

EMPIRICAL ARTICLE

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

Xinghua Wang<sup>1</sup> and Daniel Navarro-Martinez<sup>2,3,4</sup>

<sup>1</sup>Institute for Advanced Economic Research, Dongbei University of Finance and Economics, Dalian, China; <sup>2</sup>Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain; <sup>3</sup>Barcelona School of Economics, Barcelona, Spain and <sup>4</sup>Barcelona School of Management, Barcelona, Spain

**Corresponding author:** Daniel Navarro-Martinez; Email: [daniel.navarro@upf.edu](mailto:daniel.navarro@upf.edu)

**Received:** 9 November 2022; **Revised:** 26 April 2023; **Accepted:** 26 April 2023

**Keywords:** economic experiments; external validity; dictator games; charitable giving; volunteering

## Abstract

There is growing concern about the extent to which economic games played in the laboratory generalize to social behaviors outside the lab. Here, we show that it is possible to make a game much more predictive of field behavior by bringing contextual elements from the field to the lab. We report three experiments where we present the same participants with different versions of the dictator game and with two different field situations. The games are designed to include elements that make them progressively more similar to the field. We find a dramatic increase in lab–field correlations as contextual elements are incorporated, which has wide-ranging implications for experiments on economic decision making.

## 1. Introduction

Research on what has been called ‘social preferences’ has attracted great interest in the last decades and has been one of the main building blocks of behavioral and experimental economics. This line of research is characterized by being largely based on studying behavior in economic games that are intended to capture different aspects of social behavior, such as altruism, cooperation, inequity aversion, reciprocity, and trust. Examples of these games include the dictator game, ultimatum game, trust game, and public good game, among others (e.g., Berg et al., 1995; Camerer, 2003; Camerer and Thaler, 1995; Charness and Rabin, 2002; Falk et al., 2022; Fehr and Gächter, 2000, 2002; Fischbacher and Gächter, 2010; Forsythe et al., 1994; Guth et al., 1982). These social preference games typically share some common features: they are closely based on game-theoretic structures and have clear game-theoretic equilibria; they are deliberately as context-free as possible, in the sense of not incorporating elements that resemble particular real-world contexts; and their outcomes consist in monetary payoffs for the players involved. This approach to social behaviors has been hugely influential and has become one of the benchmarks for the study of human interaction in economics, judgment and decision making, and beyond (e.g., Glimcher and Fehr, 2008; Henrich et al., 2001, 2004, 2010; Lee, 2008; Mazar and Zhong, 2010; Piff et al., 2010, 2015; Rand et al., 2012; Rilling and Sanfey, 2011; Sanfey, 2007; Shariff and Norenzayan, 2007; Zhong et al., 2010).

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 very tightly controlled. Moreover, this facilitates different research groups using the same tools in the same way, so that evidence can quickly accumulate. In fact, experimental economists have advocated the use of context-free games to study behavior for reasons related to internal validity since the early days of the discipline (Smith, 1976).

However, the artificial and context-free nature of these games makes them particularly low in mundane realism (Aronson and Carlsmith, 1968), which raises a 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 the social and behavioral sciences is to understand real-world social behavior beyond the confines of the lab, 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 address systematically this concern (but see Loewenstein, 1999, for a shorter discussion). These 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 the lab games and relevant 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 the external validity of social preference games (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 still 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 some issues of external validity.

These external validity problems might not be too surprising, given that behavioral research has widely documented that context plays a crucial role in economic behavior (e.g., Ariely et al., 2003, 2006; Lichtenstein and Slovic, 2006; List, 2007; Olivola et al., 2020; Slovic, 1995; 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 (Epstein, 1979; Epstein and O'Brien, 1985; Fleeson, 2001, 2004; Fleeson and Nofhle, 2009; Mischel, 1968; Ross and Nisbett, 1991). Overall, if behavior is so determined by contextual elements, it seems logical that context-free economic games struggle to 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 increase the external validity of social preference research by introducing more realistic 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 social preference games much more predictive of field behavior by taking relevant contextual elements from the field and incorporating them to the game situations. To do this, we focus on the dictator game (Kahneman et al., 1986) and charitable giving in the field. In typical implementations of the dictator game, two players are matched together and randomly assigned to the roles of Player 1 (the 'dictator') and Player 2 (the 'recipient'). Player 1 has to make an anonymous decision about how to split between the two players a certain amount of money provided by the experimenter. Player 2 is simply a passive recipient of the money allocated to them by Player 1. 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., Benz and Meier, 2008; Branas-Garza, 2006; Carpenter et al.,

2008; Eckel and Grossman, 1996; Kolstad and Lindkvist, 2012; Konow, 2010), 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 (Bardsley, 2008; List, 2007), earning the monetary endowment (Cherry et al., 2002; List, 2007), the presence of verbal feedback and face-to-face interaction (Andreoni et al., 2017; Andreoni and Rao, 2011; DellaVigna et al., 2012; Ellingsen and Johannesson, 2008), 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 United States in the year 2021 was \$485 billion (Giving USA, 2022); the annual amount in Europe has been estimated to be €88 billion (Hoolwerf and Schuyt, 2017). Charitable giving is also a widely researched topic (e.g., Andreoni, 1989, 1990; Andreoni et al., 2017; Andreoni and Payne, 2013; Auten et al., 2002; Bekkers and Wiepking, 2011; DellaVigna et al., 2012; Glazer and Konrad, 1996; Karlan and List, 2007; List, 2011; Okten and Weisbrod, 2000; Oppenheimer and Olivola, 2011; Piliavin and Charng, 1990) and, as we said, a domain that has been directly linked to the dictator game.

So, we investigated if it is possible to make a dictator game situation more predictive of charitable giving in the field. To this end, we ran a series of three interconnected experiments in which we presented the same participants with different versions of a dictator game and with two naturalistic field situations that we created, where 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 charity information. Our different versions of the dictator game were designed to incorporate step-by-step contextual elements that made the game situation 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). This 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.<sup>1</sup> None of these papers, however, compared lab and field behaviors for the same sample of participants.

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 limited predictive power in relation to specific field behaviors, but this power can be very substantially increased by incorporating appropriate contextual elements from the field to the game situations.

These findings have wide-ranging implications for social preference experiments. Our interpretation is that social preference research should seriously consider using more naturalistic contexts in laboratory experiments and be more explicit about the types of field behaviors it aims to address. This conclusion is in line with similar, long-standing proposals to increase the representativeness of experimental designs in other areas of judgment and decision making (e.g., Brunswik, 1955, 1956; Dhami et al., 2004; Hammond, 1966, 1986, 1990; Hogarth, 2005). In more practical terms, social preference experiments with more realistic designs could also be a useful way to combine some of the virtues of both lab experimentation and field experimentation, which is gaining momentum in economics (Banerjee and Duflo, 2009; Duflo and Banerjee, 2017; Harrison and List, 2004; Levitt and List, 2009). On the one hand, contextualized lab experiments can maintain the higher levels of internal validity typically associated with lab experiments and they can be more cost-effective than experiments in the field. On the other hand, these experiments can have higher external validity than context-free experiments, at least in relation to the particular field environments they target. Contextualized social

<sup>1</sup> See Kachurka et al. (2021) for a similar example in the domain of decision making under risk and uncertainty.

preference experiments could also serve as an initial test bed for particular policies and behavioral interventions.

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

Our design was based on a series of three interconnected experiments, so we explain the methods of all three together. In the next five subsections, we describe the following five elements: 1) the general structure and procedures of our three-experiment setup, 2) an initial pilot study we ran to select the charities used in our experiments, 3) the two field situations that served as our benchmark of field behavior that the lab games aimed to predict, 4) the different lab games we created to approach step-by-step one of the field situations, and 5) how all the field situations and lab tasks were organized and implemented in each of the three experiments and in the different experimental sessions, including a detailed description of the participants.

### 2.1. General structure and procedures

As [Table 1](#) shows, each experiment comprised three different days. Participants were required to come to the lab on the first 2 days. On Day 1, the participants in the main groups were presented with one or two of our four lab games (as shown in [Table 1](#) and explained in detail in [Sections 2.4](#) and [2.5](#)).<sup>2</sup> In addition, they responded to some psychological questionnaires (see [Section 2.5](#) for details). We also elicited some basic demographic variables (age, gender, study program, study year, and citizenship).

In Experiments I and III (the ones with the games that were least and most similar to our first field situation), we also included control groups of participants who did not play the games on Day 1 but participated in the rest of the experiment. This allows us to check if their field behaviors on Days 2 and 3 were different from the ones of participants who had previously played the games. On Day 1, instead of the games, these participants responded to other tasks unrelated to social preferences.

The participants came back to our lab a few days later. On this day (Day 2), all of them responded to non-incentivized filler tasks that were unrelated to social preferences and will not be used in the paper. These tasks included cognitive ability and cognitive reflection questions, and hypothetical choices between gambles.

At the end of the Day 2 session, people were paid €15 in cash for their participation in both lab sessions. When participants left the room after the session and the payment had finished and were walking through a square, they encountered the first field situation (as described in [Section 2.3.1](#)).

**Table 1.** *Structure of the study.*

|                |               | Day 1<br>(lab games)             | Day 2<br>(field) | Day 3<br>(field) | <i>N</i><br>(Day 1) | <i>N</i> (Days 2<br>and 3) |
|----------------|---------------|----------------------------------|------------------|------------------|---------------------|----------------------------|
| Experiment I   | Main group    | Games 1<br>and 2<br>(randomized) |                  |                  | 112                 | 102                        |
|                | Control group | None                             | Solicitor        | Email            | 49                  | 46                         |
| Experiment II  | Main group    | Game 3                           |                  |                  | 110                 | 101                        |
| Experiment III | Main group    | Game 4                           |                  |                  | 116                 | 98                         |
|                | Control group | None                             |                  |                  | 53                  | 45                         |

<sup>2</sup>All the lab games were implemented using oTree (Chen et al., 2016).

**Table 2.** Pilot results for the four charities used in the experiments.

| Charity   | Main aim of the charity                        | Familiarity | Allocation |
|-----------|--|-------------|------------|
| Charity 1 | Helping children in developing countries       | 96%         | €12.27     |
| Charity 2 | Helping disadvantaged children and adolescents | 100%        | €12.25     |
| Charity 3 | Protecting human life and health               | 100%        | €8.60      |
| Charity 4 | Providing medical-humanitarian assistance      | 99%         | €13.12     |

A few weeks later, the participants received an email with a link to information about volunteering opportunities that we used to implement the second field situation (as described in Section 2.3.2). In case some participants missed the email, we sent another identical message about 1 month after the first one.

## 2.2. Pilot study

In our experiments, we used four different charities (as detailed in the following subsections). To select these charities, we ran a pilot study with 106 local university students. We asked the participants questions about 12 charities operating in the city, including international and local ones pursuing various causes. Specifically, they were asked how well they knew these charities (with four response options: ‘I never heard of it’, ‘I know it a bit’, ‘I know it well’, and ‘I know it very well’), and how they would hypothetically allocate €100 among them.

Based on the results, we chose the four charities to be used in our experiment following two steps. First, we selected the charities that at least 95% of participants had heard of before. Six charities met this standard. Then we further selected out of those six the four charities that were allocated the highest amounts of money. Table 2 summarizes the main results obtained for the four chosen charities. The familiarity score indicates the percentage of participants that had heard of a charity before; the allocation score refers to the average amount of money allocated to a charity.

## 2.3. The field situations

We used two naturalistic field situations related to charities, in which participants could behave pro-socially: one in which they had the chance to donate money to a charity and one in which they could show interest in volunteering by checking information.

### 2.3.1. Charitable giving

The first situation took place in a large square on our university campus. A research assistant solicited donations there for Charity 1 (Table 2). This square is a popular location in the university, where students gather between classes, and it is often used to organize various activities, including collecting donations for different causes. Therefore, this was a natural place to find a charity solicitor. The square was also next to the room where we conducted our experiment on Day 2, so that we knew participants had to walk through the square when they left the room at the end of the session. This allowed us to present all our participants with the same field situation in a naturalistic setting.

The assistant (always female) was wearing an official university T-shirt and was standing next to a table with a professional charity bucket that had a large sticker with the logo of the charity. She also had a laminated color-printed leaflet with information about the charity and its projects. As each participant approached the location of our assistant, she spoke to them using always the same initial words: ‘Hello, I am collecting donations for Charity 1, would you like to contribute?’ In the case that the participants wanted more information, she was instructed to show them the leaflet. If the participants decided to donate, they simply put the money they wanted inside the bucket, and this was typically the end of the interaction. Then the assistant recorded the donations made by the participants, together with any other relevant comments about their behavior.

At the end of the Day 2 session, before encountering this field situation, participants were paid €15, using always the same bill and coin denominations: two €5 bills, one €2 coin, two €1 coins, one 50 cent coin, two 20 cent coins, and one 10 cent coin. These denominations can be used to make up any amount from €0 to €15 in increments of €0.1, which assured that all participants had cash to make any donation to charity up to €15 when they faced this field situation. We paid people one by one and made sure that each participant left the room approximately every 3 minutes, so that there was enough time to complete the field situation.

Before implementing this field situation, we were granted approval by the university and the charity. At the end of the study, all the money donated was sent to the charity.

### **2.3.2. Interest in volunteering**

For the second field situation, a few weeks after the Day 2 sessions, we sent our participants an email with a link to a website where they could check information about volunteering opportunities at three different charities. The email was sent from an address linked to the university but unrelated to our lab or our experiments. The subject of the email was the name of the university plus the word ‘volunteering’. The email said: *‘We have constructed a platform that provides information to students at our university about volunteering opportunities at charities. Each time we select three volunteering opportunities that suit students. If you are interested, please click the following link to learn more’*.

If participants clicked the link, they were directed to a website where they could see the names of three real volunteering projects managed by three real charities (Charities 1–3 described in [Table 2](#)). Participants were able to click the volunteering project they wanted and go to a page where they could see a brief description of the project. At the bottom of that page, people were asked *‘Are you interested in volunteering for this program?’* They could answer ‘yes’ or ‘no’ to this question. This was followed by a page where participants could click a link to go to the charity’s website if they wanted to learn more about that particular volunteering project or to sign up for it. After having checked the information about one volunteering project, participants were automatically redirected back to the initial page where they could click on the links for the other volunteering projects if they wanted. This setup allowed us to measure people’s interest in volunteering by tracking their information search behavior (i.e., the links they clicked).

One of the three projects was about ‘digital volunteering’. This allowed enrollees to volunteer at home and collaborate in many activities, such as translating documents, helping to create and manage websites, or working with old people and AIDS patients from a long distance. The other two projects needed volunteers to help organize events or do office work.

### **2.4. The lab games: bridging the gap between the lab and the field**

To investigate whether dictator games can be made more predictive of field behavior by bringing contextual elements from the field into the lab, we created a series of four different game situations. Each of them incorporates one additional field element that makes the lab environment more similar to our first field situation. Our methodology requires us to focus on one single field situation to try to approximate its specific characteristics. We chose our first field situation for this because, as explained in the Introduction, charitable giving is a domain of high economic and social relevance, an extensively researched topic, and a setting that has been directly linked to the dictator game. Our second field situation, which has markedly different characteristics, will serve as a comparison benchmark to illustrate the importance of the specific contextual features of the situation.

To create our lab environments, we focused on three main aspects in which the standard dictator game, as typically implemented in economic experiments, differs from our first field situation: 1) participants in the standard game share the money with another study participant, as opposed to people in serious need; 2) they have to allocate ‘house money’ assigned to them by the experimenters, as opposed to ‘earned money’ that belongs to them; and 3) they do not engage in face-to-face interaction when making their donation. There are of course more differences between the standard dictator game

**Table 3.** *The lab games.*

| Games  | Description                                       |
|--------|---|
| Game 1 | Standard dictator game                            |
| Game 2 | Charity as a recipient                            |
| Game 3 | Charity + earned money                            |
| Game 4 | Charity + earned money + face-to-face interaction |

and our first field situation, but we decided to focus on these three because we believe that they are of particular relevance, and all three have been shown to significantly affect behavior in dictator game settings (e.g., Andreoni et al., 2017; Eckel and Grossman, 1996; List, 2007).

Table 3 summarizes the structure of our four game situations. Game 1 in the series was the standard dictator game. Player 1 (the ‘dictator’) was endowed with €5 and had to split that amount between themselves and a passive Player 2 (the ‘recipient’). In this case, both players were randomly matched participants. In Game 2, we changed the recipient of the money to a charity involved in medical-humanitarian actions (Charity 4 described in Table 2). As shown in Table 2, this charity was similar but not the same as Charity 1 (used in the first field situation). This was meant to reduce cross-contamination between lab and field decisions while keeping them comparable. In Game 3, participants first earned their money (always €5) in a real-effort task and then decided how much to give to the same charity. The real-effort task involved entering sequences of letters and symbols on the computer. In Game 4, we further included face-to-face interaction. Specifically, instead of inputting the money they wanted to donate on the computer, participants had to give the money to an assistant (always female) collecting donations for the same charity, who was standing at the door of the lab with a bucket. More details of these game situations will be explained in the following subsection.

This pyramidal structure of our experimental design allows us to investigate how the lab–field correlations evolve as we include the aforementioned contextual factors step-by-step. The three factors were included in this particular order because this is arguably the most meaningful and convenient way to organize the experiments. First, the element of having a charity as a recipient (Game 2) is the one that fits best together with the standard game (Game 1) in the same experiment. Having the same dictator game decision with and without earned money is likely to lead to more cross-contamination, and having a person collecting donations requires that the money goes to a charity. Second, having someone collecting donations is the least usual element in economic experiments and the one that changes the physical setup the most, so we decided to include it last. Nevertheless, it is of course possible that there are interactions between the three contextual elements that we included, so that different orders of incorporating them can lead to different incremental effects of each factor on the lab–field correlation. In any case, this would not undermine the validity of the test of our main hypothesis, which is that it is possible to make a lab game more predictive of a given field behavior by eliminating contextual differences between the lab game and the field situation.

## 2.5. Detailed experimental procedures

In this subsection, we explain how all of the lab tasks and the field situations were organized and implemented in the three different experiments and the different experimental sessions. We explain the details of each experiment in turn and then provide detailed information about our participants.

### 2.5.1. Experiment I

In Experiment I, participants in the main group faced two of our four game situations on Day 1: the standard dictator game (Game 1) and a modified game in which the recipient was Charity 4 (Game 2). In these two games, people made decisions about splitting a €5 endowment between themselves and

either another randomly matched participant (Game 1) or Charity 4 (Game 2). Game 1 was played for two rounds. In the second round, the roles of the first and second players were reversed and people were rematched with other participants. Therefore, all the participants got to play the game as the first player, who had to decide how to split the money. In Game 2, participants were first shown a brief description of Charity 4, and then they made their decision of how much of their €5 endowment to donate. The two lab games were presented in a random order. At the end of the Day 1 session, one of the game rounds was randomly selected for each participant and they all received the amount resulting from the game.

Participants in the control group did not face the games on Day 1, as indicated in [Section 2.1](#). Instead of the games, they responded to a series of hypothetical filler choices about inter-temporal trade-offs that were unrelated to social preferences, and they were paid a fixed amount of €3 at the end of the session. This control group can be used to check if playing the games on Day 1 affected in any way the field behavior of the participants on Days 2 and 3.

Participants in both groups also responded to the Interpersonal Reactivity Index scale (Davis, 1980) and a Big-Five personality questionnaire (John et al., 1991), and they provided their demographic information. The Interpersonal Reactivity Index measures four dimensions of empathy: Perspective Taking, Empathic Concern, Fantasy, and Personal Distress. These aspects of empathy are important determinants of social interactions, and they have been shown to correlate with pro-social behavior (e.g., Borman et al., 2001; Dovidio and Penner, 2001; Penner, 2002). The Big-Five questionnaire covers five personality dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These five aspects of personality are fundamental constructs in psychology, and they have been shown to explain a variety of behaviors (e.g., Barrick and Mount, 1991; Giluk and Postlethwaite, 2015; Paunonen, 2003; Prinzie et al., 2009). The main goal of these questionnaires was to compare their predictive power in relation to field behaviors with that of our games.

A few days after Day 1, participants went on to participate in Day 2 and then Day 3, as described in [Sections 2.1](#) and [2.3](#).

### 2.5.2. Experiment II

On the first day of Experiment II, participants were presented with Game 3, which was like Game 2 (i.e., with Charity 4 as the recipient) but introducing the element of earned money. To earn their money, participants did a real-effort task which consisted in entering 17 sequences of letters and numbers on the computer. This real-effort task has been used and validated in several previous experiments (see, e.g., Charness et al., 2018; Dickinson, 1999, for a broader review of real-effort tasks). All participants received €5 for completing the real-effort task. Then, like in Experiment I, they were shown the description of Charity 4 and were asked how much of their €5 they wanted to donate.

As in Experiment I, participants responded to the Interpersonal Reactivity Index and Big-Five questionnaires and provided demographic information. They afterward participated in Day 2 and then Day 3.

### 2.5.3. Experiment III

On Day 1 of Experiment III, participants faced Game 4. This game was like Game 3 but introducing face-to-face interaction. So, first people earned their money in the same real-effort task as in Experiment II. In this case, they earned €4.9 in this task (instead of €5) to make it more natural to pay them in coins. After that, they were shown the description of Charity 4 and they were told that we had an assistant collecting money for this charity next to the outside door of the lab and that they could donate money to the charity if they wanted. Then participants were paid their €4.9 and left the lab one by one. They were always paid in the same coin denominations: one €2 coin, two €1 coins, one 50 cent coin, one 20 cent coin, and two 10 cent coins. This allowed the participants to donate any amount they wanted from €0 to €4.9 in increments of €0.1.

We had an assistant (always female) standing next to the outside door of the lab and holding a bucket labeled as ‘donation box’. She was instructed to ask each participant that came out of the lab the same question: ‘*Hello, I am collecting donations for Charity 4, would you like to contribute?*’ The assistant



recorded the amounts donated and was instructed not to ask for any personal or identifying information. Note that this last version of the game situation (Game 4) is much more similar to our first field situation than the standard dictator game, but everything happens in the lab and people know at all times that they are participating in an experiment.

Also, in this experiment, participants filled out the Interpersonal Reactivity Index and Big-Five questionnaires and we elicited their demographic information. They also participated in Days 2 and 3, like in the other experiments.

As in Experiment I, we included a control group of people who did not participate in Game 4 on Day 1 of the experiment, which allows us to check for any effect of having faced the game on subsequent field behavior. People in the control condition were paid €4 for completing the real-effort task and this was their payment for the session.

#### 2.5.4. Participants

Overall, 440 people participated in our experiments. All of them were recruited from our university subject pool, which consists of about 7,000 people who are mainly current and former students of the university. People could only participate in one of the experiments reported here, and there were no other eligibility or exclusion criteria.

A total of 161 people participated in Experiment I on the first day, 112 in the main group and 49 in the control condition. Due to attrition, we lost some participants on the second day, so that we ended up with a sample of 148 people who participated in the whole experiment, 102 in the main group and 46 in the control condition. A total of 110 people participated in the first session of Experiment II, and 101 of them participated in both sessions. In Experiment III, we had a total of 169 participants in the first session, 116 in the main group and 53 in the control condition. Due to attrition, we ended up with 98 participants in the main group and 45 in the control condition. The attrition rate in Experiment III was higher because the university campus was closed at the scheduled time for some of our Day 2 sessions due to a strike. Overall, our participants were 60% female, with an age range between 18 and 54 and an average age of 21.21.

### 3. Results

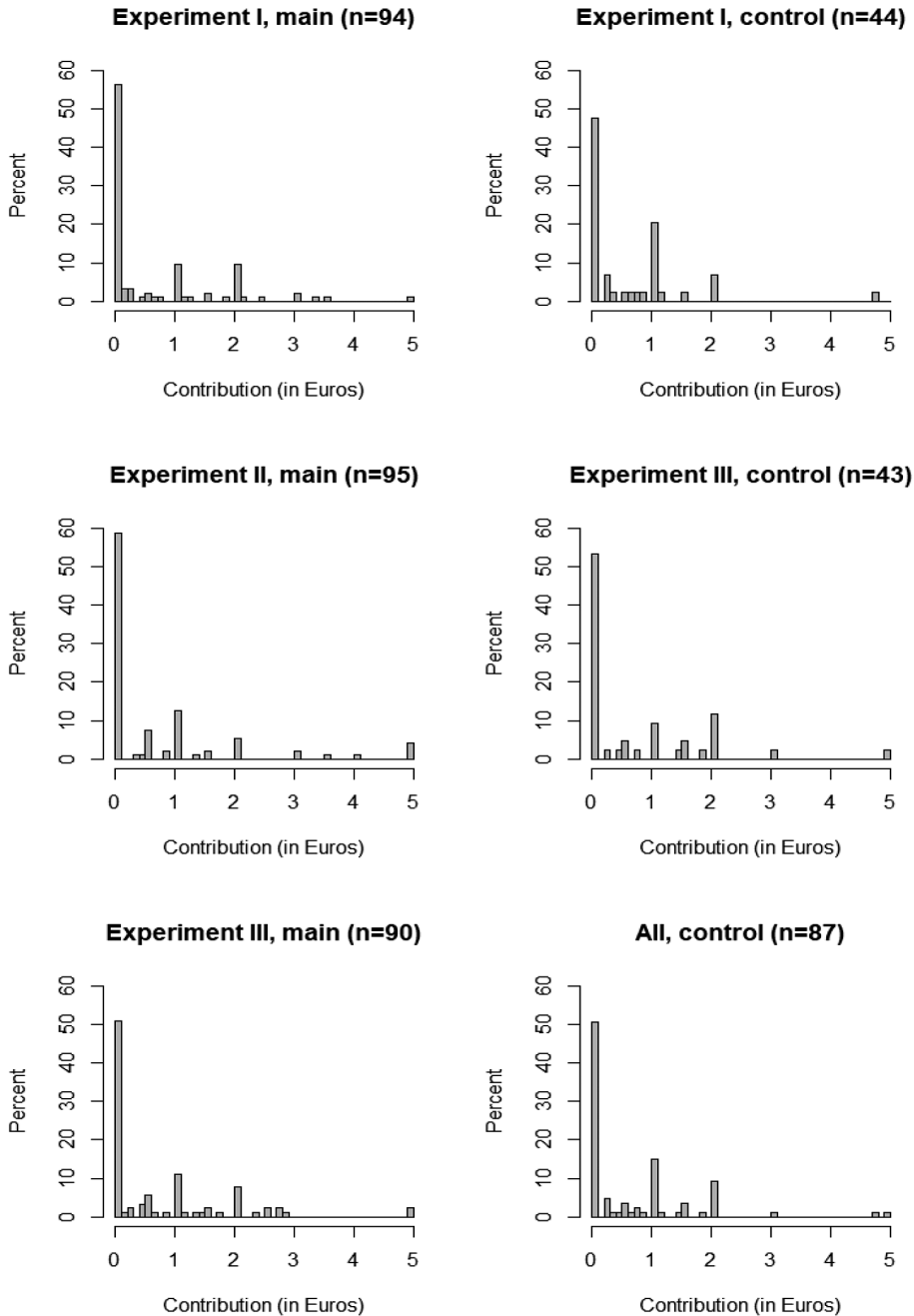
We now present the results, organized in five separate subsections. In [Sections 3.1](#) and [3.2](#), we describe the patterns of decisions observed in the two field situations. Then, in [Section 3.3](#), we introduce the decisions made in the four lab games. In [Section 3.4](#), we present our main analysis of the predictive power of the different lab games in relation to the field behaviors. In [Section 3.5](#), we analyze the relationship between the two different field situations. Finally, in [Section 3.6](#), we explore the explanatory power of the additional psychological measures we elicited in relation to the field behaviors.

#### 3.1. Donations in the field

The left column of [Figure 1](#) shows the distributions of donations in the main groups (excluding the control conditions) in our first field situation across the three experiments; the right column presents the distributions of donations in the control groups of Experiments I and III, separately and together.<sup>3</sup>

As [Figure 1](#) shows, donation behavior in the first field situation was remarkably similar across the main groups of all three experiments. This indicates that the specific games played in the lab on Day 1, which is the main difference between experiments, did not substantially affect decisions in the first

<sup>3</sup>The first field situation has some missing observations due to incidental factors that made it impossible for our assistant to properly approach some particular participants. We also excluded from our main analysis four observations for which the participants explicitly told the solicitor that they did not want to donate because they had recently donated to the charity used on Day 1 (Charity 4). Including these four observations does not change the fundamental pattern of results obtained (the analysis is available from the authors on request).



**Figure 1.** Distributions of donations in the main groups and the control conditions.

field situation. This conclusion is further reinforced by the distributions of field donations in the control conditions, which are also very similar to the ones in the main groups (Mann–Whitney test comparing main group and control in Experiment I:  $U = 1,964, p = 0.61$ ; in Experiment III:  $U = 1,969, p = 0.86$ ).

In terms of the shape of the distributions, they are all markedly right-skewed, with high spikes at zero (the percentages of zeros in the six distributions shown are 56%, 59%, 50%, 45%, 53%, and 49%, respectively). This type of distribution broadly matches previous donation data in similar settings (e.g., Galizzi and Navarro-Martinez, 2019). All the distributions have maximum values of 5.

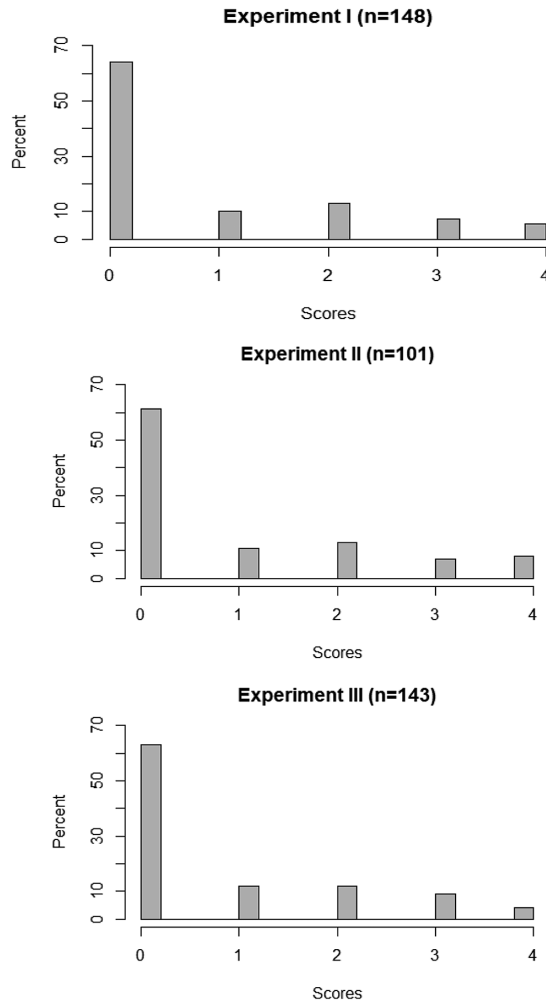


Figure 2. Distributions of interest in volunteering.

### 3.2. Interest in volunteering

Figure 2 shows the distributions of interest in volunteering scores in the second field situation across the three experiments. We coded behavior in this situation as a numeric variable between 0 and 4. If a participant did not even click the initial link embedded in the email, the variable takes a value of 0. If the participant clicked the link but did not check information about any of the volunteering opportunities, the value is 1. We gave a value of 2 to the participants who checked one volunteering opportunity but did not express interest in it; that is, they either did not answer the question ‘Are you interested in volunteering for this program?’ or answered ‘no’. If a participant checked one volunteering opportunity and also expressed interest in it, by answering ‘yes’ to the question, or answered ‘no’ but checked at least one other opportunity, the variable takes a value of 3. Finally, a value of 4 indicates that the participant either answered ‘yes’ to at least one opportunity and also checked at least another one, or that they clicked at least one of the links leading to the websites of the charities.<sup>4</sup> As explained in Section 2, we sent the second identical message about 1 month after the first one. The responses

<sup>4</sup>The main patterns obtained for interest in volunteering are robust to different ways of coding this variable. Additional analyses are available from the authors on request.

were coded the same way, and the highest value of the two rounds was recorded as the final interest in volunteering score of the participant.

The distributions of interest in volunteering are very similar across experiments. They are also right-skewed, with many participants showing no interest (64%, 61%, and 63% in the three experiments, respectively). A smaller fraction of people clicked the link embedded in the email but left the page without checking any volunteering opportunities (10%, 11%, and 12%, respectively), and a similar proportion checked the information about one program but did not express interest (13%, 13%, and 12%). A slightly smaller fraction of the participants expressed interest or checked more than one opportunity (7%, 7%, and 9%). Finally, some participants were very interested (5%, 8%, and 4%): they expressed interest and checked more than one program or visited the websites of the charities to learn more about them or sign up.

### 3.3. *Decisions in the lab games*

Figure 3 presents the distributions of decisions in our four game situations. The distribution in the standard dictator game (Game 1) is broadly in line with the results obtained in the literature: it is positively skewed with a relatively high spike at zero (25%) and a large proportion of people around half of the endowment (24% at 2 and 13% at 2.5) (Camerer, 2003). The distribution for Game 2 is notably flatter than the one for Game 1, with less people at zero (15%) and more at the higher end of the distribution, including 13% of the people at 5 (the full endowment). This means that participants gave on average substantially more when the recipient was a charity (the average donation was 1.38 in Game 1 and 2.24 in Game 2; Wilcoxon signed-rank test:  $W = 674, p = 0.00$ ), which is consistent with previous literature (e.g., Eckel and Grossman, 1996).<sup>5</sup>

In Games 3 and 4, the distributions are much more positively skewed, with higher spikes at zero (47% in Game 3 and 40% in Game 4). The average contribution in these games was 1.34 and 1.21, respectively. It is also apparent that the distributions in these games look much more similar to the distributions of donations obtained in the first field situation.

### 3.4. *Do the lab games predict the field behaviors?*

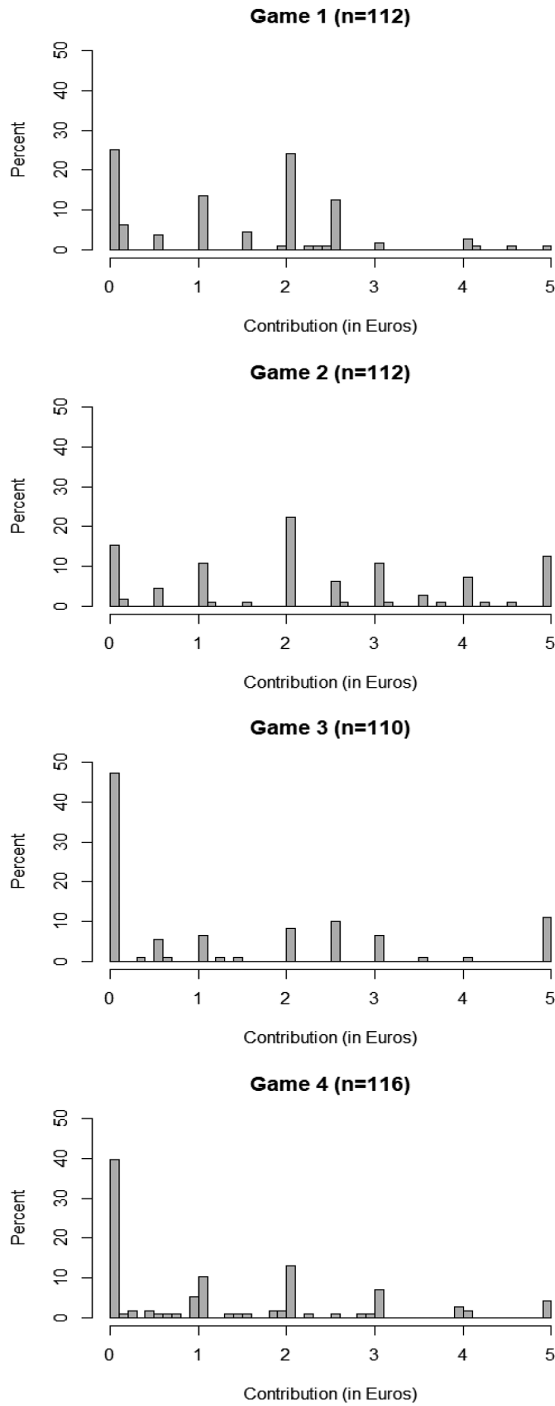
We now turn to our key objective of analyzing the extent to which the different lab games can predict the field behaviors. This will reveal the predictive power of the standard dictator game and also the improvement when the additional contextual elements from the field are incorporated. In the next two subsections, we analyze our two field behaviors in turn.

#### 3.4.1. *Lab games and donations in the field*

To begin with, Table 4 reports Pearson correlations between the different game decisions and the donations in the field, 95% credible intervals, and Bayes factors.<sup>6</sup> Our analyses include elements of both traditional frequentist approaches and Bayesian inference, which is gaining ground in behavioral research (e.g., Lee and Wagenmakers, 2013; Morey et al., 2016). As we can see in the table, the standard dictator game (Game 1) shows a small to moderate correlation of 0.21 with the field donations. Across the other games, adding contextual elements to make them more similar to our first field situation dramatically increased the correlation, reaching a level of 0.65 in Game 4. By any standards, this is a remarkably high correlation between two behaviors (see, e.g., Mischel, 1968; Ross and Nisbett, 1991, for reviews of correlations of behaviors across situations and of correlations

<sup>5</sup>Games 1 and 2 were played in the same session in a randomized order, but the results were not significantly affected by the order (Mann–Whitney test comparing Game 1 when it was the first and when it was the second:  $U = 1,730, p = 0.32$ ; for Game 2:  $U = 1,811, p = 0.14$ ).

<sup>6</sup>Table A1 in the Appendix provides guidelines to interpret Bayes factors (Lee and Wagenmakers, 2013).



**Figure 3.** Distributions of decisions in the four lab games.

**Table 4.** Correlations between game decisions and field donations (Pearson), credible intervals, and Bayes factors.

|        | Pearson | Credible interval, 95% | BF <sub>10</sub> | N  |
|--------|---------|------------------------|------------------|----|
| Game 1 | 0.21**  | [0.00, 0.39]           | 0.93             | 94 |
| Game 2 | 0.18*   | [-0.03, 0.36]          | 0.53             | 94 |
| Game 3 | 0.58*** | [0.42, 0.70]           | >1,000           | 95 |
| Game 4 | 0.65*** | [0.51, 0.75]           | >1,000           | 90 |

Note: For all the tests, the alternative hypothesis ( $H_1$ ) is that the correlation is not equal to 0. BF<sub>10</sub> indicates the Bayes factor in favor of  $H_1$  over  $H_0$  (the null hypothesis). \*, \*\*, and \*\*\* stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

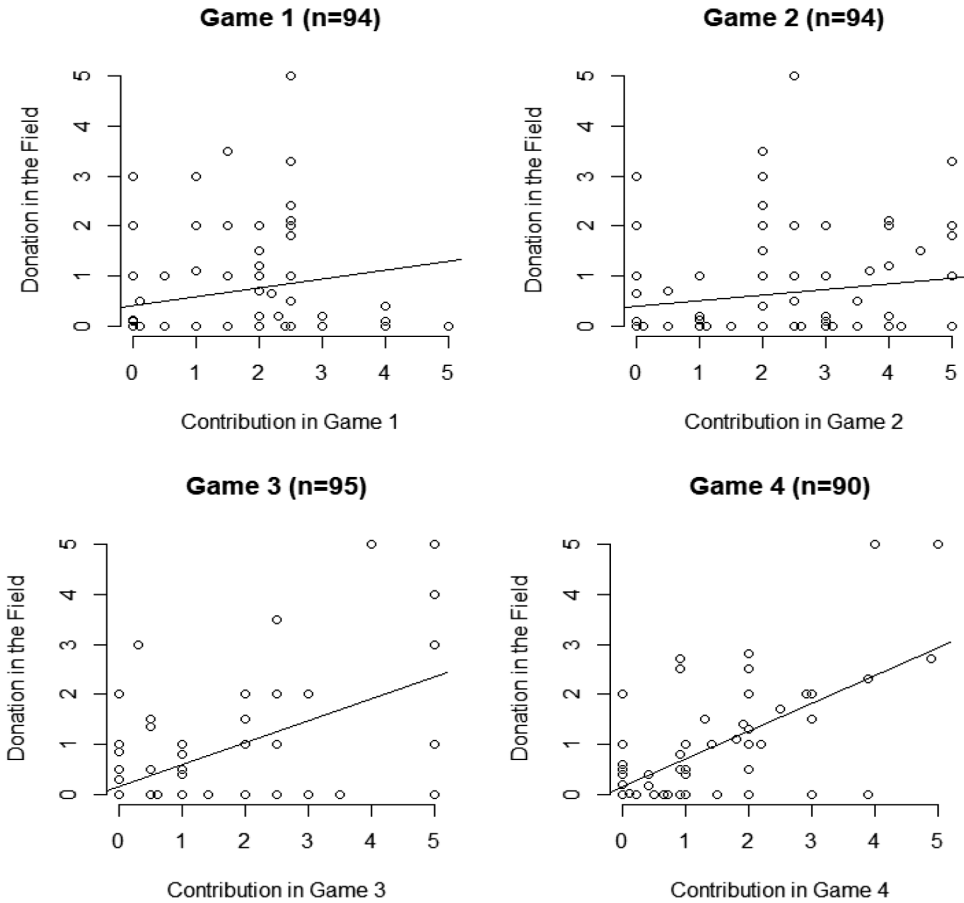
between different behavioral variables).<sup>7</sup> The 95% credible intervals and Bayes factors also confirm the increased predictive power of the game decisions. According to the guidelines to interpret Bayes factors (Table A1), there is no evidence that favors the alternative hypothesis of a nonzero correlation for the first two games, but there is extreme evidence that favors this hypothesis for the last two games.

Game 2, however, is an exception to this increasing trend. The correlation for Game 2 is slightly lower than, but not significantly different from, the one for Game 1 (Meng et al.'s, 1992, test comparing correlated correlations:  $z = 0.26$ ,  $p = 0.79$ ). As reflected in Figure 3, the distribution of contributions in Game 2 is the least similar to the field distributions (Figure 1). This is produced by the fact that, when we introduced Charity 4 as the recipient (instead of another participant), fewer people gave zero and more people gave higher amounts, resulting in a flatter distribution and higher average contributions. More specifically, participants who did not donate any money in the field donated on average €1.96 in Game 2. This pattern of donating more to charity in the lab than in the field is consistent with previous literature (e.g., Benz and Meier, 2008). Then, in Game 3, introducing the element of earned money pulled the distribution back toward zero and away from the higher contributions. Incorporating earned money in Game 3 produced the largest change in the lab–field correlation, with an increase of 0.40 (Fisher's  $z$  comparing independent correlations:  $z = -3.25$ ,  $p = 0.00$ ). Including face-to-face interaction in Game 4 produced a further increase of 0.07 in the lab–field correlation (Fisher's  $z$  comparing the correlations in Games 3 and 4:  $z = -0.75$ ,  $p = 0.45$ ).

Figure 4 provides a graphical demonstration of the relationship between decisions in the four games and donations in the field. The scatterplots confirm the correlations reported in Table 4. The association between contributions in the first two games and donations in the field is relatively weak. The lab–field association becomes much stronger as we move to Game 3, and it reaches the highest level in Game 4. In the last scatterplot, the points are much more concentrated around the tendency line and the number of outliers is small.

In Table 5, we present the results of an analysis based on standard and Bayesian Tobit regressions, which takes into account the censored nature of our donation data and provides an additional examination of the proportions of variance in the field behavior explained by the different games. We conducted eight regressions (four standard and four Bayesian) using always the field donations as the dependent variable and regressing them on each of the four different games. The standard and Bayesian Tobit regressions generated very similar coefficients. The pattern of these coefficients broadly replicates the one obtained in the correlation analysis. In terms of the proportions of variance explained ( $R^2$  in the standard models), we can see that Games 1 and 2 explain a very small proportion (0.04 and 0.03,

<sup>7</sup>We report here the simplest correlation analysis, but we also performed an analysis using nonparametric Kendall correlations (see Table A2 in the Appendix). The pattern obtained with Kendall correlations is broadly in line with the correlations discussed here.



**Figure 4.** Game decisions and donations in the field.

**Table 5.** Regression analysis: Donations in the field.

|                | Game 1 |                | Game 2 |                | Game 3  |                | Game 4  |                |
|----------------|--------|----------------|--------|----------------|---------|----------------|---------|----------------|
|                | Tobit  | Bayesian Tobit | Tobit  | Bayesian Tobit | Tobit   | Bayesian Tobit | Tobit   | Bayesian Tobit |
| Coefficient    | 0.49** | 0.51           | 0.25*  | 0.25           | 0.70*** | 0.72           | 0.81*** | 0.82           |
|                | (0.19) | (0.21)         | (0.14) | (0.15)         | (0.14)  | (0.14)         | (0.12)  | (0.13)         |
| R <sup>2</sup> | 0.04   |                | 0.03   |                | 0.34    |                | 0.43    |                |
| N              | 94     | 94             | 94     | 94             | 95      | 95             | 90      | 90             |

Note: Standard errors are reported in parentheses. ‘\*’, ‘\*\*\*’, and ‘\*\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively. Here, the R<sup>2</sup> refers to the multiple squared correlation, that is, the square of the correlation between the dependent variable and its predicted value.

respectively), whereas Games 3 and 4 explain a much larger proportion, reaching 0.43 in Game 4. In other words, Game 4 explains more than 10 times more variance than Game 1.<sup>8</sup>

<sup>8</sup>We ran additional analyses to see how well the games predict donations in the field after adjusting for the Big-Five, empathy, and demographic variables. Table A3 in the Appendix summarizes the results. The pattern obtained is broadly in line with the one discussed here.

**Table 6.** Correlations between game decisions and interest in volunteering (Pearson), credible intervals, and Bayes factors.

|        | Pearson | Credible interval, 95% | BF <sub>10</sub> | N   |
|--------|---------|------------------------|------------------|-----|
| Game 1 | 0.16    | [−0.04, 0.34]          | 0.45             | 102 |
| Game 2 | 0.14    | [−0.05, 0.33]          | 0.35             | 102 |
| Game 3 | −0.04   | [−0.23, 0.16]          | 0.13             | 101 |
| Game 4 | 0.03    | [−0.16, 0.23]          | 0.13             | 98  |

*Note:* For all the tests, the alternative hypothesis ( $H_1$ ) is that the correlation is not equal to 0. BF<sub>10</sub> indicates the Bayes factor in favor of  $H_1$  over  $H_0$  (the null hypothesis). ‘\*’, ‘\*\*’, and ‘\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

Overall, our analyses show that the limited predictive power of the standard dictator game when it comes to field behavior can be very substantially increased by introducing in the lab contextual elements from the field. In Game 4, we actually reached a level that made the game a very good predictor of the field behavior we wanted to address. In our setup, it seems that the earned money element was the factor that made most of the work of increasing the predictive power (from Game 2 to Game 3).

### 3.4.2. Lab games and interest in volunteering

Table 6 reports Pearson correlations between the game decisions and participants’ interest in volunteering scores, 95% credible intervals, and Bayes factors. The standard dictator game (Game 1) shows again a relatively small correlation (0.16) with this second field situation. This correlation is somewhat smaller than in the case of the first field situation. Game 2 presents a slightly lower, but very similar correlation (0.14). In this case, Games 3 and 4 show an even smaller correlation with interest in volunteering. Both correlations (−0.04 and 0.03, respectively) are very close to zero. None of these correlations are significant in our sample.<sup>9</sup> The credible intervals deliver the same message as the Pearson correlations. The first two Bayes factors (Games 1 and 2) provide anecdotal evidence for the null hypothesis that the correlation is 0 (see Table A1 for interpretation guidelines). The last two Bayes factors (Games 3 and 4) are smaller, implying moderate evidence for the null hypothesis.

In Table 7, we present an analysis based on standard and Bayesian ordered logistic regressions that takes into account the ordinal nature of our interest in volunteering data. As shown in Table 5, we ran eight separate regressions (four standard and four Bayesian), using always interest in volunteering scores as our dependent variable and regressing it on the four different games. The standard and Bayesian regressions produced very similar coefficients. The pattern obtained in these coefficients is broadly in line with the one obtained in the correlations, and the proportions of variance explained (McFadden’s pseudo- $R^2$  in the standard regressions) are all close to zero (0.01, 0.01, 0.00, and 0.00). In other words, none of the games explain any substantial variance in this field behavior.<sup>10</sup>

The results reported in this section show that, when contextual elements from a field situation are incorporated into lab games to increase their predictive power in relation to that particular situation, this does not increase predictive power in relation to field settings that have different characteristics. In fact, making the games more context-specific might even diminish their predictive power in relation to other contexts, as we somewhat see in Games 3 and 4. This is again entirely consistent with the idea that behavior is highly context-dependent.

<sup>9</sup>As with the first field situation, here we reported the simplest analysis, but we included Kendall correlations in Table A4 in the Appendix. They replicate the pattern shown in Table 6.

<sup>10</sup>We also examined how well the games predict interest in volunteering after adjusting for the Big-Five, empathy, and demographic variables. The results are summarized in Table A5 in the Appendix. They suggest a similar pattern to the one obtained in Table 7.



**Table 7.** Regression analysis: Interest in volunteering.

|                          | Game 1         |                | Game 2         |                | Game 3          |                 | Game 4         |                |
|--------------------------|----------------|----------------|----------------|----------------|-----------------|-----------------|----------------|----------------|
|                          | OL             | BOL            | OL             | BOL            | OL              | BOL             | OL             | BOL            |
| Coefficient              | 0.23<br>(0.17) | 0.22<br>(0.17) | 0.14<br>(0.13) | 0.14<br>(0.13) | -0.04<br>(0.12) | -0.04<br>(0.12) | 0.09<br>(0.14) | 0.09<br>(0.14) |
| McFadden's pseudo- $R^2$ | 0.01           |                | 0.01           |                | 0.00            |                 | 0.00           |                |
| $N$                      | 102            | 102            | 102            | 102            | 101             | 101             | 98             | 98             |

Note: OL and BOL refer to standard and Bayesian ordered logistic regressions. Standard errors are reported in parentheses. \*\*, \*\*\*, and \*\*\*\* stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

### 3.5. Relationship between the two field behaviors

In this subsection, we look into the relationship between our two field behaviors to analyze the extent to which behavior in one field situation is generalizable to the other one with different characteristics.

For this, we can use our whole sample, that is, participants in all three experiments, including also the control conditions. After excluding the missing values, we have a total of 366 participants with complete observations. The Pearson correlation between the two field behaviors is 0.02 (Kendall correlation: -0.02). Therefore, there is essentially no correlation between these field behaviors.

We also ran additional standard and Bayesian ordered logistic analyses, regressing the interest in volunteering scores on the donations in the field. Again, the standard and Bayesian regression models generated very similar coefficients. The standard coefficient is nonsignificant, and McFadden's pseudo- $R^2$  is 0.00. Therefore, virtually no proportion of variance in our second field situation is explained by behavior in the first one.

In line with the findings reported in the previous subsection, these results show that behavior in one particular situation (in the field in this case) does not translate into behavior in another setting with different characteristics. This highlights again the power of context in determining behavior.

### 3.6. The Big-Five and empathy

In this subsection, we analyze the predictive power of our psychometric measures (the Big-Five and empathy) in relation to our field situations and we compare it with the predictive power of our games. To begin with, Table 8 summarizes the pairwise Pearson correlations between the different psychological constructs and the two field situations, including also 95% credible intervals and Bayes factors.

As the table shows, the Big-Five and empathy measures have small correlations with behaviors in the two field situations. In the case of the first situation, three (out of nine) of the psychometric variables are significantly correlated with the field donations at the conventional 5% level. Out of the Big-Five, Conscientiousness and Agreeableness show significant correlations, negative in the case of Conscientiousness and positive in the case of Agreeableness. In addition, Neuroticism shows a positive correlation that is only significant at 10%. Among the empathy measures, only Empathic Concern shows a significant positive correlation at 5%. All these correlations are, however, fairly small, with a maximum absolute value of 0.13. In relation to interest in volunteering, two of the psychometric variables display correlations that are significant at the 5% level: Extraversion with a negative sign and Conscientiousness with a positive one. These correlations are again quite small, with a maximum absolute value of 0.15.

The Bayes factors also suggest that there is generally a very weak association between the psychometric measures and the field behaviors. The highest value is 4.17, corresponding to the correlation between Extraversion and the interest in volunteering scores. This is the only value providing moderate

**Table 8.** *Correlations between psychological constructs and field behaviors (Pearson), credible intervals, and Bayes factors.*

|          |                    | Donations in the field            | Interest in volunteering           |
|----------|--------------------|-----------------------------------|------------------------------------|
| Big-Five | Openness           | −0.07<br>([−0.17, 0.04], 0.14)    | −0.04<br>([−0.14, 0.06], 0.08)     |
|          | Conscientiousness  | −0.13**<br>([−0.23, −0.03], 1.75) | 0.11**<br>([0.01, 0.20], 0.62)     |
|          | Extraversion       | 0.02<br>([−0.08, 0.12], 0.07)     | −0.15***<br>([−0.24, −0.05], 4.17) |
|          | Agreeableness      | 0.13**<br>([0.03, 0.23], 1.66)    | 0.01<br>([−0.09, 0.11], 0.07)      |
|          | Neuroticism        | 0.09*<br>([−0.02, 0.19], 0.26)    | −0.04<br>([−0.14, 0.06], 0.09)     |
| Empathy  | Perspective Taking | 0.02<br>([−0.08, 0.12], 0.07)     | −0.08<br>([−0.18, 0.02], 0.21)     |
|          | Empathic Concern   | 0.11**<br>([0.01, 0.21], 0.59)    | 0.07<br>([−0.03, 0.17], 0.16)      |
|          | Fantasy            | −0.08<br>([−0.18, 0.03], 0.19)    | −0.06<br>([−0.16, 0.04], 0.14)     |
|          | Personal Distress  | 0.00<br>([−0.10, 0.11], 0.07)     | 0.02<br>([−0.08, 0.12], 0.07)      |
| <i>N</i> |                    | 366                               | 392                                |

*Note:* Credible intervals at the 95% level and Bayes factors (indicating the degree of support for the alternative hypothesis) are reported in parentheses. For all the analyses, the alternative hypothesis ( $H_1$ ) is that the correlation is not equal to 0. \*\*, \*\*\*, and \*\*\*\* stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

evidence for the alternative hypothesis that the correlation is not equal to 0. Two other values are larger than one (1.66 and 1.75), implying anecdotal evidence for the alternative hypothesis. All other values are smaller than one, providing some degree of evidence for the null hypothesis.

To explore further these relationships and the proportions of variance explained, we ran a more structured regression analysis, including different combinations of psychological variables. For the first field situation, we ran three separate standard Tobit regression models and three Bayesian ones, which are summarized in Table 9. In the first two regression models of the table (one standard and one Bayesian), we used the Big-Five personality traits as predictors; in the third and fourth models, we used the empathy measures; and in the last two, we included all the variables.

As Table 9 shows, the standard and Bayesian Tobit regressions again produced very similar coefficients. The regression coefficients are broadly in line with the correlations in Table 8, but with some differences. In terms of the Big-Five, Neuroticism and Conscientiousness are both significant at the 5% level across regressions, Neuroticism with a positive sign and Conscientiousness with a negative one. Among the empathy measures, Empathic Concern shows significant (and positive) coefficients at 5% across regressions, and Fantasy shows a significant (and negative) coefficient at 5% in the fifth model and at 10% in the third one.

More importantly, the proportions of variance explained are all very low, with  $R^2$ s of 0.04, 0.03, and 0.07 in the three standard regressions, respectively. All this suggests that the psychometric measures we considered have quite limited power to predict behavior in our first field situation. Compared to the results obtained with the games, the explanatory power of the psychological variables seems to be in the same ballpark as that of Games 1 and 2. In other words, the standard dictator game (Game 1) does not seem to do worse than psychometric personality measures, either general ones such as the Big-Five or more specific ones linked to social interaction such as the empathy measures of the Interpersonal

**Table 9.** Regression analysis: Field donations and psychometric measures.

|                    | Tobit 1           | Bayesian<br>Tobit 1 | Tobit 2          | Bayesian<br>Tobit 2 | Tobit 3           | Bayesian<br>Tobit 3 |
|--------------------|-------------------|---------------------|------------------|---------------------|-------------------|---------------------|
| Openness           | -0.04<br>(0.03)   | -0.04<br>(0.03)     |                  |                     | -0.03<br>(0.03)   | -0.03<br>(0.03)     |
| Conscientiousness  | -0.07**<br>(0.03) | -0.07<br>(0.03)     |                  |                     | -0.08**<br>(0.03) | -0.08<br>(0.04)     |
| Extraversion       | -0.04<br>(0.05)   | -0.04<br>(0.05)     |                  |                     | -0.05<br>(0.05)   | -0.05<br>(0.05)     |
| Agreeableness      | 0.05<br>(0.04)    | 0.05<br>(0.04)      |                  |                     | 0.05<br>(0.04)    | 0.05<br>(0.04)      |
| Neuroticism        | 0.10**<br>(0.04)  | 0.10<br>(0.05)      |                  |                     | 0.10**<br>(0.04)  | 0.10<br>(0.05)      |
| Perspective Taking |                   |                     | -0.02<br>(0.03)  | -0.02<br>(0.03)     | -0.03<br>(0.03)   | -0.03<br>(0.03)     |
| Empathic Concern   |                   |                     | 0.08**<br>(0.03) | 0.08<br>(0.03)      | 0.07**<br>(0.03)  | 0.08<br>(0.03)      |
| Fantasy            |                   |                     | -0.04*<br>(0.02) | -0.05<br>(0.02)     | -0.05**<br>(0.02) | -0.05<br>(0.02)     |
| Personal Distress  |                   |                     | -0.02<br>(0.02)  | -0.02<br>(0.02)     | 0.00<br>(0.03)    | 0.00<br>(0.03)      |
| $R^2$              | 0.04              |                     | 0.03             |                     | 0.07              |                     |
| $N$                | 366               | 366                 | 366              | 366                 | 366               | 366                 |

Note: Standard errors are reported in parentheses. Here, the  $R^2$  refers to the multiple squared correlation, that is, the square of the correlation between the dependent variable and its predicted value. ‘\*’, ‘\*\*’, and ‘\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

Reactivity Index. The predictive power of Games 3 and 4 is clearly and substantially superior to that of the personality measures, highlighting again the importance of including the right contextual elements to be able to explain behavior in specific situations.

Table 10 summarizes the results of using the same type of regression approach for our second field situation, in this case using standard and Bayesian ordered logistic regressions and interest in volunteering scores as the dependent variable. Again, the standard and Bayesian regressions produced very close coefficients. The regression results are broadly consistent with the correlations in Table 8, but with some differences. In the case of the Big-Five measures, Extraversion is significant (and negative) at the 5% level across regressions, Conscientiousness is significant (and positive) at the 5% level in the first model and at 10% in the fifth one, and Agreeableness is significant (and positive) at the 5% level in the first model. Among the empathy measures, Empathic Concern is significant (and positive) at 5% across regressions, Perspective Taking is significant (and negative) at 5% in the third model, and Fantasy is significant (and negative) at 10% in the third model.

The proportions of variance explained are very low across the interest in volunteering regressions (McFadden’s pseudo- $R^2$  of 0.02, 0.01, and 0.03 in the standard regressions). Also, for this field situation, the explanatory power of the psychological measures seems broadly similar to that of Games 1 and 2. In this case, however, it is important to recall that Games 3 and 4 did not produce an increase, but rather a slight decrease, in predictive power. This is one more time in line with the context-dependent nature of pro-social behavior, as explained in previous sections.<sup>11</sup>

<sup>11</sup>Table A6 in the Appendix contains an additional analysis, focused on the correlations between the psychometric measures and the game decisions.

**Table 10.** *Regression analysis: Interest in volunteering and psychometric measures.*

|                             | OL 1              | BOL 1           | OL 2              | BOL 2           | OL 3               | BOL 3           |
|-----------------------------|-------------------|-----------------|-------------------|-----------------|--------------------|-----------------|
| Openness                    | -0.01<br>(0.03)   | -0.01<br>(0.03) |                   |                 | -0.01<br>(0.03)    | -0.01<br>(0.03) |
| Conscientiousness           | 0.07**<br>(0.03)  | 0.07<br>(0.03)  |                   |                 | 0.06*<br>(0.03)    | 0.06<br>(0.03)  |
| Extraversion                | -0.11**<br>(0.04) | -0.11<br>(0.04) |                   |                 | -0.12***<br>(0.05) | -0.12<br>(0.05) |
| Agreeableness               | 0.07**<br>(0.04)  | 0.07<br>(0.04)  |                   |                 | 0.06<br>(0.04)     | 0.06<br>(0.04)  |
| Neuroticism                 | -0.06<br>(0.04)   | -0.06<br>(0.04) |                   |                 | -0.06<br>(0.04)    | -0.06<br>(0.04) |
| Perspective Taking          |                   |                 | -0.05**<br>(0.03) | -0.05<br>(0.03) | -0.04<br>(0.03)    | -0.04<br>(0.03) |
| Empathic Concern            |                   |                 | 0.07**<br>(0.03)  | 0.07<br>(0.03)  | 0.08**<br>(0.03)   | 0.08<br>(0.03)  |
| Fantasy                     |                   |                 | -0.04*<br>(0.02)  | -0.04<br>(0.02) | -0.03<br>(0.02)    | -0.03<br>(0.02) |
| Personal Distress           |                   |                 | -0.01<br>(0.02)   | -0.01<br>(0.02) | -0.01<br>(0.02)    | -0.01<br>(0.02) |
| McFadden's<br>pseudo- $R^2$ | 0.02              |                 | 0.01              |                 | 0.03               |                 |
| <i>N</i>                    | 392               | 392             | 392               | 392             | 392                | 392             |

Note: OL and BOL refer to standard and Bayesian ordered logistic regressions. Standard errors are reported in parentheses. \*\*, \*\*\*, and \*\*\*\* stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

#### 4. Discussion and conclusions

As Levitt and List (2007) pointed out, ‘perhaps the most fundamental question in experimental economics is whether findings from the lab are likely to provide reliable inferences outside of the laboratory’ (p. 170). The extent to which the current lab experiments on economic decisions can provide such inferences is still to be determined, but an increasing number of papers suggest that experiments based on social preference games have issues of external validity (see Galizzi and Navarro-Martinez, 2019, for a systematic review and meta-analysis).

In this paper, we have demonstrated (focusing on the dictator game) that even though so-called context-free lab games have limited power to predict specific field behaviors, this power can be very substantially increased by bringing relevant contextual elements from the field into the lab. Specifically, in the case of the game situation that we designed to be most similar to our target field situation, the lab–field correlation reached a remarkable 0.65. To the best of our knowledge, this is the first paper to show that an economic lab game can be modified to reach such a high correlation with a specific field behavior. At the same time, this game situation did not correlate well with a different type of field situation that it was not designed to address. Moreover, behaviors in our two different field situations did not correlate at all with each other.

All this suggests that social behavior is highly context-specific and cannot be properly explained without the contextual elements that are associated with it. As explained in the Introduction, this conclusion about the importance of context is in line with a large volume of literature in psychology and in behavioral and experimental economics showing that human behavior strongly depends on context.

A clear piece of advice seems to naturally follow from these conclusions: researchers who conduct economic experiments should seriously consider incorporating more contextual elements from outside

the lab that make their experimental settings more similar to the types of field environments they aim to address. This advice is similar to that of Alekseev et al. (2017) on the more specific aspect of experimental instructions. These authors suggest that context-rich language, as opposed to the typical abstract language used in economic experiments, can enhance the understanding of laboratory tasks and also affect behavior in a positive way by making it more related to the underlying research question; for example, in cases where the context of interest evokes particular emotions or social norms. More generally, our suggestion is also in agreement with well-known proposals to increase the representativeness of experimental designs in other areas of judgment and decision making (e.g., Dhimi et al., 2004; Hammond, 1990; Hogarth, 2005), and with the current interest in field experimentation in economics (Banerjee and Duflo, 2009; Duflo and Banerjee, 2017; Harrison and List, 2004; Levitt and List, 2009), which by its nature takes place in more contextualized environments. Ideally, there should also be a closer interaction between lab and field experimentation on economic decisions. Given that lab experiments typically allow for a greater degree of control (i.e., internal validity) and are easier and cheaper to implement, more contextualized lab experiments could be a valuable tool to understand what to expect in the field. Lab findings could then be investigated in the relevant field environments to make sure that the lab paradigms used tap well into the field behaviors they aim to understand.

A natural question to ask in relation to our advice is which are the relevant contextual elements from the field that researchers should consider incorporating into economic games to achieve a higher lab–field correlation. Unfortunately, this question is difficult to answer in a general and comprehensive way and it is beyond the scope of this paper, but there is already a large volume of research pointing to important contextual elements. In our own experiments, making participants earn the money they had to allocate (in combination with changing the recipient to a charity) seemed to be the most important factor in increasing the lab–field correlation. This is in line with previous research highlighting the importance of earned money in dictator games (e.g., Carlsson et al., 2013; Cherry et al., 2002; Cherry and Shogren, 2008; Fershtman et al., 2012; Korenok et al., 2017; List, 2007; Oxoby and Spraggon, 2008), and also in other lab games and tasks (e.g., Durham et al., 2014; Kroll et al., 2007; Muehlbacher and Kirchler, 2009). Other important examples of contextual elements that matter include the degree of anonymity (Hoffman et al., 1994, 1996; Winking and Mizer, 2013); the framing of the game (Burnham et al., 2000; Dufwenberg et al., 2011; Liberman et al., 2004; Zhong et al., 2007); or the available actions (Bardsley, 2008; List, 2007).

Apart from systematically studying and incorporating specific contextual factors, simply moving away from context-free games and trying to make laboratory environments similar to field situations of interest, that is, increasing the level of mundane realism (Aronson and Carlsmith, 1968), should maximize the chances of a good lab–field correspondence. This seems especially relevant in cases in which the experimenters establish explicit links between their research and particular types of behaviors outside the lab, which we believe is the case in most papers using economic experiments.

Some previous literature has also highlighted differences between laboratory tasks and real-world situations that go beyond specific contextual elements (e.g., Loewenstein, 1999; Levitt and List, 2007). An example of such a difference is implied social norms. When participants are making decisions in a lab environment, they infer from it the norms they think they are expected to follow. For instance, participants in an experiment related to social preferences may be under the impression that when they are asked about giving money in a laboratory environment, they are expected to behave more pro-socially. Another example of these differences is that decisions made in laboratory environments may seem more hypothetical than decisions in the field, even if there is real money at stake, and this in turn can lead to behavior that corresponds more to hypothetical decisions (see, e.g., Camerer and Hogarth, 1999; Harrison and Rutström, 2008; Murphy et al., 2005; Pronin et al., 2008, for differences between hypothetical and incentivized choices).

We believe that increasing mundane realism and using more contextualized laboratory tasks is also likely to reduce these differences between the lab and the field and increase external validity. The differences may never be eliminated completely, but creating environments that closer resemble their

real-world counterparts is bound to trigger more similar norms, lower degrees of hypothetical biases, and so forth.

A final question we want to address is whether incorporating contextual elements from the field into the lab is always the right thing to do. In our view, there are at least two important cases where this may not be the way to go. The first case is when experiments directly test context-free theory. If the goal of an experiment is just to conduct a test of a theory that is formulated in context-free terms, as is often the case with economic theory, the use of context-free experimental tasks that match the theory seems justified. The second case is when research questions are not linked to particular types of real-world behaviors. A well-known result in psychology is that broad psychological measures, such as personality traits, are bad predictors of particular behaviors but have a much higher predictive power in relation to measures constructed by aggregating multiple behaviors (Epstein, 1979; Epstein and O'Brien, 1985; Fleeson, 2001, 2004). The same has been shown to be true for context-free economic games (Wang and Navarro-Martinez, *forthcoming*). These games have limited predictive power in relation to specific pro-social behaviors in the field, but they can correlate fairly well with aggregated measures of behavior. Therefore, as we explained above, if an experiment is related to particular types of field behaviors, making the lab games more similar to the contexts in which those behaviors take place will increase external validity. However, if an experiment aims to capture a general trait or a general behavioral tendency across contexts, context-free games might be a good option. In this case, aggregating across multiple game rounds or different games is likely to help reduce measurement error and increase predictive power (Haesevoets et al., 2020; Wang and Navarro-Martinez, *forthcoming*). Therefore, there certainly is legitimate room for context-free games in economic experiments.

To conclude, while more research is needed on how incorporating contextual elements from field settings can improve the external validity of laboratory tasks, our research clearly points to the importance of seriously rethinking the role of context in social preference experiments. Using more naturalistic environments that closer resemble the world outside the lab might be one of the keys to avoid important external validity issues that could undermine the relevance of lab experiments in this area. We hope that our findings help to make room for more contextualized experiments and a closer and more systematic interaction between the lab and the field.

**Data availability statement.** Data, experimental instructions, and analysis code are available via the following link: [https://osf.io/du8vr/?view\\_only=5e3274054cd64cc8aeca64008ae8ed13](https://osf.io/du8vr/?view_only=5e3274054cd64cc8aeca64008ae8ed13).

**Acknowledgements.** The authors thank Verónica Benet-Martínez, Yan Chen, Gert Cornelissen, Matteo Galizzi, Josep Gisbert Rodríguez, Antonio Filippin, Tomáš Jagelka, Gaël Le Mens, John List, Rahil Hosseini, Elia Soler, and Jan Stoop for helpful comments and suggestions. They also express gratitude to Marta Araque, Xenia Dalmau, Analía García, and Pablo López for their assistance in conducting this study.

**Funding statement.** The authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (grant PID2019-105249GB-I00), the BBVA Foundation (grant Fundacion BBVA-El-2019-D.Navarro), and the Ramon Areces Foundation (grant Fundacion Ramon Areces 2019-Navarro).

**Competing interest.** The authors declare no competing interests.

## References

- Alekseev, A., Charness, G., & Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, 134, 48–59.
- Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *The Journal of Political Economy*, 97(6), 1447–1458.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401), 464–477.
- Andreoni, J., & Payne, A. (2013). Charitable giving. In Auerbach, A. J., Chetty, R., Feldstein, M., & Saez, E. (Eds.), *Handbook of public economics* (Vol. 5, pp. 1–50). New York: Elsevier.
- Andreoni, J., & Rao, J. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics*, 95, 513–520.

- Andreoni, J., Rao, J., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *The Journal of Political Economy*, *125*(3), 625–653.
- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73–106.
- Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*(1), 1–10.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In Lindzey, G., & Aronson, E. (Eds.), *The handbook of social psychology* (2nd ed., Vol. 2, pp. 1–79). Reading, MA: Addison-Wesley.
- Auten, G. E., Sieg, H., & Clotfelter, C. T. (2002). Charitable giving, income, and taxes: An analysis of panel data. *The American Economic Review*, *92*, 371–382.
- Banerjee, A., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics*, *1*, 151–178.
- Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*, 122–133.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26.
- Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973.
- Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field? Evidence from donations. *Experimental Economics*, *11*, 268–281.
- Berg, J., Dickhaut, J. W., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 166–193.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*(1–2), 52–69.
- Branas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior & Organization*, *60*(3), 306–320.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.
- Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, *43*(1), 57–73.
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Camerer, C. (2011). The promise and success of lab–field generalizability in experimental economics: A critical reply to Levitt and List. Working Paper. Pasadena, CA: California Institute of Technology.
- Camerer, C., & Thaler, R. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, *9*(2), 209–219.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor production framework. *Journal of Risk and Uncertainty*, *19*(1–3), 7–42.
- Carlsson, F., He, H., & Martinsson, P. (2013). Easy come, easy go. The role of windfall money in lab and field experiments. *Experimental Economics*, *16*, 190–207.
- Carpenter, J., Connolly, C., & Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics*, *11*, 282–298.
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.
- 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*, *9*, 88–97.
- Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, *92*(4), 1218–1221.
- Cherry, T. L., & Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters*, *101*(1), 69–72.
- 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*, *100*(2), 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., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, *127*(1), 1–56.
- Dhali, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988.
- Dickinson, D. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, *17*(4), 638–670.

- Dovidio, J. F., & Penner, L. A. (2001). Helping and altruism. In Brewer M., & Hewstone M. (Eds.), *Blackwell international handbook of social psychology: Interpersonal processes* (pp. 162–195). Cambridge, MA: Blackwell.
- Duflo, E., & Banerjee, A. (Eds.). (2017). *Handbook of field experiments* (Vol. 1). New York: Elsevier.
- Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2), 459–478.
- Durham, Y., Manly, T. S., & Ritsema, C. (2014). The effects of income source, context, and income level on tax compliance decisions in a dynamic experiment. *Journal of Economic Psychology*, 40, 220–233.
- Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, 16(2), 181–191.
- Ellingsen, T., & Johannesson, M. (2008). Anticipated verbal feedback induces altruistic behavior. *Evolution and Human Behavior*, 29(2), 100–105.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37(7), 1097.
- Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin*, 98(3), 513–537.
- Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2022). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science* (published online in Articles in Advance, 31 October 2022).
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4), 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Fershtman, C., Gneezy, U., & List, J. A. (2012). Equity aversion: Social norms and the desire to be ahead. *American Economic Journal: Microeconomics*, 4(4), 131–144.
- Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments. *The American Economic Review*, 100(1), 541–556.
- 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(6), 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.
- Fleeson, W., & Nofle, E. (2009). In favor of the synthetic resolution to the person–situation debate. *Journal of Research in Personality*, 43(2), 150–154.
- Forsythe, R., Horowitz, J. L., Savin N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347–369.
- Franzen, A., & Pointner S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, 81(1), 74–81.
- Galizzi, M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: A systematic lab–field study. *Management Science*, 65(3), 976–1002.
- Giluk, T. L., & Postlethwaite, B. E. (2015). Big five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences*, 72, 59–67.
- Giving USA. (2022). *Giving USA 2022: The annual report on philanthropy for the year 2021*. McLean, VA: Giving USA Foundation.
- Glazer, A., & Konrad, K. A. (1996). A signaling explanation for the charity. *The American Economic Review*, 86(4), 1019–1028.
- Glimcher, P. W., & Fehr, E. (Eds.) (2008). *Neuroeconomics: Decision making and the brain*. London: Elsevier.
- Guth W., Schmittberger R., & Schwarze B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.
- Haesevoets, T., Van Hiel, A., Dierckx, K., & Folmer, C. R. (2020). Do multiple-trial games better reflect prosocial behavior than single-trial games? *Judgement and Decision Making*, 15(3), 330–345.
- Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In Hammond, K. R. (Ed.), *The psychology of Egon Brunswik* (pp. 15–80). New York: Holt, Rinehart and Winston.
- Hammond, K. R. (1986). Generalization in operational contexts: What does it mean? Can it be done? *IEEE Transactions on Systems, Man, and Cybernetics*, 16, 428–433.
- Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved? In Hogarth, R. M. (Ed.), *Insights in decision making* (pp. 227–261). Chicago, IL: University of Chicago Press.
- Harrison, G., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42, 1009–1055.
- Harrison, G. W., & Rutström, E. (2008). Experimental evidence on the existence of hypothetical bias in value elicitation methods. In Plott, C. R., & Smith, V. L. (Eds.), *Handbook of experimental economics results* (Vol. 1, pp. 752–767). Amsterdam: Elsevier.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (Eds.) (2004). *Foundations of human sociality*. Oxford: Oxford University Press.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2), 73–78.



- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484.
- Hoffman, E., McCabe, K., Shachat, K., & Smith, V. L. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7, 346–380.
- Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, 86(3), 653–660.
- Hogarth, R. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, 12, 253–263.
- Hoolwerf, B., & Schuyt, T. (Eds.) (2017). *Giving in Europe. The state of research on giving in 20 European countries*. Amsterdam: Lente Publishers.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory—versions 4a and 54*. Berkeley, CA: Institute of Personality and Social Research, University of California, Berkeley.
- Kachurka, R., Krawczyk, M., & Rachubik, J. (2021). State lottery in the lab: An experiment in external validity. *Experimental Economics*, 24, 1242–1266.
- Kahneman, D., Knetsch, J., & Thaler, R. (1986). Fairness and the assumptions of economics. *The Journal of Business*, 59(4), S285–S300.
- Karlan, D., & List, J. A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *The American Economic Review*, 97(5), 1774–1793.
- Kolstad, J. R., & Lindkvist, I. (2012). Pro-social preferences and self-selection into the public health sector: Evidence from an economic experiment. *Health Policy and Planning*, 28(3), 320–327.
- Konow, J. (2010). Mixed feelings: Theories of and evidence on giving. *Journal of Public Economics*, 94(3–4), 279–297.
- Korenok, O., Millner, E., & Razzolini, L. (2017). Feelings of ownership in dictator games. *Journal of Economic Psychology*, 61, 145–151.
- Kroll, S., Cherry, T., & Shogren, J. (2007). The impact of endowment heterogeneity and origin on contributions in best-shot public good games. *Experimental Economics*, 10, 411–428.
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nature Neuroscience*, 11, 404–409.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press. [https://books.google.com/books?hl=en&lr=&id=Gq6kAgAAQBAJ&oi=fnd&pg=PR10&dq=info:LnJEgs-Ult4J:scholar.google.com&ots=tyJeECFqsr&sig=TFdHFNdf5F17SVzHZQl\\_TKRq6Ks](https://books.google.com/books?hl=en&lr=&id=Gq6kAgAAQBAJ&oi=fnd&pg=PR10&dq=info:LnJEgs-Ult4J:scholar.google.com&ots=tyJeECFqsr&sig=TFdHFNdf5F17SVzHZQl_TKRq6Ks)
- Levitt, S., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world. *Journal of Economic Perspectives*, 21(2), 153–174.
- Levitt, S., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1), 1–18.
- Lieberman, V., Samuels, S. M., & 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.
- Lichtenstein, S., & Slovic, P. (Eds.) (2006). *The construction of preference*. Cambridge: Cambridge University Press.
- List, J. A. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *The Journal of Political Economy*, 114(1), 1–37.
- List, J. A. (2007). On the interpretation of giving in dictator games. *The Journal of Political Economy*, 115(3), 482–493.
- List, J. A. (2011). The market for charitable giving. *Journal of Economic Perspectives*, 25(2), 157–180.
- Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economics Journal*, 109(453), 25–34.
- Mazar, N., & Zhong, C. (2010). Do green products make us better people? *Psychological Science*, 21(4), 494–498.
- Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111(1), 172–175.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin and Review*, 23, 103–123.
- Muehlbacher, S., & Kirchler, E. (2009). Origin of endowments in public good games: The impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics*, 2(1), 59–67.
- Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3), 313–325.
- Okten, C., & Weisbrod, B. (2000). Determinants of donations in private nonprofit markets. *Journal of Public Economics*, 75(2), 255–272.
- Olivola, C. Y., Kim, Y., Merzel, A., Kareev, Y., Avrahami, J., & Ritov, I. (2020). Cooperation and coordination across cultures and contexts: Individual, sociocultural, and contextual factors jointly influence decision making in the volunteer's dilemma game. *Journal of Behavioral Decision Making*, 33, 93–118.
- Oppenheimer, D., & Olivola, C. (2011). *The science of giving: Experimental approaches to the study of charity*. New York: Psychology Press.

- Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, *65*, 703–713.
- Paunonen, S. (2003). Big five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology*, *84*(2), 411–424.
- Penner, L. A. (2002). Dispositional and organizational influences on sustained volunteerism: An interactionist perspective. *Journal of Social Issues*, *58*(3), 447–467.
- Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. (2015). Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology*, *108*(6), 883–899.
- Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784.
- Piliavin, J. A., & Charng, H.-W. (1990). Altruism: A review of recent theory and research. *Annual Review of Sociology*, *16*, 27–65.
- Prinzin, P., Stams, G., Deković, M., Reijntjes, A., & Belsky, J. (2009). The relations between parents' big five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology*, *97*(2), 351–362.
- Proinin, E., Olivola, C. Y., & Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: Psychological distance and decision making. *Personality and Social Psychology Bulletin*, *34*(2), 224–236.
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.
- Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23–48.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
- Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*(5850), 598–602.
- Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*(5), 364–371.
- Smith, V. L. (1976). Experimental economics: Induced value. *The American Economic Review*, *66*(2), 274–279.
- Stewart, N., Reimers, S., & 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*(3), 687–705.
- Stoop, J. (2014). From the lab to the field: Envelopes, dictators and manners. *Experimental Economics*, *17*(2), 304–313.
- Stoop, J., Noussair, C. N., & Soest, D. (2012). From the lab to the field: Cooperation among fishermen. *The Journal of Political Economy*, *120*(6), 1027–1056.
- Wang, X., & Navarro-Martinez, D. (forthcoming). Increasing the external validity of social preference games by reducing measurement error. *Games and Economic Behavior*.
- Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, *34*(4), 288–293.
- Zhong, C., Bohns, V., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, *21*(3), 311–314.
- Zhong, C., Loewenstein, J., & Murnighan, J. K. (2007). Speaking the same language: The cooperative effects of labeling in the prisoner's dilemma. *Journal of Conflict Resolution*, *51*(3), 431–456.

## Appendix

**Table A1.** Guidelines for interpreting Bayes factors (Lee and Wagenmakers, 2013).

| BF <sub>10</sub> | Evidence                                |
|------------------|---|
| <1/100           | Extreme evidence for H <sub>0</sub>     |
| 1/100 to 1/30    | Very strong evidence for H <sub>0</sub> |
| 1/30 to 1/10     | Strong evidence for H <sub>0</sub>      |
| 1/10 to 1/3      | Moderate evidence for H <sub>0</sub>    |
| 1/3 to 1         | Anecdotal evidence for H <sub>0</sub>   |
| 1                | No evidence                             |
| 1 to 3           | Anecdotal evidence for H <sub>1</sub>   |
| 3 to 10          | Moderate evidence for H <sub>1</sub>    |
| 10 to 30         | Strong evidence for H <sub>1</sub>      |
| 30 to 100        | Very strong evidence for H <sub>1</sub> |
| >100             | Extreme evidence for H <sub>1</sub>     |

**Table A2.** Correlations between game decisions and field donations (Kendall).

|        | Pearson | Credible interval, 95% | BF <sub>10</sub> | N  |
|--------|---------|------------------------|------------------|----|
| Game 1 | 0.24*** | [0.09, 0.36]           | 35.07            | 94 |
| Game 2 | 0.15*   | [0.02, 0.28]           | 1.43             | 94 |
| Game 3 | 0.35*** | [0.21, 0.47]           | >1,000           | 95 |
| Game 4 | 0.46*** | [0.31, 0.58]           | >1,000           | 90 |

Note: For all the tests, the alternative hypothesis (H<sub>1</sub>) is that the correlation is not equal to 0. BF<sub>10</sub> indicates the Bayes factor in favor of H<sub>1</sub> over H<sub>0</sub> (the null hypothesis). ‘\*’, ‘\*\*\*’, and ‘\*\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table A3.** Regression analysis adjusting for other factors: Donations in the field.

|                | Game 1            |                | Game 2          |                | Game 3            |                | Game 4            |                |
|----------------|-------------------|----------------|-----------------|----------------|-------------------|----------------|-------------------|----------------|
|                | Tobit             | Bayesian Tobit | Tobit           | Bayesian Tobit | Tobit             | Bayesian Tobit | Tobit             | Bayesian Tobit |
| Coefficient    | 0.51***<br>(0.19) | 0.64<br>(0.27) | 0.31*<br>(0.16) | 0.40<br>(0.22) | 0.68***<br>(0.12) | 0.77<br>(0.17) | 0.73***<br>(0.12) | 0.80<br>(0.16) |
| R <sup>2</sup> | 0.15              |                | 0.13            |                | 0.42              |                | 0.49              |                |
| N              | 94                | 94             | 94              | 94             | 95                | 95             | 90                | 90             |

Note: In each regression, the amount donated in the field is regressed on decisions in one of the games, adjusting for the Big-Five, the four dimensions of empathy, and two basic demographic variables (age and gender). Standard errors are reported in parentheses. ‘\*’, ‘\*\*\*’, and ‘\*\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively. Here, the R<sup>2</sup> refers to the multiple squared correlation, that is, the square of the correlation between the dependent variable and its predicted value.

**Table A4.** Correlations between game decisions and interest in volunteering (Kendall).

|        | Pearson | Credible interval, 95% | BF <sub>10</sub> | N   |
|--------|---------|------------------------|------------------|-----|
| Game 1 | 0.13    | [0.00, 0.26]           | 0.90             | 102 |
| Game 2 | 0.12    | [−0.02, 0.24]          | 0.57             | 102 |
| Game 3 | 0.01    | [−0.12, 0.14]          | 0.13             | 101 |
| Game 4 | 0.01    | [−0.12, 0.15]          | 0.14             | 98  |

Note: For all the tests, the alternative hypothesis (H<sub>1</sub>) is that the correlation is not equal to 0. BF<sub>10</sub> indicates the Bayes factor in favor of H<sub>1</sub> over H<sub>0</sub> (the null hypothesis). ‘\*’, ‘\*\*\*’, and ‘\*\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table A5.** Regression analysis adjusting for other factors: Interest in volunteering.

|                                  | Game 1          |                | Game 2         |                | Game 3         |                | Game 4         |                |
|----------------------------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                                  | OL              | BOL            | OL             | BOL            | OL             | BOL            | OL             | BOL            |
| Coefficient                      | 0.32*<br>(0.19) | 0.32<br>(0.19) | 0.10<br>(0.16) | 0.11<br>(0.15) | 0.05<br>(0.14) | 0.04<br>(0.14) | 0.10<br>(0.16) | 0.09<br>(0.16) |
| McFadden’s pseudo-R <sup>2</sup> | 0.10            |                | 0.09           |                | 0.10           |                | 0.04           |                |
| N                                | 102             | 102            | 102            | 102            | 101            | 101            | 98             | 98             |

Note: In each regression, the interest in volunteering score is regressed on decisions in one of the games, adjusting for the Big-Five, the four dimensions of empathy, and two basic demographic variables (age and gender). OL and BOL refer to standard and Bayesian ordered logistic regressions. Standard errors are reported in parentheses. ‘\*’, ‘\*\*\*’, and ‘\*\*\*\*’ stand for statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table A6.** Correlations between psychological constructs and game decisions (Pearson), credible intervals, and Bayes factors.

|          |                   | Game 1                            | Game 2                              | Game 3                             | Game 4                            |
|----------|-------------------|-----------------------------------|-------------------------------------|------------------------------------|-----------------------------------|
| Big-Five | Openness          | 0.06<br>([−0.12, 0.24],<br>0.15)  | −0.02<br>([−0.20, 0.17],<br>0.12)   | 0.08<br>([−0.11, 0.26],<br>0.17)   | −0.03<br>([−0.22, 0.17],<br>0.13) |
|          | Conscientiousness | −0.06<br>([−0.24, 0.12],<br>0.15) | 0.08<br>([−0.10, 0.26],<br>0.17)    | −0.09<br>([−0.27, 0.10],<br>0.18)  | 0.02<br>([−0.18, 0.21],<br>0.13)  |
|          | Extraversion      | 0.02<br>([−0.16, 0.21],<br>0.12)  | 0.12<br>([−0.07, 0.30],<br>0.26)    | −0.02<br>([−0.21, 0.17],<br>0.12)  | 0.04<br>([−0.16, 0.24],<br>0.14)  |
|          | Agreeableness     | 0.17*<br>([−0.02, 0.34],<br>0.57) | 0.03<br>([−0.15, 0.22],<br>0.13)    | 0.05<br>([−0.14, 0.23],<br>0.13)   | 0.07<br>([−0.13, 0.26],<br>0.16)  |
|          | Neuroticism       | 0.14<br>([−0.05, 0.31],<br>0.33)  | 0.22**<br>([0.04, 0.39],<br>1.87)   | 0.05<br>([−0.14, 0.23],<br>0.14)   | 0.11<br>([−0.09, 0.30],<br>0.23)  |
|          | Empathy           | Perspective Taking                | 0.14<br>([−0.04, 0.37],<br>0.37)    | 0.23**<br>([0.04, 0.39],<br>2.09)  | 0.20**<br>([0.02, 0.37],<br>1.15) |
|          | Empathic Concern  | 0.05<br>([−0.13, 0.23],<br>0.14)  | 0.29***<br>([0.11, 0.45],<br>13.83) | 0.14<br>([−0.05, 0.31],<br>0.33)   | −0.00<br>([−0.20, 0.20],<br>0.13) |
|          | Fantasy           | 0.15<br>([−0.04, 0.32],<br>0.38)  | 0.16*<br>([−0.03, 0.33],<br>0.49)   | 0.26***<br>([0.08, 0.42],<br>5.09) | −0.04<br>([−0.24, 0.15],<br>0.14) |
|          | Personal Distress | −0.15<br>([−0.32, 0.04],<br>0.39) | 0.03<br>([−0.16, 0.21],<br>0.12)    | −0.07<br>([−0.25, 0.12],<br>0.16)  | 0.04<br>([−0.16, 0.24],<br>0.14)  |
|          | <i>N</i>          | 112                               | 112                                 | 110                                | 98                                |

*Note:* Credible intervals at the 95% level and Bayes factors (indicating the degree of support for the alternative hypothesis) are reported in parentheses. For all the analyses, the alternative hypothesis ( $H_1$ ) is that the correlation is not equal to 0. \*\*, \*\*\*, and \*\*\*\* stand for statistical significance at the 10%, 5%, and 1% levels, respectively.