Bridging the gap between the lab and the field: Dictator games and donations

Xinghua Wang

Pompeu Fabra University;
Barcelona Graduate School of Economics.

Daniel Navarro-Martinez

Pompeu Fabra University;
Barcelona Graduate School of Economics;
Barcelona School of Management.

Abstract

There is a growing concern about the extent to which laboratory games generalize to social behaviors outside the lab. In this paper, we show that it is possible to make lab games much more predictive of field behavior by bringing the right contextual elements from the field into the lab. For this, we focus on dictator games and charitable giving in the field. We present a series of three experiments in which we face the same participants with different versions of the dictator game and with two different field situations outside the lab: one in which they can donate money to a charity and one in which they can show interest in volunteering. The games are designed to include step by step contextual elements that make them similar to our first field situation. We find a dramatic increase in the lab-field correlation with the first situation as contextual elements are incorporated. However, making the games more similar to the first situation does not increase correlation with the second one, showing that pro-social behavior is highly context-specific. Our results demonstrate that context-free games have small predictive power in the field, but this gap can be substantially bridged by incorporating appropriate contextual elements from the field. This highlights the importance of establishing close links between lab and field research.

1. Introduction

Research on what is known as "social preferences" has attracted great interest in the last decades and has been one of the main building blocks of behavioral and experimental economics. Social preferences mainly fall into two categories: distributive preferences and reciprocal preferences. For agents with distributive preferences, their utilities are not narrowly defined over their own payoffs but over distributions of payoff in certain groups. That is, for the well-being of these agents, not only material wealth matters, but also some spiritual pursuit, e.g., equity, efficiency, and altruism. Reciprocal preferences then take into account decision makers' desire to reward or punish others beyond mere material consequences, e.g., gift-exchange and punishing those who are unfair to the decision maker, even at a cost.

A central feature of this line of research is that it has been strongly based on studying behavior in "social preference games", experimental games designed to capture different aspects of social behavior, such as altruism, inequity aversion, reciprocity, trust, cooperation, etc. Examples of these games include dictator game, ultimatum game, trust game, and public goods game, among others (see Camerer, 2003, for a review). These social preference games typically share some common characteristics: they are closely based on game-theoretic structures and have clear game-theoretic equilibria; they are deliberately as context-free as possible; and their outcomes consist in monetary payoffs for the players involved. This approach to social behaviors has been hugely influential and has become a benchmark for the study of human interaction in economics and beyond.

Context-free games have some clear virtues. They provide very stylized and controlled environments that allow researchers to achieve high levels of internal validity in the laboratory. That is, the variables being modified in experiments and the elements being affected by them can be very clear and can be very tightly controlled. Moreover, this facilitates that different research groups use exactly the same tools in the same way and that evidence can quickly accumulate.

However, the artificial and context-free nature of these games raises one very fundamental question: are social preference games externally valid? In other words, do social preference games tap into the principles that determine social behavior outside the laboratory?

To the extent that the aim of economics, and of the social and behavioral sciences more generally, is to understand real-world social behavior outside the lab and not how people play artificial games, this is a crucial question to respond. There is currently increasing concern that social preference games cannot deliver on external validity. Levitt and List (2007) were the first to systematically discuss this concern (but see Loewenstein, 1999, for a shorter discussion). The authors presented a theoretical framework that can be used to organize different factors that are likely to limit the external validity of social preference games, including differences between lab games and field settings in terms of scrutiny, anonymity, context, stakes, participants, and restrictions on choice sets and time horizons. After Levitt and List's paper, however, there has only been a relatively small number of papers that have investigated the issue of external validity (for reviews, see Camerer, 2011; Galizzi and Navarro-Martinez, 2019). In a meta-analysis of this literature, Galizzi and Navarro-Martinez (2019) concluded that only 39.7% of the lab-field correlations reported in the papers they analyzed were statistically significant, and the average lab-field correlation obtained was 0.14. The authors also conducted a systematic experiment comparing various lab games against several field behaviors and found no significant correlation between them. It is probably too soon to reach final conclusions about the external validity of all these games, but it seems safe to say that social preference games have issues of external validity.

We actually do not find these external validity problems too surprising, given that behavioral research has widely documented that context plays a crucial role in economic behavior (see, e.g., Slovic, 1995; Ariely et al, 2003, 2006; Lichtenstein and Slovic, 2006; List, 2007; Stewart et al., 2015). This also relates to the long-standing person-situation debate in psychology, in which social and personality psychologists have shown that the cross-situational consistency of behavior is typically very low and behavior is highly dependent on particular situational cues (see Mischel, 1968; Ross and Nisbett, 1991; Fleeson, 2001, 2004). Overall, if behavior is so determined by contextual elements, it seems logical that context-free lab games cannot provide a good account of behavior when it is put in context outside the lab.

Given the context-free nature of social preference games and the context-dependent nature of behavior, one interesting possibility arises: it might be possible to bridge the gap

between the lab and the field and improve the external validity of the games by introducing contextual elements in the lab. This brings us to the main goal of our paper.

In this paper, we show that it is possible to make lab games much more predictive of field behavior by taking the right contextual elements from the field and incorporating them to the games. For this, we focus on the dictator game (Kahneman et al., 1986) and on charitable giving in the field. The dictator game is one of the most influential games in social preference research and arguably the simplest one. It is also a game that has been conceptually related to real-world social behaviors, such as charitable giving (e.g. Eckel and Grossman, 1996; Branas-Garza, 2006; Benz and Meier, 2008; Carpenter et al., 2008; Carpenter and Myers, 2010; Konow, 2010; Kolstad and Lindkvist, 2012), and a game in which behavior has been shown to be sensitive to contextual elements. For instance, giving in the game has been shown to be affected by the degree of anonymity (Dana et al., 2006; Franzen and Pointner, 2012), adding a taking option (List, 2007; Bardsley, 2008), earning the monetary endowment (List, 2007), the presence of verbal feedback and face to face interaction (Ellingsen and Johannesson, 2008; Andreoni and Rao, 2011; DellaVigna et al. 2012; Andreoni et al., 2017), or the type of recipient of the shared amount (Eckel and Grossman, 1996; Konow, 2010). All this makes the dictator game a suitable candidate for our investigation.

In the field, we decided to focus on charitable giving, which is a domain of notable economic and social relevance. The estimated amount of donation to charities in the USA in the year 2018 was $428 billion (Giving USA 2019 report); the annual estimated amount in Europe is €87.5 billion (Hoolwerf and Schuyt, 2017). Charitable giving is also a widely researched topic in economics (see, e.g., Andreoni, 1989, 1990; Glazer and Konrad, 1996; Okten and Weisbrod, 2000; Auten et al., 2002; Karlan and List, 2007; List, 2011; DellaVigna et al., 2012; Andreoni and Payne, 2013; Andreoni et al., 2017), and like we said, a domain that has been directly linked to the dictator game.

So, we investigated if it is possible to make the dictator game more predictive of charitable giving in the field. To this end, we run a series of three interconnected experiments in which we presented the same participants with different versions of the dictator game and with two naturalistic field situations that we created, in which they could behave pro-socially.

In the first field situation, participants were approached by a solicitor and had the opportunity to donate money to charity; in the second one, they could show interest in volunteering by checking information about volunteering opportunities at various charities. Our different versions of the dictator game were designed to incorporate step by step additional contextual elements that made the games more similar to our first field situation. In particular, we incorporated the following three elements: a recipient in real need (as opposed to another student), a monetary endowment that was earned by the participant (as opposed to simply assigned by the experimenter), and face to face interaction (as opposed to anonymous giving).

Our approach bears some similarities to that of List (2006), Stoop et al. (2012) and Stoop (2014), where the authors designed lab games to match particular field environments. None of these papers, however, correlated lab and field behaviors for the same sample of participants.

By presenting the first within-subject study bridging the lab-field gap for certain types of social behaviors, we obtained some informative novel findings. We found a dramatic increase in the correlation between the lab games and our first field situation as more contextual elements were incorporated. These elements, however, did not increase the correlation with our second field situation (which they were not intended to address), confirming that pro-social behavior is highly context-specific. Our results show that context-free lab games have very limited predictive power in the field, but this power can be very substantially increased by incorporating appropriate contextual elements from the field to the games. Our interpretation of these findings is that social preference research should move in the direction of including more context in laboratory experiments and be more explicit about the types of field behaviors it aims to address. Combining lab and field experiments can also be an important way to reduce external validity issues. Remaining in the realm of context-free games could put social preference research at risk of being disconnected from real-world behaviors.

The rest of the paper is organized as follows: Section 2 describes our research methods, Section 3 presents the results, and Section 4 discusses our findings and concludes.

2. Methods

We conducted a series of three interconnected experiments. In subsection A, we describe the general procedure that these experiments follow. In subsection B, we present the two field situations that will serve as our benchmark of field behavior that the lab games will aim to predict. In subsection C, the different lab games that we created to approach step by step one of the field situations were explained. In subsection D, E, and F, we demonstrate that how all these elements were organized in the three different experiments.

A. General procedure

Participants in our study were required to come to our lab twice. At the beginning of our first session, the participants were assigned anonymous codes. They were told to keep the codes until we finish all the sessions. While those in the control groups did some tasks unrelated to social preferences on the first day, those in the treatment groups played some (one or two) lab games. In the first session, all the participants also filled out the Interpersonal Reactivity Index (Davis, 1980), which measures the four dimensions of empathy, i.e. Perspective Taking, Empathic Concern, Fantasy, and Personal Distress. A big-five personality questionnaire (John, 1991) and some demographic questions were also included in the first session.

In the second session, the participants came to a classroom, which was temporally used as our lab. Tasks included in this session were not related to social preferences. Specifically, participants did a short cognitive ability test (Condon and Revelle, 2014), two short cognitive reflection tests (Primi et al., 2016 and Thomson and Oppenheimer, 2016) and answered some hypothetical questions regarding choices of lotteries (Holt and Laury, 2002). All the tasks were printed on papers. When the participants answered the tasks, they were required to put their lab codes assigned earlier on the first page of the task paper.

After finishing the tasks, each participant brought the task paper to the experimenter. Then each participant was paid €15 that promised for attending all the two sessions. The €15 was always paid with the exactly same bill and coin denominations: two €5 bills, one €2 coin, two €1 coins, one 50 cents coin, two 20 cents coins, and one 10 cents coin. These bills and coins in various denominations can make up any amount from €0 to €15 with an increment of €0.1. This payment method thus made sure that each participant can make any amount of donations (below €15) in cash when the participant met our solicitor outside of the lab.

After getting paid, each participant was required to sign a receipt and a consent form that entitles us to use his or her data for research. On these forms, the participants needed to put their names and email addresses. Then we could match the lab codes on the task papers with the participants' email addresses.

When the participants walked out of the lab, they encountered our research assistant soliciting donations for a charity. About two weeks after all the participants in the same experiment have finished the second session, we sent the participants emails, embedded in which a link to our website, on which they can check volunteering opportunities at three different charities.

The whole study was conducted in Spanish. We used the Spanish version of Interpersonal Reactivity Index translated by Pérez-Albénez et al. (2003). The Spanish version of the big five questionnaire is the one provided by Benet-Martínez and John (1998).

The software we used to implement our lab games was oTree (Chen et al., 2016).

B. The field situations

B.1 Children's charity

Inside the lab, the experimenter let the participants out approximately every three minutes, to give the previous participant enough time to complete the field situation. The experimenter also wrote down the lab codes of the participants in the order they left.

Outside of the lab is a square, on which students at our university gather and relax after class, and advertise all kinds of activities (including collecting donations for different causes). We put a table and a yellow-wooden chair on the square (the table and chair were provided by the university just for events on the square). Our assistant stood beside the table soliciting money for a leading charity devoted to helping children in developing countries. For convenience, let us call this charity "Charity F". She was wearing an official university T-shirt. On the table was put a professional (sealed) charity bucket with a large sticker with the logo of both the charity and the university. We also prepared the assistant a laminated color-printed leaflet, which introduced the charity and programs that the donations would be used for.

There is only one way going out the lab. On his or her way leaving, a participant had to

pass our assistant. When a participant was almost at the location of our assistant, our assistant talked to the participant. She was instructed to ask each participant the same words, “I am collecting donation for Charity F, are you interested in learning more?”

The assistant recorded the behavior of each participant in the field situation, including the amount donated and a brief summary of what the participant said to her. Later we match the lab codes on the form recorded inside the lab with the recorded behaviors outside of the lab.

Since the field situation was very similar to the usual events on the square, very few participants would connect the situation with our study.

Before conducting the study, we asked approval from both the charity and the university. All the donations were sent to the charity.

B.2 Volunteering opportunities

In each experiment, about two weeks after all the participants have finished the second session, the participants received an email (in Spanish), embedded in which a link to our website, on which they could check volunteering opportunities at three different charities.

The subject of the email was “UPF Volunteering”. In the email, we wrote: “We are constructing a platform that provides information to UPF students about volunteering opportunities at charities. Every time we select three volunteering opportunities that suit UPF students best. If you are interested, please click the following link to learn more.”

If a participant clicked the link, he or she would be directed to our website. On the homepage, he or she would see the name of the volunteering programs. A participant could choose the volunteering program he or she likes and click its name. If he or she clicked, the participant would be shown a brief introduction of the volunteering program. At the end of the page, the participant would be asked a question, “Are you interested in volunteering for this program?”. The participant could choose either “yes” or “no”. On the next page, the participants were told that they can click a link to go to the charity’s website if they want to learn more about that volunteering program or to sign up for it.

After having checked the information about one volunteering program, a participant would be automatically directed back to the homepage of our website. So he or she could further check information about more programs.

The three charities we worked with in this field situation were Charity F and another two famous charities in Barcelona. (See footnote 1 for the procedure that we followed to choose the charities we worked with.) One programme of these three is a "digital volunteering" programme. This programme allows its attendees to volunteer at home and affords the attendees many volunteering activities, like translating documents, helping with creating and managing websites, and working with the old or AIDS patients from long distance. The other two programs need their volunteers to help to organize big events or to do some paperwork in the office. To volunteer for these two programs, the volunteers need to sign up first and then will be informed every time when some volunteers are needed.

We used individual links to trace the responses of each participant. We coded the responses as numeric values ranging from 0 to 4. If a participant didn't click the link, a value of "0" was recorded; a value of "1" was recorded if a participant clicked the link, but didn't check the information about any volunteering activity; and a value of "2" was recorded if a participant checked one volunteering activity but showed no interest; and a value of "3" was recorded if a participant checked one volunteering activity and also showed interest; and a value of "4" was recorded if a participant was interested in more than one volunteering activities or if he or she clicked one of the links leading to the website of the charities.

Just in case that some participants didn't notice the email, we sent another identical email about one month after the first email. The responses were coded in the same way. The largest score a participant got in these two rounds was taken as his or her final score in this field situation.

C. The Lab games: Bridging the gap between the lab and the field

We noticed that there are three main aspects in which the standard dictator game differs from our first field situation (donating to a children's charity). These differences are: house money vs. earned money, student as recipient vs. recipient in serious need, and absence vs. presence of face to face interaction. We suspect that these differences may result in the low lab-field correlations in the literature (e.g., Galizzi and Navarro-Martinez, 2019). We thus designed four lab games to eliminate these differences step by step in our lab games, and to see whether we can bridge the gap between the lab and the field.

First, we used the standard dictator game as our benchmark, in which participants made decisions with house money, having their peers as recipients and face to face interaction being absent. In our second lab game, the recipient was changed to a charity engaging in medical-humanitarian actions. In the third game, participants first earned in a real-effort task the same amount of money as in the second game, then decided how much to donate to the same charity. In our final game, we further added the presence of face to face interaction to the previous game. More specifically, the participants made donations not by entering the amount to donate in the computer, but by giving money to our assistant at the door of our lab when leaving. More details of these games will be explained in the following subsections.

The pyramidal structure of our experimental design allows us to investigate the role of the aforementioned three factors in driving the changes in the lab-field correlations.

To minimize across-contamination among the four lab games, and also to make the implementation of our design easier, we further separated the games into three experiments, the details of which will be explained in the following subsections.

D. Experiment I

Each participant in the treatment group of Experiment I played two lab games, the standard dictator game and making donations to a charity with house money. In these two games, the participants made decisions about splitting €5 between themselves and either another random participant or a charity. The dictator game was played two rounds. Across which, the roles of players were switched and players re-matched. In the donating game, the participants were first shown a brief introduction of the charity before they made their decisions. The two lab games were presented in a random order. For each participant, we randomly chose one task from the two rounds of dictator game and the donating task to pay him or her at the end of the first session.

The charity involved engages in medical-humanitarian actions, mainly in developing countries. We worked with this charity in all the lab games that involved making donations. All the donations to this charity were sent to it. To distinguish this charity from the one we worked with in the field, let us call this charity "Charity L". (Remember that the other charity

was referred to as "Charity F".)[1]

Experiment I followed the general procedure described in Subsection A. More specifically, participants in the treatment group played the lab games and answered the questionnaires on the first day. Four days later, they returned to the lab, doing the tasks in the second session and getting paid. Outside of the lab, they were solicited for donations. Later they received emails about volunteering opportunities.

To check whether there was cross-contamination between donating in the lab and making donations in the field, we included a control group in Experiment I. Participants in the control group didn't do any task related to social preference on the first day. They only answered some hypothetical questions regarding time preferences. They were paid a fixed amount of €3 at the end of the first session. The field parts were the same for those in the control group as for those in the treatment.

In total, 161 students participated in Experiment I on the first day: 112 in the treatment and 49 in the control. We lost some participants on the second day. We ended up with 102 participants in the treatment and 46 in the control.

E. Experiment II

In the first session of this experiment, participants first answered the Interpersonal Reactivity Index (IRI), the big-five questionnaire, and some demographic questions, then did a real effort task. In this task, participants were offered a chance to earn €5 by entering sequences of letters and symbols in the computer. Then they were shown the introduction of the charity and asked to make decisions on computers about how much to donate out of the €5 they just earned.

Ten days after the first session, our participants returned to the lab, doing the tasks in that session and getting paid. Outside of the lab, they were solicited for donations. Later they

[1] Before we conducted our study, we did a pilot survey with 106 students at our university. In this survey, we asked the students some questions about the popularity of 12 charities in Barcelona (including a hypothetical allocation of €100 among these charities). Over 96% of students knew well about charity L, charity F, and another two charities we worked with in the second field situation. These four charities were also allocated similar amounts of money.

received emails about volunteering opportunities.

We didn't include a control group in this experiment. The reason is the following. When we run the experiments, we actually did Experiment III before Experiment II. But to facilitate the discussion, we present here Experiment II before Experiment III. Experiment II and Experiment III followed the same procedure except for one part, that is how we solicited donations in the lab. Thus, if we had included a control group in Experiment II, participants in the control group would have been faced with the same tasks as those in the control group in Experiment III. Therefore, there was no need to include another control group in Experiment II; we could just use the control group in Experiment III to detect whether there was cross-contamination between donating in the lab and making donations in the field.

A total of 110 students participated in the first session of Experiment II. We lost 9 participants on the second day. Thus, 101 people participated in both sessions.

F. Experiment III

Participants in Experiment III went to either the treatment group or the control group. For those in the treatment, on the first day, they answered the Interpersonal Reactivity Index (IRI), the big-five questionnaire, and some demographic questions, then did a real effort task. The real effort task was the same one as in Experiment II. The only difference was that the participants earned €4.9, instead of €5, from accomplishing the real effort task. We made this small change just to make it more natural to pay participants with coins.

Then we showed the participants the introduction of charity L, and told them that they can give their donations to our assistant later on their way out if they want to help. We then started to pay each participant that had finished the tasks. The €4.9 was always paid with coins of the same denominations: namely, one €2 coin, two €1 coins, one 50 cents coin, one 20 cents coin, and two 10 cents coin. This payment method allowed the participants to donate any amount from €0 to €4.9 with an increment of €0.1.

We had an assistant stood at the door of our lab (the outer side). She was instructed to ask each participant that came out the lab the same question, "I am collecting donations for charity L, do you want to contribute?" She recorded the amounts of donations and words the participants told her, and were instructed not to ask the names or lab codes of the participants.

The money donated were put in a normal opaque bucket with a hand-written sticker, “donation box”. We also let the participants come out one by one. So each participant’s decision couldn’t be observed by the other participants.

Inside the lab, the experimenter organized the receipts signed in that session in the order the participants left. Later we matched identities on the receipts with the responses recorded by the assistant.

To check whether there was cross-contamination between donating in the lab and making donations in the field, we also included a control group in this experiment. For the participants in this group, the tasks in the first session were the same as the one for those in the treatment, excepting for not presented with the chance to donate. Participants in the control were paid a fixed amount of €4 at the end of the first session.

Ten days after the first session, participants in both groups returned to the lab, doing the tasks in the second session and getting paid. Outside of the lab, they were solicited for donations. Later they received emails about volunteering opportunities.

A total of 169 students participated in the first session of Experiment II: 116 in the treatment and 53 in the control. We lost some participants on the second day. We ended up with 98 participants in the treatment and 45 in the control.

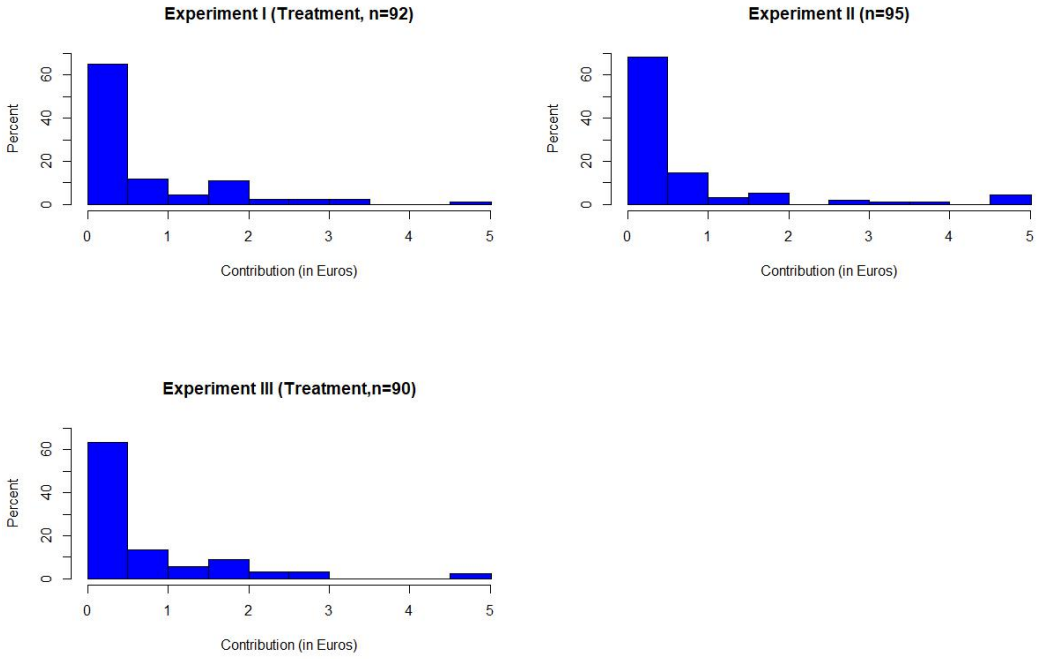
3. Results

We present the results in five separate subsections. In subsections A and B, we describe the patterns of decisions observed in the two field situations. Then we show the distributions of decisions in our four lab games in subsection C. Then, in subsection D, we explore the predictive power of lab decisions on decisions in the two field situations. In subsection E, we check the predictive power of decisions in one field situation on decisions in the other. In subsection F, we correlate pro-social personality measures with decisions both in the lab and in the field.

A. Donations in the field

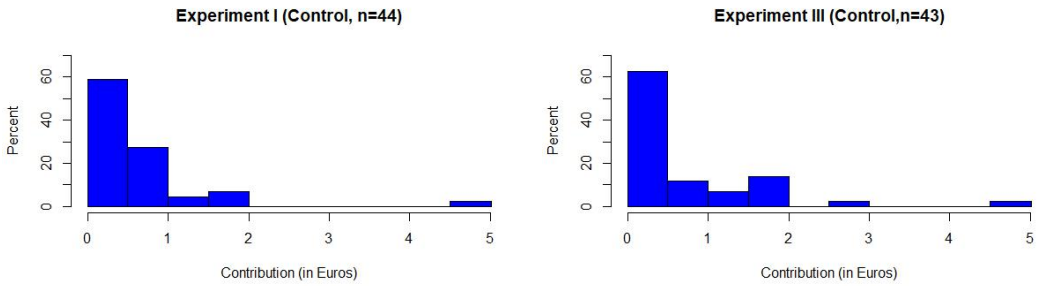
Figure 1a consists of three panels, which show the distributions of donations from participants in the treatment groups in our first field situation.[2]

Figure 1a: Distributions of Donations from the Treatment Groups

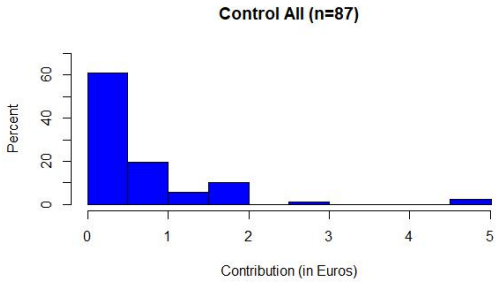


The three panels of Figure 1b show the distributions of donations from participants in the two control groups separately and as a whole.

Figure 1b: Distributions of Donations from the Control Groups



[2] Some observations are missing in the first field situation. The main reasons are that some incidental factors made it impossible for our assistants to approach some participants or to have enough time to explain to them in detail the activity.



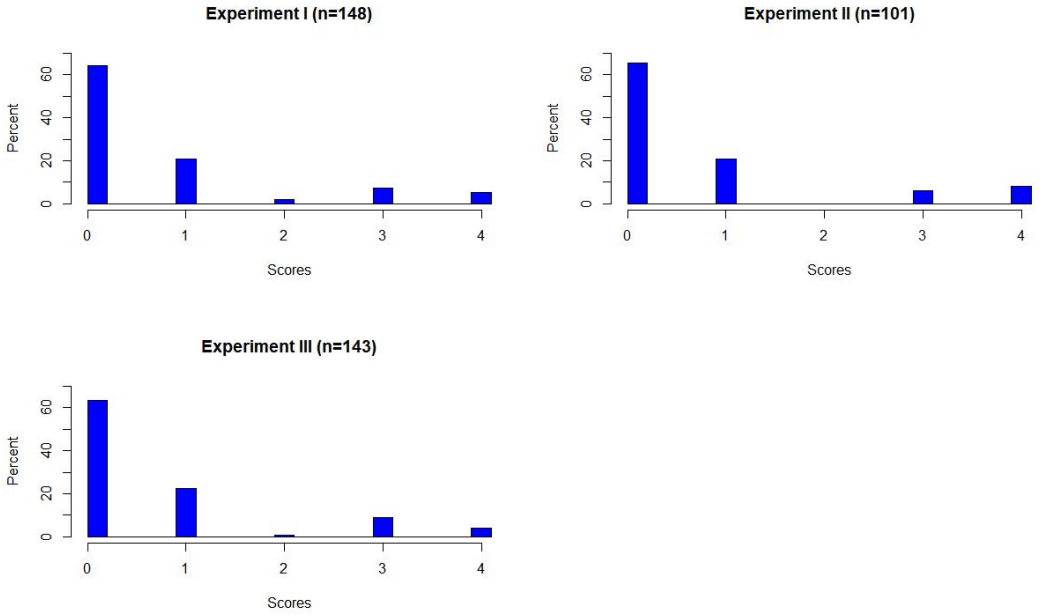
From Figures 1a and 1b, we see that the distributions of donations in our field situation are quite similar across experiments and groups, especially that there is no distinguishable difference between the treatment groups and the control groups. Indeed, this is also what the statistical analysis tells us. In Experiment I, the Wilcoxon rank sum test of the difference between the two distributions of donations from the treatment group and from the control is not statistically significant (p-value is 0.69). Similarly, the difference between the distributions from the two groups in Experiment III is also not statistically significant (p-value is 0.86 for the Wilcoxon rank sum test). And the difference between the distribution of donations in Experiment II and the distribution of donations from the control group in Experiment III is also not statistically significant (p-value is 0.53 for the Wilcoxon rank sum test).

B. Interests in the volunteering opportunities

Figure 2 consists of three panels, which show the distributions of scores that participants in the three experiments got in our second field situation. For how we coded the responses of our participants as numeric values, please refer to the subsection B of Section 2.

From the three pictures, we can see that the distributions of interests in the volunteering opportunities are quite similar across experiments. About 65% of participants were not interested at all; about 20% of participants clicked the link embedded in the email but left very quickly before checking any volunteering opportunity on our website; about 9% of participants checked the information about one volunteering opportunity and showed some interest; and about 6% of participants were very interested: they checked two or all the volunteering opportunities and some of them visited websites of the charities to learn more or sign up.

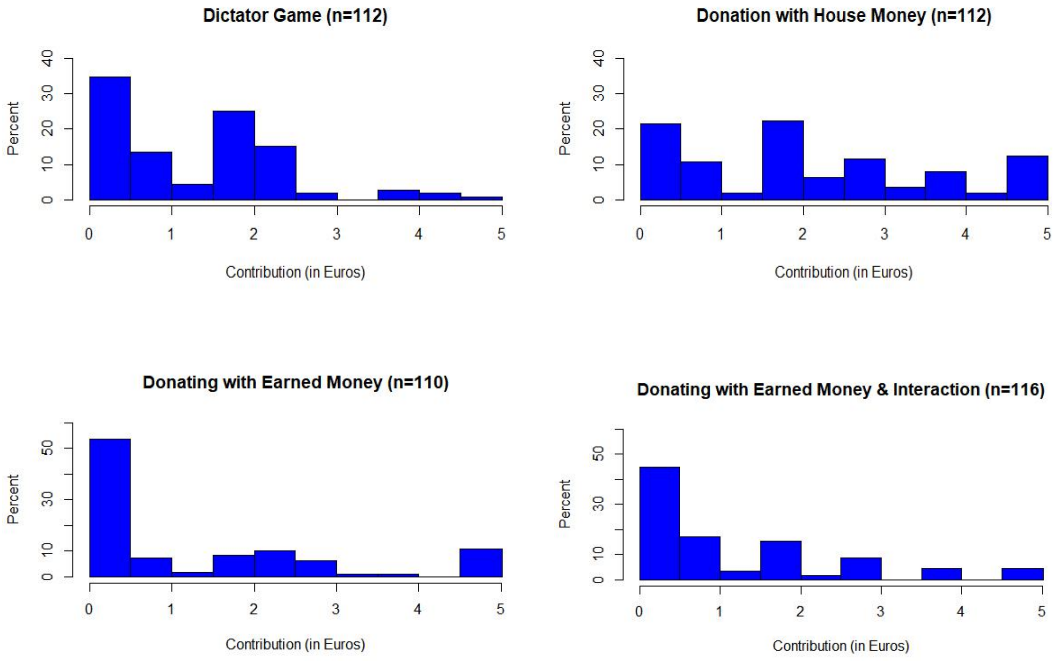
Figure 2: Distributions of Interests in Volunteering



C. Decisions in the lab games

The four panels of Figure 3 show the distributions of decisions in our four lab games.

Figure 3: Distributions of Decisions in the Four Lab Games



The distributions are broadly in line with those reported in the literature. In the standard dictator game, the participants that acted as player 1 on average shared € 1.38 (out of € 5) with player 2; the participants on average donated € 2.24 (out of € 5) when they made donations with house money. So the participants were more generous when they shared money with a charity than they were when sharing money with another participant of the study. This pattern has been well confirmed in the literature (e.g., Eckel and Grossman, 1995).

Remember that we presented the first two lab games together, in a random order, in Experiment I. So we checked whether there was some order effect. The distributions of shared amount of money in the dictator game was not affected by whether the game appeared as the first game or the second game: in the Wilcoxon rank sum test, the p-value is 0.32. This result also holds for the distributions of donations with house money: the p-value is 0.14.

In the latter two games, the participants were less generous when they made donations with earned money rather than with house money. The means of donations in the two lab games were € 1.34 (out of € 5) and € 1.21 (out of € 4.9) respectively. The difference between the distributions of decisions in the latter two games were not statistically significant (the p-value of the Wilcoxon rank sum test is 0.97).

D. Do the lab decisions predict the field behaviors?

We now turn the two key research questions of this study: Could decisions in the standard dictator game predict field behaviors it may want to address? If not, could we bridge the lab-field gap by introducing some important contextual elements into the lab? In the following subsections, we correlate the lab decisions with the two field behaviors in our study.

D.1. Donations in the field

Table 1 reports the Pearson correlations between the game decisions with the donations in the field.

Table 1. Correlations Between the Game Decisions and Donations in the Field (Pearson)

Dictator Game	0.20*

Donating with house money	0.17*
Donating with earned money	0.58***
Donating with earned money & interaction	0.65***

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

The confidence interval at 95% level for the first correlation coefficient is [0,0.39]; it is [-0.03,0.36] for the second correlation coefficient; it is [0.43,0.70] for the third correlation coefficient; and it is [0.51,0.75] for the last correlation coefficient.

Now let us turn to the regression analysis. In the OLS regression models, we regress donations in the field on the lab decisions. We report the results in Table 2. The results in Table 2 preserve more or less the results in Table 1.

Table 2. Regression Analysis (OLS): Donations in the Field

	Coefficient	Variance Explained	No. of observations
Dictator Game	0.18* (0.09)	0.04	92
Donating with house money	0.11* (0.07)	0.03	92
Donating with earned money	0.44*** (0.06)	0.34	95
Donating with earned money & interaction	0.55*** (0.07)	0.43	90

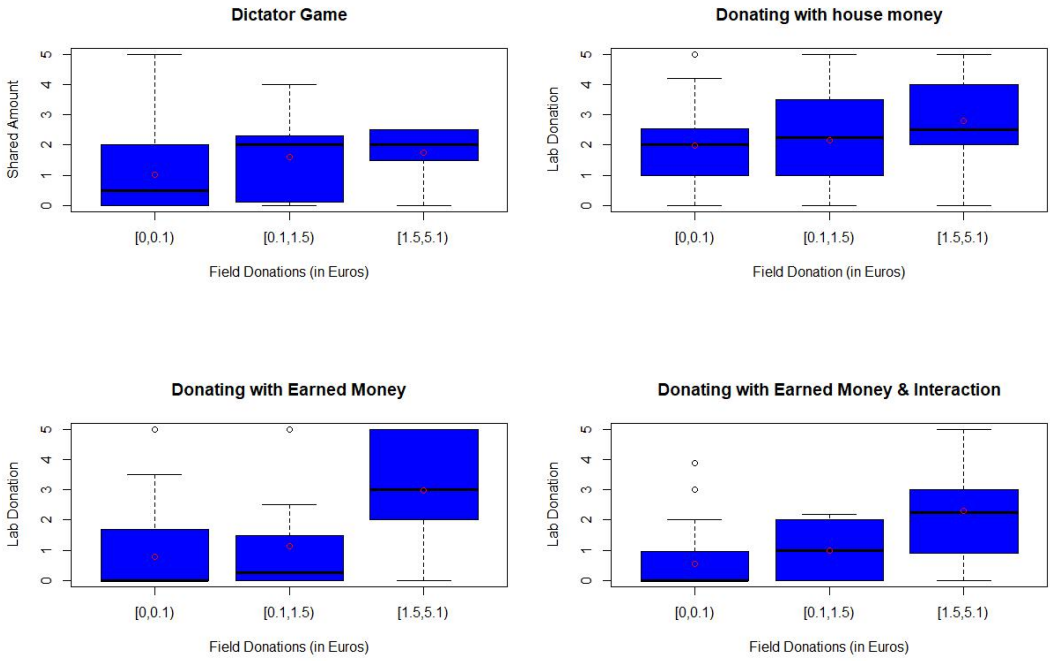
Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

We can infer two main results from the above correlations and regression results. First, decisions in the dictator game has some moderate predictive power on donations in the field. Second, as we eliminate more lab-field differences, the predictive power of the lab decisions increases substantially.[3]

[3] As robustness tests, we also checked Kendall rank correlations between the field donations and the lab decisions, and regression results in Tobit models. Please refer to Table A1 and Table A2 for the results.

But there is an exception with the second game, donating with house money. After we have changed the receipt of the decision makers' contribution from participants in our study to a charity, the predictive power of the lab decisions didn't increase but dropped a bit (even though not statistically significantly regarding Pearson correlations and OLS estimates). How could this happen? In Figure 4, we present four box plots. The patterns in each plot and across the plots may help us answer the question.

Figure 4: Box Plots of Lab-Field Correlations for the Four Lab Games



The horizontal axis in each plot denotes the amount donated in the field; the vertical axis denotes contributions in each lab game. For each lab game, we categorized the participants into three groups by the amounts they donated in the field. The first group consists of participants who didn't donate in the field; the second group consists of those whose donations was between € 0.1 and € 1.5; the third group is composed of those that donated more than € 1.5. The solid black line refers to the median of lab decisions in each group, while the red dot refers to the mean.

We can see that in the second plot that, when participants made donations with house money, those that wouldn't donate in the field were quite generous: on average they donated

€ 2 (out of € 5), while the mean for the whole sample was € 2.19. In the other lab games, those didn't donate anything in the field were not that generous in the lab: their average contribution didn't exceed € 1. The difference in lab contributions across groups in the other games is also more significant.

The above findings are consistent with the literature. Benz and Meier (2008) recorded their participants' donations to two charities in the lab with house money, and they also recorded the participants' donations to the two charities in the four semesters before their experiment and the four semesters after the experiment. They found that the participants that never donated before the experiment on average donated 60 percent of their endowment in the lab; and that those that didn't donate anything after the experiment on average donated 50 percent of their endowment in the lab.

Our understanding of the smaller difference in lab contribution across groups is that not contributing hurts the participants' self-image more when they make donations with windfall money than it does when they share windfall money with other participants or when they make donations with earned money. A large literature in economics (e.g., Konow, 2001; Levitt and List, 2007) argued that deviation from the perceived fair allocations gives the decision maker disutility and that the perceived fair allocations are determined by some principles. In our case, that how much the receiver needs money to maintain his or her standard of life corresponds to the "need" principle in Konow (2001); the "earned money" element then corresponds to the "accountability" principle in Konow (2001), i.e., how much the decision maker and the receiver contributed to the group wealth each.

D.2. Interests in the volunteering opportunities

Table 3 reports the Pearson correlations between the game decisions with the scores the participants got in the second field situation.

Table 3. Correlations Between the Games Decisions and the Scores (Pearson)

Dictator Game	0.15
Donating with house money	0.13
Donating with earned money	-0.01
Donating with earned money & interaction	0.03

Significance level: "*", $0.05<p<0.10$; "**", $0.01<p<0.05$; "***", $p<0.01$

The confidence interval at 95% level for the first correlation coefficient is [-0.05,0.33]; it is [-0.06,0.32] for the second correlation coefficient; it is [-0.21,0.19] for the third correlation coefficient; and it is [-0.17,0.23] for the last correlation coefficient.

Now let us turn to the regression analysis. In the regression models, we regress the scores the participants got on the lab decisions. We report the results from the OLS regressions in Table 4.

Table 4. Regression Analysis (OLS): Interests in Volunteering

	Coefficient	Variance Explained	No. of observations
Dictator Game	0.14 (0.09)	0.02	102
Donating with house money	0.09 (0.07)	0.02	102
Donating with earned money	-0.01 (0.07)	0.00	101
Donating with earned money & interaction	0.03 (0.08)	0.00	98

Significance level: “*”, 0.05<p<0.10; “**”, 0.01<p<0.05; “***”, p<0.01

The above results indicate that none of the four lab games has significant predictive power for the interests in the volunteering opportunities. The low predictive power is probably due to the significant difference in the situational elements between the lab games and the second field situation. For example, the stake in the lab games is money while the

stake in the field situation is time and effort, and also the games are one-short while the latter involves a longer time horizon. The results show again the low cross-situational consistency of behaviors.[4]

E. The relationship between the two field behaviors

In subsection E, we check the relationship between the two field behaviors to see whether behavior in one situation is generalizable to another situation.

We use the whole sample of our study, i.e., participants in all the three experiments. After excluding missing values, we have in total 364 complete observations. The Pearson correlation between the two behaviors is 0.01, and the Kendall correlation is -0.01.

We run both the OLS regression and negative binomial regression. In the regression models, we regress the scores recorded in the second field situation on donations in the field. In the OLS regression, the estimated coefficient is 0.01, and the fraction of explained variance is 0.00. In the negative binomial regression, the estimated coefficient is 0.02, and the McFadden pseudo R-squared is 0.00.

The above results further confirm the low lab-field correlation for the second field situation reported in subsection D, especially the low correlation between donations in our final lab game and the interests in the volunteering opportunities. The results highlight again the power of situational elements.

F. The predictive power of pro-social personality traits

In this subsection, we check how well pro-social personality traits predict decisions in the four lab games and behaviors in the two field situations. Here we report the predictive power of three pro-social personality measures (conceptually the most relevant three ones among psychological measures we have): Agreeableness from the big five questionnaire, Perspective Taking and Empathic Concern from the Interpersonal Reactivity Index.

F.1. On the lab decisions

---

[4] As a robustness test, we also checked Kendall correlations and run negative binomial regressions. The results are quite similar to what we have in Table 3 and 4. Please refer to Table A3 and A4 for the results.

We first report how well the three pro-social personality measures predict decisions in the four lab games. The number of observations is 112 for the first two games; it is 110 for the third game; and it is 98 for the final game.

Table 5 contains the Pearson correlations between the three pro-social personality measures and decisions in the four lab games.[5]

Table 5. Correlations Between Pro-social Personality Measures and the Lab Decisions (Pearson)

	Dictator game	Donation with house money	Donation with earned money	Donation with earned money & interaction
Agreeableness	0.16*	0.04	0.05	0.07
Perspective Taking	0.15	0.22**	0.20**	-0.01
Empathic Concern	0.05	0.29***	0.14	0.00

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Now let us regress the lab decisions on our three pro-social personality measures. Table 6 and Table 7 report respectively the estimated coefficients and the proportion of explained variance in the OLS regressions.

Table 6. Regression Analysis (OLS): Lab Decisions

	Dictator game	Donation with house money	Donation with earned money	Donation with earned money & interaction

[5] Please refer to Table A5 for the Kendall correlations between the pro-social personality measures and the lab decisions.

Agreeableness	0.05*	0.02	0.02	0.02
	(0.03)	(0.04)	(0.05)	(0.04)
Perspective	0.03	0.07**	0.08**	0.00
Taking	(0.02)	(0.03)	(0.04)	(0.03)
Empathic	0.02	0.12***	0.05	0.00
Concern	(0.03)	(0.04)	(0.03)	(0.03)

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table 7. The Proportion of Explained Variance of the Lab Decisions (OLS)

	Dictator game	Donation with house money	Donation with earned money	Donation with earned money & interaction
Agreeableness	0.02	0.01	0.00	0.00
Perspective Taking	0.02	0.05	0.04	0.00
Empathic Concern	0.00	0.08	0.02	0.00

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

From the above analysis, we can see that except for the last game, all the game decisions has some moderate correlations with the three pro-social personality measures.

F.2. On the field behaviors

Now we report how well the three pro-social personality measures predict decisions in the two field situations. Putting all the observations in the three experiments together, we have 364 observations for the first field situation and 392 observations for the second.

We start with correlations. Table 8 summarizes the correlations between the three pro-social personality measures and the two field behaviors (donation in the field and the

score that measures the interest in the three volunteering opportunities).[6]

Table 8. Correlations Between the Pro-social Personality Measures and the Field Behaviors (Pearson)

	Field Donations (Pearson's ρ)	Interests in volunteering (Pearson's ρ)
Agreeableness	0.14***	0.01
Perspective Taking	0.03	-0.06
Empathic Concern	0.11**	0.06

Note: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Now we regress the two field behaviors on the three pro-social personality measures. We first report in Table 9 the results from the OLS regressions.[7]

Table 9. Regression results (OLS): Field Behaviors

	Field Donations (Estimates)	Field Donations (R-squared)	Interests in volunteering (Estimates)	Interests in volunteering (R-squared)
Agreeableness	0.04*** (0.02)	0.02	0.01 (0.02)	0.00
Perspective Taking	0.01 (0.01)	0.00	-0.02 (0.01)	0.00
Empathic Concern	0.03** (0.01)	0.01	0.02 (0.01)	0.00

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

[6] See Table A6 for the Kendall correlations between pro-social personality measures and the field behaviors.

7 To check the robustness of the results in Table 9, we run some alternative regression models: Tobit regressions for donations in the field and negative binomial regressions for interest in volunteering. Please refer to Table A7 and A8 for the results.

We can see that the pro-social personality measures have some moderate predictive power on donations in the field. If we compare the results in Table 14, 15 with the results in Table 3 and 4, we can see that the predictive power of our first two lab games only weakly dominate the predictive power of those general measures of pro-social personality traits; and that the predictive power of the last two games dominate significantly the predictive power of those general pro-social personality measures.

Like the decisions in the lab games and donations in the field, neither could the three pro-social personality measures predict well the interests in the volunteering opportunities.

4. Discussions

Like Levitt and List (2007) noted, perhaps the most fundamental question in experimental economics is whether laboratory findings could provide reliable inferences for the real world. However, for the building block of social preference studies, social preference games, the existing literature suggest that their external validity is not that high (e.g., the average correlation in the meta analysis of Galizzi and Navarro-Martinez (2019) is only 0.14).

To evaluate the external validity of social preference games, researchers need to first address the question that what correlation can be regarded as a satisfying one. Decades of research in psychology and in experimental economics showed that it is very hard for correlations of behaviors in different situations to break the barrier of 0.3 (e.g., Mischel, 1968; Ross and Nisbett, 1991); and that meanwhile the consistency of behaviors in exactly the same situation over time could be very high. For example, Benz and Meier (2008) found that the correlation between donations from their student participants in different semesters was around 0.8.

Our results in this paper confirm again the low external validity of context-free social preference game (in our case, the standard dictator game). Our results further suggest that the low lab-field correlation actually comes from some systematic differences between the lab games and the real world in social norm, stakes and other aspects. Furthermore, by identifying and eliminating these differences, not just could we bridge the lab-field gap for distributions of decisions at group level, we could also significantly enhance the predictive power of

individual decisions in the laboratory: for donations in the field, the lab-field correlation (Pearson's ρ) increased from the initial 0.2 to 0.65.

To our best knowledge, our paper is the first within-subject study that bridged the lab-field gap for some pro-social behaviors. In methodological sense, our results suggest that people's behavior actually is not that unpredictable and that experimental results could provide some reliable inference for behaviors outside the lab as long as we are cautious of the differences between the real world and our designed experiment.

But we are not the first to suspect the role of critical contextual elements in enhancing external validity of social preference games. First, a lot of researches have shown the power of some contextual elements in shaping distributions of pro-social behaviors. For example, among many others, Eckel, Grossman (1996), List (2007), and Andreoni et al. (2017) have explored the impact of social norms, earned money, and face-to-face interactions on distributions of givings. Second, some research, for example, List (2006), even has shown that eliminating some important lab-field differences could make the distributions of behaviors in experiments more close to the reality and pin down the true motivations behind some observed pro-social behaviors. Furthermore, Levitt and List (2007) has warned that some systematic lab-field differences could result in low external validity of laboratory experiments studying pro-social behaviors.

We are just surprised that very few work has been done in investigating systematically the external validity of context-free games and that even fewer has been done to test whether eliminating critical lab-field differences could enhance the external validity of lab studies.

Alekseev et al. (2017) did a literature review of experimental studies that used context-rich instructions instead of abstract ones. They suggested that context-rich language could enhance understanding of tasks; and when it changes behavior of the participants, the change is usually beneficial for the research purpose, for example, in the case that certain emotions or social norms need to be evoked. Our results suggest that experimental economics should go one step further: not only using context-rich instructions is often very necessary but also is incorporating critical elements of the situation to be studied.

In this paper, we focused on dictator games and donations. But we believe that the insights we revealed are not limited to only these cases. Our methodology could be applied in

a broader range of experimental studies of social behaviors. Some potential interesting research questions could be bridging the gap between cooperating in public goods games and cooperating in real social dilemmas,[8] the gap between reciprocating in gift-exchange games and reciprocating in the workplace,[9] the gap between cheating in lab games and cheating in various situations in the real life, etc.

Decades of research in psychology and experimental economics have accumulated tons of evidence suggesting that when critical elements of the situation change, behavior will also change, usually in predictable ways. All of these evidence suggest that experimental method is a useful and powerful tool for studying behaviors. Meanwhile, these evidences also warn we researchers to be cautious of the power of critical contextual elements when we design laboratory experiments to study social behaviors or when we try to generalize the existing laboratory findings to the real world.

Appendix

Table A1. Correlations Between the Game Decisions and Donations in the Field(Kendall)

Dictator Game	0.23***
Donating with house money	0.15*
Donating with earned money	0.35***
Donating with earned money & interaction	0.46***

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A2. Regression Analysis: Donation in the Field (Tobit)

	Coefficient	Variance Explained	No. of observations
Dictator Game	0.47**	0.04	92
	(0.19)		

[8] Stoop et al. (2012) tried to bridge the gap between cooperating in public goods game and cooperating in a real social dilemma. But their methodology is different from ours: they focused on distributions of behaviors at group level and on different aspects of lab-field differences such as subject pool, laboratory setting and contextualization.

[9] For example, Gneezy and List (2006) found that the correlation between the offered wage and workers' productivity in their two field experiments is much lower than the typical ones in gift-exchange games.

Donating with house money	0.24* (0.14)	0.03	92
Donating with earned money	0.70*** (0.14)	0.34	95
Donating with earned money & interaction	0.78*** (0.13)	0.40	90

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A3. Correlations Between the Game Decisions and Interest in Volunteering (Kendall)

Dictator Game	0.13
Donating with house money	0.10
Donating with earned money	0.02
Donating with earned money & interaction	0.02

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A4. Negative Binomial Regression Analysis: Interests in Volunteering

	Coefficient	McFadden Pseudo R^2	Number of observations
Dictator Game	0.21 (0.14)	0.01	102
Donating with house money	0.14 (0.11)	0.01	102
Donating with earned money	-0.01 (0.10)	0.00	101
Donating with earned money & interaction	0.04 (0.14)	0.00	98

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A5. Correlations Between the Pro-social Personality Measures and the Lab Decisions (Kendall)

	Dictator game	Donation with house money	Donation with earned money	Donation with earned money & interaction

Agreeableness	0.10	0.03	0.00	0.04
Perspective Taking	0.08	0.18***	0.16**	0.00
Empathic Concern	0.04	0.19***	0.09	-0.02

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A6. Correlations Between the Pro-social Personality Measures and the Field Behaviors (Kendall)

	Field donations (Kendall's τ)	Interest in volunteering (Kendall's τ)
Agreeableness	0.09**	0.00
Perspective Taking	0.00	-0.06
Empathic Concern	0.06	0.04

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A7. Regression Analysis: Donations in the Field (Tobit)

	Estimates	R-squared
Agreeableness	0.08** (0.03)	0.02
Perspective Taking	0.01 (0.03)	0.00
Empathic Concern	0.05* (0.03)	0.01

Significance level: "*", 0.05<p<0.10; "**", 0.01<p<0.05;"***", p<0.01

Table A8. Regression Analysis: Interests in Volunteering (Negative Binomial)

	Estimates	McFadden Pseudo R^2
Agreeableness	0.00 (0.03)	0.00

Perspective Taking	-0.03 (0.02)	0.00
Empathic Concern	0.03 (0.02)	0.00

Significance level: “*”, 0.05<p<0.10; “**”, 0.01<p<0.05;“***”, p<0.01

References

Aleksandr, A., Charness, G., Gneezy, U., 2017. Experimental methods: When and why contextual instructions are important. Journal of Economic Behavior & Organization, vol.134,48–59.

Andreoni, J., 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. Journal of Political Economy, vol. 97, 1447-1458.

Andreoni, J., 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. The Economic Journal. Vol. 100, No. 401, pp. 464-477.

Andreoni, J. and Rao, J., 2011. The power of asking: How communication affects selfishness, empathy, and altruism. Journal of Public Economics, vol.95, 513--520.

Andreoni, J. and Payne, A.A., 2013. Charitable giving. Handbook of Public Economics. vol. 5.

Andreoni, J., Rao, J., Trachtman, H., 2017. Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. Journal of Political Economy, vol. 125, no. 3, 625-653.

Ariely, D., Loewenstein, G., and Prelec, D., 2003. “Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences. The Quarterly Journal of Economics, Volume 118, Issue 1, Pages 73–106.

Ariely, D., Loewenstein, G., and Prelec, D., 2006. Tom Sawyer and the construction of value. Journal of Economic Behavior & Organization, 60(1), 1-10.

Auten, G.E., Sieg, H., Clotfelter, C.T., 2002. Charitable giving, income, and taxes: An analysis of panel data. American Economic Review, 92(1): 371-82.

Bardsley, N., 2008. Dictator game giving: altruism or artefact? Experimental Economics 11, 122--133.

Benet-Martínez, V. and John, O. P., 1998. Los Cinco Grandes across cultures and ethnic groups: Multitrait–multimethod analyses of the Big Five in Spanish and English. Journal of Personality and Social Psychology, 75, 729–750.

Benz, M. and Meier, S., 2008. Do people behave in experiments as in the field?---evidence from donations. Experimental Economics, 11: 268--281.

Branas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. Journal of Economic Behavior and Organization, 60, 306–320.

Camerer, C., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton.

Camerer, C., 2011. The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Working paper, California Institute of Technology, Pasadena.

Carpenter, J., Connolly, C. & Myers, C. K., 2008. Altruistic behavior in a representative dictator experiment. Experimental Economics 11(3):282–98.

Carpenter, J.P., and Myers, C.K. (2010). Why volunteer? Evidence on the role of altruism, reputation, and incentives. *Journal of Public Economics*, 94, 911-920.

Charness, G. and Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics, 117, 817-69.

Chen, D.L., Schonger, M., Wickens, C., 2016. oTree - An open-source platform for laboratory, online and field experiments. Journal of Behavioral and Experimental Finance, vol 9: 88-97.

Condon, D. and Revelle, W., 2014. The international cognitive ability resource: Development and initial validation of a public-domain measure. Intelligence, 43, 52-64.

Dana, J., Cain, D.M., Dawes, R.M., 2006. What you don't know won't hurt me: Costly (but quiet) exit in dictator games. Organizational Behavior and Human Decision Processes, Volume 100, Issue 2, Pages 193-201.

Davis, M. H., 1980. A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

DellaVigna, S., List J.A., and Malmendier, U., 2012. Testing for altruism and social pressure in charitable giving. The Quarterly Journal of Economics Vol. 127, 1 – 56.

Dillenberger, D. and Sadowski, P., 2012. Ashamed to be selfish. Theoretical Economics, Volume 7, Issue 1, Pages 99--124.

Ellingsen, T. and Johannesson, M., 2008. Anticipated verbal feedback induces altruistic behavior. Evolution and Human Behavior 29, 100--105.

Eckel, C.C. and Grossman, P.J., 1996. Altruism in anonymous dictator games. Games and Economic Behavior, 16, 181-191.

Fleeson, W., 2001. Towards a structure- and process-integrated view of personality: Traits as density distributions of states. Journal of Personality and Social Psychology, 80, 1011--1027.

Fleeson, W., 2004. Moving personality beyond the person-situation debate the challenge and the opportunity of within-person variability. Current Directions in Psychological Science, 13(2), 83-87.

Franzen, A. and Pointner S., 2012. Anonymity in the dictator game revisited. Journal of Economic Behavior & Organization, 81 (2012), 74--81.

Franzen, A., and Pointner, S. (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*, 16, 155-169.

Galizzi, M. and Navarro-Martinez, D., 2019. On the external validity of social preference games: a systematic lab-field study. Management Science 65(3): 976-1002.

Glazer, A. and Konrad, K.A. 1996. A signaling explanation for the charity. The American Economic Review. Vol. 86, No. 4, pp. 1019-1028.

Gneezy, U. and List J., 2006. Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. Econometrica, vol. 74, No. 5, pp. 1365-1384.

Giving USA: The Annual Report on Philanthropy for the Year 2018 (2019). Chicago: Giving USA Foundation.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., et al. (2006). Costly punishment across human societies. Science, 312, 1767–1770.

Holt, C. A. and Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92(5): 1644-1655.

Hoolwerf B., Schuyt T. (eds.), 2017. Giving in Europe. The State of Research on Giving in 20 European Countries. Lenthe Publishers. Amsterdam.

John, O.P., Donahue, E. M., & Kentle, R. L., 1991. The "Big Five" Inventory--Versions 4a and 54. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.

Karlan, D. and List, J.A., 2007. “Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment.” American Economic Review, 97(5): 1774-1793.

Lichtenstein, S., and Slovic, P. (Eds.). 2006. The construction of preference. Cambridge University Press.

Levitt, S., and List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world. Journal of Economic Perspectives, 21, 153-174.

List, J.A., 2006. The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. Journal of Political Economy, Vol. 114, No. 1, pp. 1-37.

List, J.A., 2007. On the interpretation of giving in dictator games. Journal of Political Economy 115 No.3, 482-493.

List, J.A., 2011. The Market for Charitable Giving. Journal of Economic Perspectives, Vol. 25, No. 2, pp. 157-180.

Loewenstein, G., 1999. Experimental economics from the vantagepoint of behavioural economics. Econom. J. 109(453):25–34.

Kahneman, D., Knetsch, J. and Thaler, R., 1986. Fairness and the Assumptions of Economics. The Journal of Business, Vol. 59, No. 4, pp.S285-S300.

Kolstad, J.R., and Lindkvist, I., 2012. Pro-social preferences and self-selection into the public health sector: evidence from an economic experiment. Health Policy and Planning, 1-8.

Konow, J., 2001. Fair and square: the four sides of distributive justice. Journal of Economic Behavior & Organization, Vol. 46, 137–164.

Konow, J., 2010. Mixed feelings: Theories of and evidence on giving. Journal of Public Economics 94, 279--29.

Mischel, W., 1968. Personality and Assess- ment. New York; Wiley.

Ottoni-Wilhelm, M., Vesterlund, L., and Xie, H., 2017. Why Do People Give? Testing Pure and Impure Altruism. American Economic Review, 107(11): 3617-3633.

Okten, C. and Weisbrod, B., 2000. Determinants of donations in private nonprofit markets. Journal of Public Economics. Volume 75, Issue 2, Pages 255-272.

Pérez-Albéniz, A., Paúl, J., Etxeberría, J. Montes, M.P. y Torr, E. 2003. Adaptación del Interpersonal Reactivity Index (IRI) al español. Psicothema, 15, 267-272.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. ,2015. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). Journal of Behavioral Decision Making, 29: 453–469.

Ross, L. and Nisbett, R.E., 1991. The Person and the Situation: Perspectives of Social Psychology. New York: McGraw-Hill.

Slovic, P., 1995. The construction of preference. American psychologist, 50(5), 364-371.

Stewart, N., Reimers, S., and Harris, A. J. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. Management Science, 61, 687-705.

Stoop, J., Noussair, C. N. and Soest, D., 2012. From the Lab to the Field: Cooperation among Fishermen. Journal of Political Economy, Vol. 120, No. 6, pp. 1027-1056.

Stoop, J. (2014). From the lab to the field: Envelopes, dictators and manners. *Experimental Economics,* Volume 17, Issue 2, pp 304–313.

Thomson,K. and Oppenheimer, D., 2016. Investigating an alternate form of the cognitive reflection test. Judgment and Decision Making, Vol. 11, No. 1,99–113.