity and envy

Following Espín et al. (2018), we obtained a self-reported measure of solidarity and envy. Figure S3 shows the screenshots for both questions.

### *Solidarity (Top part of Figure S3)*

Please use the scale to indicate to what extent you agree or disagree with the following statement: “I do not care about how much money I have; what concerns me is that there are people who have less money than I have”. Left: strongly disagree / Right: strongly agree.

### *Envy (Bottom part of Figure S3)*

Please use the scale to indicate to what extent you agree or disagree with the following statement: “I do not care about how much money I have; what concerns me is that there are people who have more money than I have”. Left: strongly disagree / Right: strongly agree

⋮

Valora en la escala cómo de acuerdo estas con la siguiente afirmación acerca de ti: "No me preocupa cuánto dinero tengo, lo que me preocupa es que otros tienen menos que yo" \*

1 2 3 4 5 6 7 8 9 10

Totalmente en desacuerdo           Totalmente de acuerdo

Valora en la escala cómo de acuerdo estás con la siguiente afirmación respecto a ti: "No me preocupa cuánto dinero tengo, lo que me preocupa es que otros tienen más que yo" \*

1 2 3 4 5 6 7 8 9 10

Totalmente en desacuerdo           Totalmente de acuerdo

Figure S3: Screenshots of experimental instructions II. Top: instructions for Solidarity. Bottom: instructions for Envy.