

External Validity of Experimental Social Preference Games

Sarah Xu

April 2018

1 Introduction

The standard economic model assumes individuals' actions are motivated purely by self-interest. However, from simple observation it is clear that people actually care about the well-being of others: people volunteer, people donate to causes, and social programs exist. If someone is not solely motivated by material self-interest but also cares about the material payoffs of others, we say that the person has social preferences. The last few decades have seen a strong increase of interest in this topic, with numerous research studies providing evidence of social preferences. Many papers have identified various aspects of social preferences such as altruism, social welfare, inequality aversion, and reciprocity (e.g., Charness and Rabin 2000; Fehr and Schmidt 1999; Andreoni and Miller 2002; Rabin 1993; Fisman, Jakiela, and Kariv 2014).

When studying social preferences, the typical approach is to conduct experiments in a laboratory setting. Participants play experimental games, where they receive monetary incentives aligned with the payoffs of the games. Researchers are able to control and influence prices, information, and actions available to the

participants, which allows researchers to rigorously target different aspects of social behavior. Below are common games used to study social preferences, along with their typical social preference interpretations:

- Dictator Game: The player (dictator) receives an endowment, m . The dictator chooses how to split the amount between themselves and a second player (the receiver) so that $\pi_s + \pi_o = m$, where π_s is how much is kept and π_o is how much is given.

The dictator's action is usually explained by concepts of altruism or fairness preferences (e.g. inequality aversion).

- Ultimatum Game: Two-player game where two players bargain over a fixed endowment, m . The first mover (proposer) divides the sum, so that $\pi_s + \pi_o = m$. The second mover (responder) chooses to either accept or reject the offer. If accepted, the proposal is implemented. If rejected, neither player receives any money.

The proposer's behavior is typically corresponded with fairness, and the responder's response can be explained by negative reciprocity or fairness preferences such as inequality aversion.

- Trust Game: Two-player game where the first player receives an endowment, m , and proposes how to divide the endowment between themselves and the second mover so that $\pi_s + \pi_o = m$. The second player (the responder) receives $k\pi_o$, where $k > 1$, and decides how much of the endowment to return to the first mover.

The proposer's action indicates their trust levels, and the responder's return indicates their trustworthiness or reciprocity.

- Public Goods Game: N-player game where each player receives the same

sum of money, m , and simultaneously decides how much to put into a public fund. The payoff function is given by $P_i = m - g_i + \beta \sum_{n=1}^N g_j$, where g_i is the amount that player i donated to the fund, β is the marginal payoff of the public fund ($0 < \beta < N\beta$), and $\sum_{n=1}^N g_j$ is the sum of the individual donations to the public fund.

The players' actions are usually explained by their altruism, cooperation, or trust levels.

There is an abundance of literature that study social preferences using the mentioned experimental games (along with other papers using games not mentioned). For example, Forsythe (1994) recruited undergraduate students from the University of Iowa, where they played single dictator and ultimatum games against anonymous opponents. The authors found that, contrary to the subgame perfect Nash equilibrium that both proposers and responders would not send any money, most players in fact gave away proportions of the endowment given to them. Bert et al (1995) recruited undergraduate students from the University of Minnesota to participate in trust games. 94% of the proposers sent money, and over one-third of the responders returned an amount greater than what was given to them, showing evidence of trust and reciprocity. Hermann, Thoni, and Gächter (2008) conducted public goods experiments across 16 participant pools, and found that that opportunities for punishment led to varying cooperation levels. used public goods game to provide evidence of punishment and reciprocity. These papers are only a few examples from an abundance of literature (with many papers using social preference games not mentioned above), making it clear that these games has become one of the building blocks of experimental and behavioral economics.

However, laboratory experiments have a disadvantage: they are abstract and remote from realistic situations. Therefore, an important question is the extent to which the experimental games approach to social preferences can be generalized

to real world situations. Human behavior may be influenced by a variety of factors that differ between practical and laboratory settings. One issue is that in a laboratory setting, the subject's actions are under the scrutiny of the researcher. Hoffman et al (1994) found that almost 50% of their subjects donated at least \$3 (out of a \$10 pie) when playing the dictator game. However, when the authors implemented a "double-blind" treatment where both the experimenter and other subjects could not observe the dictator's actions, they found that only 16% of subjects gave at least \$3. There are other papers that also find that subjects are more prosocial in the lab than they are in the field (e.g., List, 2006; Gneezy et al., 2004). Individuals are also influenced heavily on the context of the situation. For example, in a two-person trust game, Burnham, McCabe, and Smith (2000) switched between calling the responder "partner" and "opponent", and found a significant difference in trust and trustworthy behavior. Similarly, Ross and Ward (1996) found that simply calling a prisoner's dilemma game either a "community" or "Wall Street" game wildly varied each treatment group's rates of defection. There are also other studies that find that slightly changing the narrative could result in significant changes in behavior (e.g., Andreoni 1995; Gintis 2001; Henrich et al 2005). Studies also find evidence that varying the level of stakes may lead to significantly different behaviors. For example, Carpenter, Verhoogen, and Burks (2005) find that increasing the stakes from \$10 to \$100 decreased the median offer in the dictator game from 40% to 20%. Slonim and Roth (1998) and Parco, Rapoport, and Stein (2002) also find that higher stakes leads to significantly different results than with lower stakes. Interestingly, however, Cherry, Frykblom, and Shogren (2002) find no differences in offers when increasing the stakes from \$10 to \$40. Scrutiny, context, and stakes are only a few factors that can lead to deviations in behavior; Levitt and List (2007) discuss other factors – with supporting literature – that can lead to deviations in behavior between laboratory

and field settings.

The fact that varying factors such as scrutiny, context, and stakes can yield significant differences in behavior calls into question the external validity of using experimental games in a laboratory setting to study social preferences. There are a few studies that have examined how subjects' behavior in a lab setting is compared to the same subjects' behavior in a real world situation, and found that in-lab behavior does predict field behavior. Baran et al (2010) compared MBA alumni donations to their university with their reciprocity behavior when playing the trust game. The authors find that responder behavior predicted university donations. Franzen and Pointner (2013) compared behaviors from university students participating in standard dictator games to their actions when receiving a misdirected letter containing money, and found that subjects who showed prosocial behavior in the lab returned the misdirected letters more often than subjects who were selfish in the lab. Englmaier and Gebhardt (2010) conducted field experiments where they compared university students' free riding behavior at the library with free riding behavior in a public goods game, and found statistically significant correlation between field and lab measures. There are also a couple studies that used non-student subjects: Fehr and Leibbrandt (2011) conducted public goods games with Brazilian fishermen, and found that fishermen who are more cooperative in the games were also less likely to exploit the communal fishing grounds. Karlan (2005) compared Peru inhabitants' trust game behavior to their micro loans repayment behavior, and found correlation between behaviors in the lab game and in the field.

On the other hand, there are a number of papers that find lab behavior has no predictive power for field behavior. Goeschl et al (2015) examined university students' behavior in two different tasks: a public goods game and their contributions to a task about reducing CO2 emissions. The authors find that behavior

in both tasks is uncorrelated. Hill and Gurven (2004) carried out ultimatum and public goods games on Paraguay Ache Indians. They compared the behaviors to observed food production and sharing patterns to individuals outside the nuclear family and found no significant relationships to lab behavior. Gurven and Winking (2008) recruited Tsimane forager-horticulturalists in Bolivia, and compared behavior when playing dictator and ultimatum games to their food-sharing behavior. The authors find no relation between the two measures. Voors et al (2012) studied farmers in Sierra Leone, and compared their behavior in a public goods game to their behavior when asked to contribute to a real community public good. They find no meaningful correlation in behavior between the lab and field measures.

With other studies finding statistically significant, not statistically significant, and mixed results, the currently accumulated evidence is ambiguous. Galizzi and Navarro-Martinez (2017) conducted a systematic review and meta-analysis of literature on within-subjects lab-field studies: about 40% of reported lab-field correlations and regressions found statistically significant association between lab and field behaviors. It is clear that evidence for the external validity of experimental games is weak, justifying that further research is needed. Galizzi and Navarro-Martinez (2017) note, however, that the previous studies compare only one social preference game to one specific field measure. The abstract and context-free nature of the games makes it difficult to theoretically map the game to the field measure, and so it is crucial to have a more systematic approach. The authors proceed to carry out their own study where participants answered questions about social behaviors exhibited in the past, played various social preference games, and encountered naturalistic field situations. The field situations included a research assistant asking for help carrying boxes down the stairs, asking to use the participant's phone to make a brief phone call, asking for donations to a children's

charity, asking for donations to an environmental charity, or asking for donations to the lab's research fund. The authors' overarching conclusion is that behavior when playing the experimental games does a poor job predicting both the survey questions and the field behaviors. They do mention, however, that more systematic studies are needed to draw a definite conclusion.

It is critical to think about why some studies found correlations between lab and field measures and some studies did not, in order to inform the design of my current experiment. An underlying theme for those experiments that found no correlations is higher stakes. Voors et al (2012) note that an important difference between the lab and field experiment are that the stakes are much higher in the field experiment: a month's wage versus a day's wage, and the public good affected the entire village rather than two other players. In Gurven and Winking (2008)'s study, there are different stakes between the experimental and field measures. The authors note that "money is fungible and easier to hide than agricultural produce or meat and is shared differently than other resources". In addition, stakes are especially higher for food-sharing behavior since food-sharing is heavily ingrained in the culture. Lastly, while playing the experimental games for Hill and Gurven (2004)'s study, participants expressed worry that their choices would make the receiver upset. In particular, the tribal group's culture was heavily focused on community, and was well known for extensive food sharing (which was the field measure the authors used). Perhaps the participants didn't want to risk their choices creating any tension in their community, and getting punished with less food sharing and cooperation. For the last two papers, the stakes were a result of high scrutiny and low anonymity since it was easy for community members to find out the choices each subject made. For our design, an important aspect is that participants will play the games in their own time and location to ensure no scrutiny and high anonymity.

Adopting Galizzi and Navarro-Martinez (2017)’s design, I will recruit Wesleyan University seniors and recent alumni to play various experimental games. The different social preference games include the generalized dictator game, ultimatum game, trust game, and public goods game. I will further describe each game, as well as their relevance to my field measure, in the next section.

A criticism of Galizzi and Navarro-Martinez’s design is how the authors executed their field measures: since the setups occur as the participants are exiting their lab session, it is hard to believe that the participants did not connect the encounters to the research lab, especially since they previously answered questions similar to the current event. This may have led to actions that participants would not have done if they were unaware of the scrutiny. Therefore it is better to use a natural, far-removed situation. For my study, I will use Wesleyan University donations as the field measure. Further discussion on my choice of field measure is in the next section.

Similar to previous lab-field experiments, I will examine whether in-lab social behavior predicts real-life social behavior. In addition, along with playing various experimental games, participants will answer non-incentivized questions regarding their past social behavior. I will use the self-reported measures as an additional layer to evaluate the explanatory ability of the games. By comparing my results with those of past lab-field studies, I can contribute to the discussion on whether experimental games are a useful tool to examine social preferences

BRIEF OVERVIEW OF RESULTS .. IS IT IN LINE WITH GALIZZI/NAVARRO PAPER? THEY FOUND POOR JOB.

The rest of the paper is organized as follows: Section 2 describes the methods used; Section 3 presents the results obtained; and Section 4 discusses the results.

2 Methods

Each participant was presented with two sets of tasks: (i) incentivized social preference games, and (ii) survey questions regarding past social behavior. The field measure used to compare to participants' lab results is their Wesleyan University donations. The goal of this design is to evaluate the external validity of experimental social preference games to behaviors in the field. I will also examine whether another tool (self-reported measures) or both tools can better predict social preferences exhibited in real-world situations.

Wesleyan University seniors and recent alumni received an invitation email briefly explaining my research and asking for participation in my study (emails were provided by Wesleyan University Relations). The entire study was computerized, programmed and implemented using Qualtrics. Online links to both the experimental games and survey questions were included in the email, so those who chose to take part could complete the study remotely. As mentioned in the previous section, the online structure allows participants to make decisions anonymously and without scrutiny, minimizing the risk for exaggerated behavior. Participants received an overview of the two tasks they would complete, and were informed that upon completion of the study, their major, class year, and Wesleyan donations information would be released to me.

In order to incentivize completion of both the games and survey questions, as well as elicit honest game behavior, participants were also informed that completing the entire study made them eligible for a lottery prize. All games used "tickets" as the experimental currency unit: the total tickets each person earns directly corresponds with how many lottery tickets they own.

I will now explain each of the three main components.

2.1 Incentivized social preference games

Those who chose to participate in the study received a unique identifier assigned to them by Qualtrics. Participants first played four social preference games: generalized dictator game, ultimatum game, trust game, and public goods game. Before each game, they received detailed instructions along with an example to illustrate how each game worked. For the ultimatum and trust games, each participant will have the opportunity to play the roles of both proposer and responder.

Below are the games each participant played:

- Generalized Dictator Game (GDG): Each participant plays as the dictator (Player 1). They are asked to make a series of choices on how to divide tickets between themselves and an anonymous, random Player 2. Every ticket that Player 1 earns will be worth 1, 2, 3 or 4 tickets, depending on the choice. Similarly for the earnings of Player 2. Based off of Andreoni and Miller (2002)'s design, there will be nine rounds with varying prices and endowments.
- Ultimatum Game:
 - Player 1 (UG1): Each participant plays as the proposer (Player 1), and is endowed with 10 tickets. They decide how much of their endowment to send to an anonymous, random responder (Player 2). They are told that the responder may or may not reject the allocation. If the allocation is rejected, neither player receives any tickets in this round.
 - Player 2 (UG2): Each participant is now the responder. They are given a list from 0 tickets to 10 tickets (in 1-ticket increments), and are asked whether they accept or reject for each listed amount.
- Trust Game:

- Player 1 (TG1): Each participant plays as the proposer, and is endowed with 10 tickets. They decide how much of their endowment to send to an anonymous, random responder. The amount sent over is multiplied by three. The responder then decides how many tickets to return.
- Player 2 (TG2): Now each participant is the responder. They receive a list of all ten possible multiplied amounts that the proposer could have chosen to send. For each amount, they are prompted to enter the number of tickets they would like to return back to the proposer.
- Public Goods Game (PGG): Each participant is endowed with 10 tickets and decide how much of their endowment to contribute to a group fund with one other anonymous, random participant. The total tickets in the group fund is multiplied by 2 and divided evenly between the two players.

The games were played in random order, and images of each game, along with their instructions and examples, are in [Figure 1](#).

At the end of the participation deadline (February 2018), all participants were randomly paired, and ticket payoffs were calculated. I will discuss the strategy method in Section 2.1.1, and explain the payment process in further detail in Section 2.1.2.

2.1.1 Strategy Method

Since the participants are completing the games at their own availability, they were not randomly matched with other participants until after all data had been collected. As seen with the ultimatum and trust games, which required both a proposer and responder, the responder saw a list of all possible offer amounts and was prompted for their choice for each offer. This is called the strategy method.

There are several reasons why the strategy method is advantageous. It is useful

to have all of the responder’s potential choices so that payoffs can be determined. Most importantly, the strategy method gives more information. If players were matched with other players live, then we would only have the responder’s response given the proposer’s offer. The strategy method, however, provides all returned amounts for all possible donated amounts in the ultimatum game, and provides participants’ minimum accepted amount.

2.1.2 Payment

Participants were randomly paired, and ticket payoffs were determined for each game. In each assigned pair, one participant was randomly selected to be Player 1 (i.e. they were the dictator in the generalized dictator game, and proposer in both the ultimatum game and trust game), and the other was Player 2 (i.e. they were the receiver in the generalized dictator game, and the responder in both the ultimatum and trust game). I ran through each game and calculated ticket payoffs for each participant. Since the generalized dictator game consisted of nine rounds, one round was randomly selected. and each participant received the tickets based on their decision for that round. However many total tickets each participant earned from all the games corresponded with the number of tickets they owned in the lottery. Ten tickets were drawn, and each winner received \$100. [Once get more info from Linda Mascaro...](#)

2.2 Self-reported measures of past social behaviors

Participants then reported on social behaviors exhibited in the past. The questions are adapted from the Self-Report Altruism (SRA) scale introduced by Rushton, Chrisjohn, and Fekken (1981). The survey is comprised of 10 items, and participants report how frequently they have done each item in the past. Participants rate each statement as either “Never”, “Once”, “More than once”, “Often”, or

“Cery often”. Examples include: “I have donated money at the cash register when buying groceries”, “I have pointed out a clerk’s error (at the supermarket, at a restaurant) in undercharging me”, and “I have donated instead of sold my clothes/used items”. A full list of the 10 items is displayed in Table 1.

2.3 Field Measure

After the participation deadline passed, Wesleyan University Relations provided a dataset that contained each participant’s Wesleyan donations information.

Seniors are solicited around once per semester, during senior events. Recent alumni are typically solicited for donations at the end of November / beginning of December, March, and June. Since my online study was released between solicitation cycles for both seniors and recent alumni, and donations information was provided as of January 2018, donations are independent of my study.

There are several reasons why I chose to use donations as my field measure. First, many behavioral economics papers exploring the relationship between in-lab and field behavior have used donations (e.g. Falk, Meier, and Zehnder 2011; Benz and Meier 2008; Baran et al 2010). Falk, Meier, and Zehnder (2011) say they chose to use donations because “First, the measure does not rely on self-reported survey responses but on actual decisions. Second, donation decisions are made in private and never made public. Third, students are unaware that their behavior is analyzed in a research study. Fourth, and most importantly, all students at the university have to decide about the donations.”

Another reason is that donations may represent various social preference constructs. The most common representation is altruism, but donating can represent other concepts such as reciprocity, cooperation, and trust. In the context of using Wesleyan University donations, donating may represent reciprocity and trustworthiness since seniors/recent alumni may donate as thanks for providing them four

years of education and opportunities. Donating may also represent trust since donors are trusting that the University will putting their donations to good use, and cooperation because the University solicits for donations. Finally, donating may symbolize fairness preferences, more specifically inequality aversion: donors are able to target what area their donations can go towards. If the donor targets their donations to Financial Aid, then they are displaying inequality averse preferences.

Finally, donations is a more accurate field measure than creating a scenario at the end of the online session. Wesleyan donations is a far-removed situation, and is not directly related to the study. I previously mentioned that Galizzi and Navarro-Martinez (2017)’s design is unsatisfactory since their subjects encountered the field experiment shortly after completing the lab experiments, and reasoned that it would not be difficult for participants to link the two situations together. If I were to create a field situation at the end of the lab session, participants may feel compelled to act more pro-socially than they normally are since they know their actions are being scrutinized. In my design, participants’ donations behavior is not affected by their participation in the lab.

2.3.1 Theoretical Correspondences Between Games and Donations

All the games I chose for my study tap into different types of pro-social behaviors related to donating (mentioned in Section 2.3). First, the GDG and PGG responses are typically interpreted as altruism preferences. GDG, UG1, UG2, and PGG actions typically indicate fairness preferences. Lastly, TG1 represents participants’ trust levels, and TG2 represents participants’ trustworthiness and reciprocity. Therefore, each game can be theoretically mapped to donations behavior.

2.4 Participants and Sessions

Wesleyan University Relations provided a random sample of 2004 emails (334 emails for each class year from 2013 to 2018), and I sent an invitation email in early January 2018 asking for participation in my study. Participants were informed that the study consisted of several experimental games and non-incentivized survey questions, and participation should take no more than 15 minutes. The deadline for participation was mid-February 2018. A total of 397 people completed the online study. The participants were volunteers who opened the Qualtrics link, provided consent for me to receive their major, class year, and Wesleyan donations data, and completed both the survey questions and social preference games. At the end of the participation deadline, all participants were randomly paired, ticket payoffs were calculated, and 10 winners were randomly selected.

Winners were reached out WHEN

3 Results

The results are presented in four distinct sections. I first start by briefly describing the results obtained in the three main elements (social preference games, self-report measure of past social behaviors, and donations). The fourth section will focus on the main research question of the paper: the extent to which the games explain the field behavior.

3.1 Social Preference Games

Since the GDG consists of 9 rounds with varying sets of endowments and prices of giving, like Andreoni and Miller (2002) and Fisman, Kariv, and Markovits (2007), I assume each participant's giving preferences is a member of the constant elasticity of substitution (CES) utility function, written as:

$$U_s = [\alpha(\pi_s)^\rho + (1 - \alpha)(\pi_o)^\rho]^{1/\rho}$$

where α measures the relative weight on the payoff for self, ρ indicates the willingness to trade off payoffs to self and other in response to price changes, and $\sigma = 1/(\rho - 1)$ is the elasticity of substitution. Maximizing utility subject to the budget constraint ($p_s\pi_s + p_o\pi_o = m$) yields the CES demand function given by:

$$\begin{aligned}\pi_s(p, m) &= \frac{[\alpha/(1-\alpha)]^{1/(1-\rho)}}{\rho^{-\rho/(\rho-1)} + [\alpha/(1-\alpha)]^{1/(1-\rho)}} m \\ &= \frac{A}{p^r + A} m\end{aligned}$$

where $r = -\rho/(1 - \rho)$ and $A = [\alpha/(1 - \alpha)]^{1/(1-\rho)}$. This generates the following individual-level econometric specification for each participant:

$$\frac{p_{s,i}^t}{m_i^t} = \frac{A_i}{(p_i^t)^{r_i} + A_i} + \epsilon_i^t$$

where i represents each participant, t represents each of the 9 independent decision-problems participant i played in the GDG, and ϵ_i^t is assumed to be distributed normally with mean zero and variance σ_i^2 . The estimates \hat{A}_i and \hat{r}_i are generated using non-linear least squares. With the estimates, I am able to retrieve each participant's sharing propensity, α_i , and efficiency levels, ρ_i .

Results for TG1, UG1, and PGG are represented by the proposer's pass rates (percentage of the initial endowment passed to the other player or public pool). UG2 is represented by the minimum pass rate that the responder chooses to accept.

TG2 asks for the participant's return amount given 10 different offer amounts.

For each participant, I regressed their return amount on the offer amount, and retrieved the estimated slope. The estimated slope represents the participant's reciprocity, or trustworthiness level. Values closer to 1 represent more reciprocal behavior, and lower levels closer to 0 represent more selfish behavior. There may be concern that some participant's may not follow a linear trend. After plotting each participant's return amount in response to the offer amount, I found that with the exception of a few participants, each participant's data points follow a linear slope.

Figure 2a consists of 2 panels, Panels A and B, that shows the distribution of parameters, α and ρ , derived from responses in GDG. The parameter estimates vary dramatically across subjects, implying that preferences for giving are very heterogeneous. Panel A displays a high peak of 21% at $\alpha=1$, which shows that a considerable amount of participants displayed extremely selfish preferences. There is a smaller peak of 12% at $\alpha=0.5$ with more α levels above than below 0.5. To facilitate presentation of Panel B, participants with very negative ρ values were combined into the leftmost bar. About 7% of subjects have perfect substitutes preferences for giving ($\rho \approx 1$), which means the subjects prefer to give their entire endowment to Player 2 when the price of giving is less than one, and keep their entire endowment when the price of giving is greater than one. Roughly 15% of subjects possess Cobb-Douglas preferences ($\rho \approx 0$), and more than 20% of participants demonstrated Leontief preferences (ρ far below 0), i.e. they prefer splitting the endowment equally. Many subjects also have intermediate values of ρ : 24% have preferences for increasing total payoffs ($0.1 \leq \rho \leq 0.9$), and almost 20% have preferences for reducing differences in payoffs between self and others ($-0.9 \leq \rho \leq 0.9$).

Figure 2b consists of 4 panels (Panels C, D, E, and F) which shows the distribution of responses in UG1, UG2, TG1, and PGG, respectively. Panel C reports a

high 42% spike at giving half of their endowment to Player 2, and there is slightly more emphasis on giving contributions lower than half the endowment than contributions higher than half the endowment. These results are in line with the typical patterns in previous literature, which finds that a majority of offers are in the range of 25-50%. Correspondingly, Panel D shows the distribution of the minimum offers that Player 2 chooses to accept in UG2. With the exception of 1 participant who showed extreme negative reciprocity by only accepting an offer of Player 1's entire endowment, everyone accepted an amount less than or equal to half of Player 1's endowment. Surprisingly, most participants (27%) accepted an offer amount of 0 tickets. Panel E shows contributions scattered all across the range from offering 0% to 100% of their given endowment in TG1 with two spikes, both around 22%, at giving half the endowment or the entire endowment. Other contributions that have more than 10% participants are giving 30% or 40% of the given endowment. This is also broadly in line with typical findings that report average transfers of roughly 50% of endowment. Panel F shows that almost half the participants in PGG send their entire endowment into the public pool. The next most popular choice (20%) is to send half their endowment to the public pool, and only 4% of participants sent 0% into the public pool. Again, this broadly matches usual results in literature.

Figure 2c has 2 panels (Panels G and H) that displays the distribution of reciprocity levels and average repayment rate, respectively, which are derived from participants' responses in TG2. Panel G shows that, with the exception of two participants who demonstrated negative reciprocity (perhaps they did not understand the game), reciprocity levels are very heterogeneous between 0 and 1. There is a spike (27%) at reciprocity levels around 0.5, and a lower spike (14%) around 0.35. There is also a small spike (10%) at 0, which represents those who did not show any reciprocity. Panel H shows a strong peak of repayment rates of around

45%, with participants showing slightly stronger preference repaying less than half of what was received versus repaying more than half of what was received. These results are in line with typical patterns found in previous literature, which report average repayment rates of nearly 50% of the transfer.

3.2 Self-Reported Measures of Past Social Behaviors

Figure 3 shows the distribution of total scores obtained by the participants on the SRA Scale. Scores are obtained by summing across each participant’s responses for the 10 items (“Never” = 0, “Once” = 1, “More than once” = 2, “Often” = 3, “Very often” = 4). A higher SRA score indicates higher pro-social behavior. There was a wide variety in the total SRA scores obtained, ranging between 20 and 48. Scores are centered around 33 and the shape is symmetric.

3.3 Donations Behavior

Figure 4 shows the distribution of Wesleyan donations. Panel A shows that 64% of the participants have donated to Wesleyan University before, and 36% have not. Panel B shows the distribution of total donation amounts. Most of the donation amounts are between \$0 and \$50, with more donation amounts below \$25. There are few donations that are above \$100, and there are two large donations that are not shown (\$2,171 and \$16,863).

3.4 External Validity of Social Preference Games

4 Importance and Benefits

First, the findings in this experiment will add to growing literature and debate on whether experimental games are a useful tool to examine social preferences. The

current accumulated evidence is mixed - some studies find that in-lab behavior is correlated to field behavior, while others do not find a significant relationship. However, the studies map only one experimental game to one field measure. My experience, which uses a variety of experimental games along with self-reported social behavior questions, therefore provides a systematic approach to uncovering the external validity of experimental social preference games.

Along with adding to the literature on external validity of experimental social preference games, there is a potential policy implication with our findings from a research perspective. Experimental games are costly - not only is it expensive paying participants for attending the lab session and providing money aligned with the payoffs of the games (especially if more show up than expected), but it also takes a lot of time to program the experimental games. If experimental games do not better predict field measures, then perhaps eliminating experimental games and instead using non-incentivized survey questions is better for funding and logistical purposes.

Finally, on a larger scale, this study can potentially spark interest into future studies into the external validity of other behavioral economics topics where lab experiments are also commonly used. For example, using experiments to study risk preferences and time preferences is common, so it will be interesting to see whether in-lab behavior is correlated to field measures, and perhaps adopting a systematic approach (comparing a variety of lab experiments to various field measures) can prove to be beneficial.

References

- [1] Andreoni, J., and Miller, J.H. (2002). *Giving according to GARP: An experimental test of the consistency of preferences for altruism*. *Econometrica*, 70, 737-53.
- [2] Baran, N.M., Sapienza, P., and Zingales, L. (2010). *Can we infer social preferences from the lab? Evidence from the trust game*. NBER Working Paper 15654.
- [3] Benz, M., and Meier, S. (2006). *Do people behave in experiments as in the field? Evidence from donations*. *Experimental Economics*, 11, 268-81.
- [4] Berg, J., Dickhaut, J.W., and McCabe, K.A. (1995). *Trust, reciprocity, and social history*. *Games and Economic Behavior*, 90, 166-93.
- [5] Burnham, T., McCabe, K., Smith, V. (2000). *Friend-or-foe intentionality priming in an extensive form trust game*. *Journal of Economic Behavior & Organization*, 43, 57-73.
- [6] Carpenter, J.P., Verhoogen, E., and Burks, S. (2005). *The effect of stakes in distribution experiments*. *Economics Letters*, 86, 393-98.
- [7] Charness, G., and Rabin, M. (2002). *Understanding social preferences with simple tests*. *Quarterly Journal of Economics*, 117, 817-69.
- [8] Cherry, T., Fykblom, P., and Shogren, J. (2002). *Hardnose the Dictator*. *American Economic Review*, 92(4): 1218-21.
- [9] Englmaier, F., and Gebhardt, G. (2010) *Free-riding in the lab and in the field*. CESifo Working Paper No. 3612.

- [10] Fehr, E., and Leibbrandt, A. (2011) *A field study on cooperativeness and impatience in the Tragedy of the Commons*. Journal of Public Economics, 95, 1144-55.
- [11] Fehr, E., and Schmidt, K. (1999). *A theory of fairness, competition, and cooperation*. Quarterly Journal of Economics, 114, 173-68.
- [12] Fisman, R., Jakiela, P., and Kariv, S. (2014). *The distributional preferences of Americans*. NBER Working Paper.
- [13] Fisman, R., Kariv, S., and Markovits, D. (2007). *Individual Preferences for Giving*. American Economic Review, 97(5): 1858-76.
- [14] Franzen, A., and Pointner, S. (2013) *The external validity of giving in the dictator game: A field experiment using the misdirected letter technique*. Experimental Economics, 16, 155-69.
- [15] Galizzi, M., and Navarro-Martinez, D. (2017). *On the external validity of social preference games: a systematic lab-field study*. Management Science.
- [16] Gintis, H. (2000). *Strong reciprocity and human sociality*. Journal of Theoretical Biology, 206, 169-79.
- [17] Gneezy, U., Haruvy, E., and Yafe, H. (2004). *The inefficiency of splitting the bill: A lesson in institutional design*. The Economic Journal, 114(495), 265-80.
- [18] Goeschl, T., Kettner, S.E., Lohse, J., and Schwieren, C. (2015) *What do we learn from public good games about voluntary climate action? Evidence from an artefactual field experiment*. University of Heidelberg, Department of Economics Discussion Paper 595.
- [19] Gurven, M., Winking, J. (2008). *Collective action in action: prosocial behavior in and out of the laboratory*. American Anthropologist, 110(2), 179-190.

- [20] Henrich, J., et al. (2005). *“Economic Man” in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies*. Behavioral and Brain Sciences. 28(6): 795-815.
- [21] Hermann, B., Thoni, C., and Gächter, S. (2008). *Anti-social punishment across societies*. Science, 319, 1362-67.
- [22] Hill, K., and Gurven, M. (2004) *Economic experiments to examine fairness and cooperation among the Ache Indians of Paraguay*. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (Eds.), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press.
- [23] Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). *Preferences, property rights, and anonymity in bargaining games*. Games and Economic Behavior, 7(3): 346-80.
- [24] Karlan, D.S. (2005). *Using experimental economics to measure social capital and predict financial decisions*. American Economic Review, 95, 1688-99.
- [25] Levitt, S., and List, J. (2007) *What do laboratory experiments measuring social preferences reveal about the real world?*. Journal of Economic Perspectives, 21(2), 153-74.
- [26] List, John. 2006. *The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions*. Journal of Political Economy, 114(51), 1-37.
- [27] Parco, J., Rapoport, A., and Stein, W. (2002). *Effects of financial incentives on the breakdown of mutual trust*. Psychological Science.

- [28] Rabin, M. (1993). *Incorporating fairness into game theory and economics*. The American Economic Review, 83, 1281-1302.
- [29] Ross, L., and Ward, A. (1996). *Naive realism in everyday life: Implications for social conflict and misunderstanding*. Values and Knowledge, 103-35.
- [30] Rushton, J.P., Chrisjohn, R.D., and Fekken, G.C. (1981). *The altruistic personality and the self-report altruism scale*. Personality and Individual Differences, 2, 293-302.
- [31] Slonim, R., and Roth, A. (1998). *Learning in high stakes ultimatum games: An experiment in the Slovak Republic*. Econometrica, 66, 569-96.
- [32] Voors, M., Turley, T., Kontoleon, A., Bulte, E., and List, J.A. (2012). *Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone*. Economic Letters, 114, 308-311.

A Tables

SRA Questions List
1. I have allowed someone to go ahead of me in line.
2. I have donated money at the cash register when buying groceries.
3. I have given money to a stranger (or an acquaintance I don't know too well) in need.
4. I have donated instead of sold my clothes/used items.
5. I have donated to a charity.
6. I have donated blood.
7. I have done volunteer work for a charity/organization.
8. I have delayed an elevator/held door open for stranger(s).
9. I have lent an acquaintance that I don't know too well with something of value to me (clothes, tools, etc).
10. I have pointed out a clerk's error (at a supermarket, restaurant) in undercharging me.
11. I have gone out of my way to meet with someone to help them with a task (e.g. help proofread their paper, listen to their presentation, etc).
12. I have helped carry a stranger's belongings (e.g. groceries).
13. I have offered my seat on a bus/train to a stranger who was standing.
14. I have helped an acquaintance with moving in/ moving out of their dorm/apartment/house.
15. I have donated money/coins to the Salvation army bell-ringers.

B Figures

Figure 1: Online games and survey questions

You are **Player 1**. You will decide how to divide a given number of tickets between yourself and an anonymous, random Player 2. Each ticket you keep will be worth multiplied by either 1, 2, 3, or 4, and similarly for Player 2.

For each question, there is a slider that indicates how many tickets you would like to give Player 2. Above each amount are the tickets you and Player 2 will receive, represented as (tickets you receive, tickets other player receives).

Please choose how you would like to divide the tickets.

Divide 15 tickets: Hold @ 1 tickets, Pass @ 2 tickets. How many points do you want to pass to other the player?

(15, 0)	(14, 2)	(13, 4)	(12, 6)	(11, 8)	(10, 10)	(9, 12)	(8, 14)	(7, 16)	(6, 18)	(5, 20)	(4, 22)	(3, 24)	(2, 26)	(1, 28)	(0, 30)
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15



Divide 10 tickets: Hold @ 1 ticket each, Pass @ 3 tickets each. How many tickets do you want to pass to the other player?

(10, 0)	(9, 3)	(8, 6)	(7, 9)	(6, 12)	(5, 15)	(4, 18)	(3, 21)	(2, 24)	(1, 27)	(0, 30)
0	1	2	3	4	5	6	7	8	9	10



Divide 15 tickets: Hold @ 2 tickets each, Pass @ 1 ticket each. How many tickets do you want to pass to the other player?

(30, 0)	(28, 1)	(26, 2)	(24, 3)	(22, 4)	(20, 5)	(18, 6)	(16, 7)	(14, 8)	(12, 9)	(10, 10)	(8, 11)	(6, 12)	(4, 13)	(2, 14)	(0, 15)
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15



Divide 12 tickets: Hold @ 1 ticket each, Pass @ 2 tickets each. How many tickets do you want to pass to the other player?

(12, 0)	(11, 2)	(10, 4)	(9, 6)	(8, 8)	(7, 10)	(6, 12)	(5, 14)	(4, 16)	(3, 18)	(2, 20)	(1, 22)	(0, 24)
0	1	2	3	4	5	6	7	8	9	10	11	12



Divide 10 tickets: Hold @ 3 tickets each, Pass @ 1 ticket each. How many tickets do you want to pass to the other player?

(30, 0)	(27, 1)	(24, 2)	(21, 3)	(18, 4)	(15, 5)	(12, 6)	(9, 7)	(6, 8)	(3, 8)	(0, 10)
0	1	2	3	4	5	6	7	8	9	10



Divide 20 tickets: Hold @ 1 ticket each, Pass @ 1 ticket each. How many tickets do you want to pass to the other player?

(20, 0)	(19, 1)	(18, 2)	(17, 3)	(16, 4)	(15, 5)	(14, 6)	(13, 7)	(12, 8)	(11, 9)	(10, 10)	(9, 11)	(8, 12)	(7, 13)	(6, 14)	(5, 15)	(4, \$16)	(3, 17)	(2, 18)	(1, 19)	(0, 20)
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20



Divide 20 tickets: Hold @ 1 ticket each, Pass @ 2 tickets each. How many tickets do you want to pass to the other player?

(20, 0)	(19, 2)	(18, 4)	(17, 6)	(16, 8)	(15, 10)	(14, 12)	(13, 14)	(12, 16)	(11, 18)	(10, 20)	(9, 22)	(8, 24)	(7, 26)	(6, 28)	(5, 30)	(4, 32)	(3, 34)	(2, 36)	(1, 38)	(0, 40)
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20



Divide 12 tickets: Hold @ 2 tickets each, Pass @ 1 ticket each. How many tickets do you want to pass to the other player?

(24, 0)	(22, 1)	(20, 2)	(18, 3)	(16, 4)	(14, 5)	(12, 6)	(10, 7)	(8, 8)	(6, 9)	(4, 10)	(2, 11)	(0, 12)
0	1	2	3	4	5	6	7	8	9	10	11	12



Divide 16 tickets: Hold @ 1 ticket each, Pass @ 2 tickets each. How many tickets do you want to pass to the other player?

(16, 0)	(15, 2)	(14, 4)	(13, 6)	(12, 8)	(11, 10)	(10, 12)	(9, 14)	(8, 16)	(7, 18)	(6, 20)	(5, 22)	(4, 24)	(3, 26)	(2, 28)	(1, 30)	(0, 32)
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16



Divide 10 tickets: Hold @ 4 tickets each, Pass @ 1 ticket each. How many tickets do you want to pass to the other player?

(40, 0)	(36, 1)	(32, 2)	(28, 3)	(24, 4)	(20, 5)	(16, 6)	(12, 7)	(8, 8)	(4, 9)	(0, 10)
0	1	2	3	4	5	6	7	8	9	10



Divide 10 tickets: Hold @ 1 ticket each, Pass @ 4 tickets each. How many tickets do you want to pass to the other player?

(10, 0)	(9, 4)	(8, 8)	(7, 12)	(6, 16)	(5, 20)	(4, 24)	(3, 28)	(2, 32)	(1, 36)	(0, 40)
0	1	2	3	4	5	6	7	8	9	10



Next

You are **Player 1** (the proposer). You are endowed with 10 tickets. Please decide how much of your 10 ticket you would like to send to an anonymous, random Player 2.

After receiving your donated amount, Player 2 may choose to accept or reject the offer. If your offer is rejected, you both will receive 0 tickets.

Please decide how much of the 10 tickets you will give to Player 2:

Next

You are **Player 2** (the responder). Player 1 (the proposer) was endowed with 10 points and can choose to share their endowment with you.

You can choose to accept or reject Player 1's offer. **If you reject the offer, you both will receive 0 points.**

For each possible amount Player 1 can offer you, please state if you accept or reject the offer.

	Reject	Accept
Player 1 passes 0 points (keeps 10 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 1 point (keeps 9 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 2 points (keeps 8 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 3 points (keeps 7 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 4 points (keeps 6 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 5 points (keeps 5 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 6 points (keeps 4 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 7 points (keeps 3 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 8 points (keeps 2 points)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 9 points (keeps 1 point)	<input type="radio"/>	<input type="radio"/>
Player 1 passes 10 points (keeps 0 points)	<input type="radio"/>	<input type="radio"/>

You are **Player 1** (the proposer). You are endowed with 10 tickets. Please decide how much of your 10 tickets to send to an anonymous, random Player 2. The amount sent over will be multiplied by 3. Player 2 will then decide how much of the multiplied amount they will return back to you.

Please decide how much of the 10 tickets you will give to Player 2:

Next

You are **Player 2** (the responder).

Player 1 was endowed with 10 points, and can choose to share their endowment with you. The amount Player 1 donates to you will be multiplied by 3. You will be given the opportunity to return some of the points back to Player 1.

For each possible amount that Player 1 can give to you, please state how many points you would like to send back to Player 1.

	How many points you would like to return:
Player 1 passes 1 point, i.e. you receive 3 points	<input type="text"/>
Player 1 passes 2 points, i.e. you receive 6 points	<input type="text"/>
Player 1 passes 3 points, i.e. you receive 9 points	<input type="text"/>
Player 1 passes 4 points, i.e. you receive 12 points	<input type="text"/>
Player 1 passes 5 points, i.e. you receive 15 points	<input type="text"/>
Player 1 passes 6 points, i.e. you receive 18 points	<input type="text"/>
Player 1 passes 7 points, i.e. you receive 21 points	<input type="text"/>
Player 1 passes 8 points, i.e. you receive 24 points	<input type="text"/>
Player 1 passes 9 points, i.e. you receive 27 points	<input type="text"/>
Player 1 passes 10 points, i.e. you receive 30 points	<input type="text"/>

You will be randomly matched with one other anonymous, random player. You each are endowed with 10 tickets, and will decide how much of your respective endowments to contribute to a common group fund. The overall tickets in the fund will be multiplied by 2, and divided evenly between the both of you.

Now please decide how much of your 10 tickets to put into the common group fund:

Next

Please indicate the frequency you have done each statement in the past.

	Never	Once	More than once	Often	Very often
I have allowed someone to go ahead of me in line	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have donated money at the cash register when buying groceries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have given money to a stranger (or lent to an acquaintance I don't know too well) in need	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have donated instead of sold my clothes/used items	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have donated to a charity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have donated blood	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have done volunteer work for a charity/organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have delayed an elevator/held door open for a stranger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have lent an acquaintance that I don't know too well with something of value to me (clothes, tools, etc)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have pointed out a clerk's error (at a supermarket, restaurant) in undercharging me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have gone out of my way to meet with someone to help them with a task (e.g. help proofread their paper, listen to their presentation, etc)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have helped carry a stranger's belongings (e.g. groceries)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have offered my seat on a bus/train to a stranger who was standing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have helped an acquaintance with moving in/ moving out of their dorm/apartment/house	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have donated money/coins to the Salvation army bell-ringers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 2a: Distribution of CES parameters from GDG

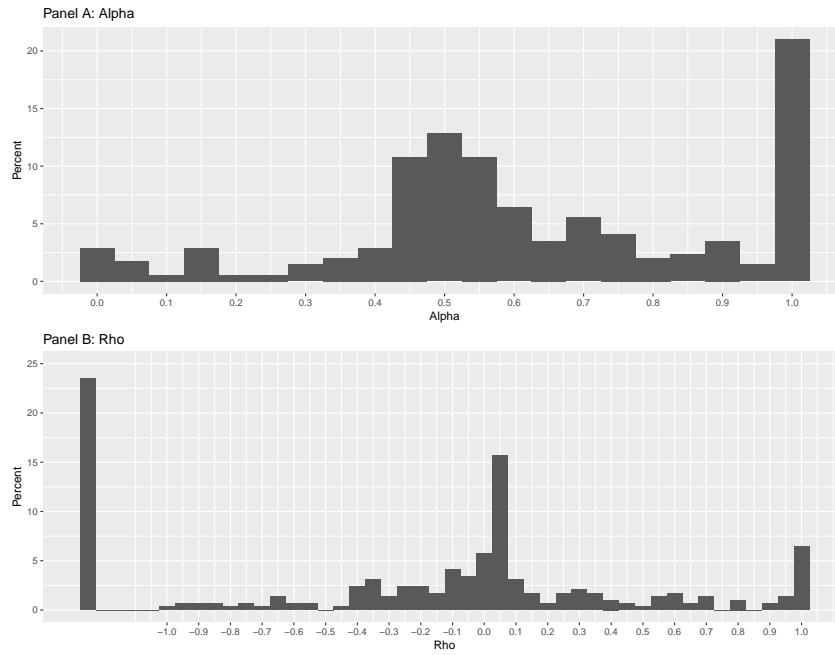


Figure 2b: Distribution of responses in UG1, UG2, TG1, and PGG

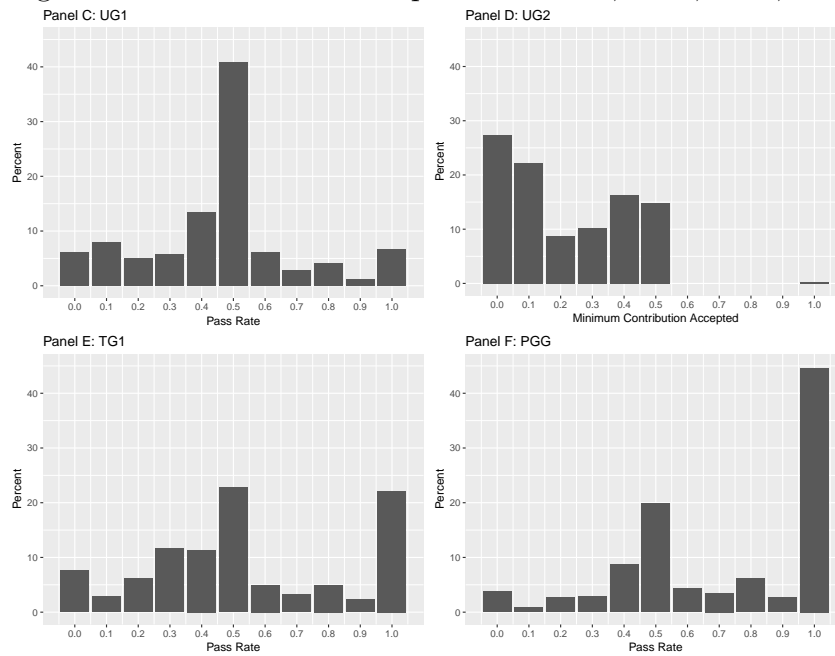


Figure 2c: Distribution of responses from TG2

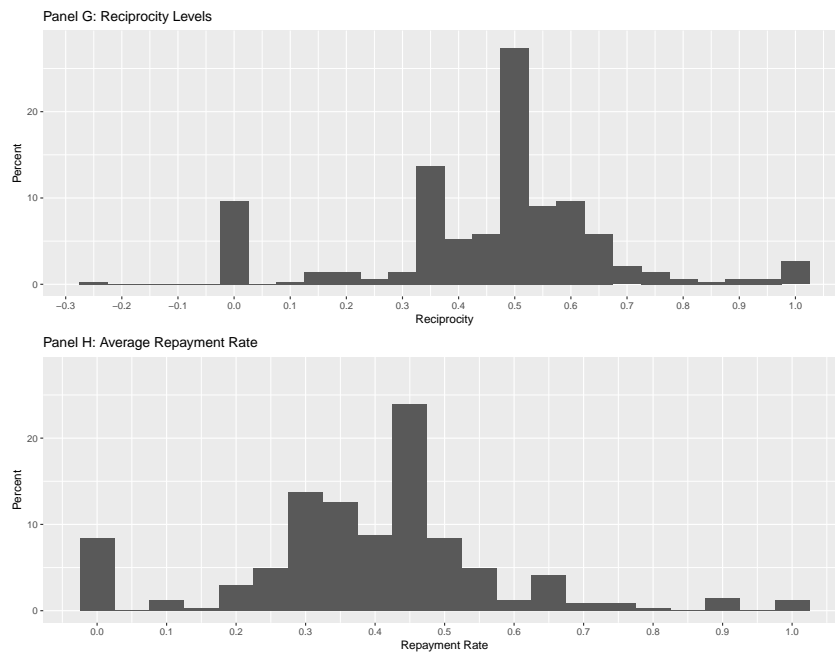


Figure 3: Total SRA scores

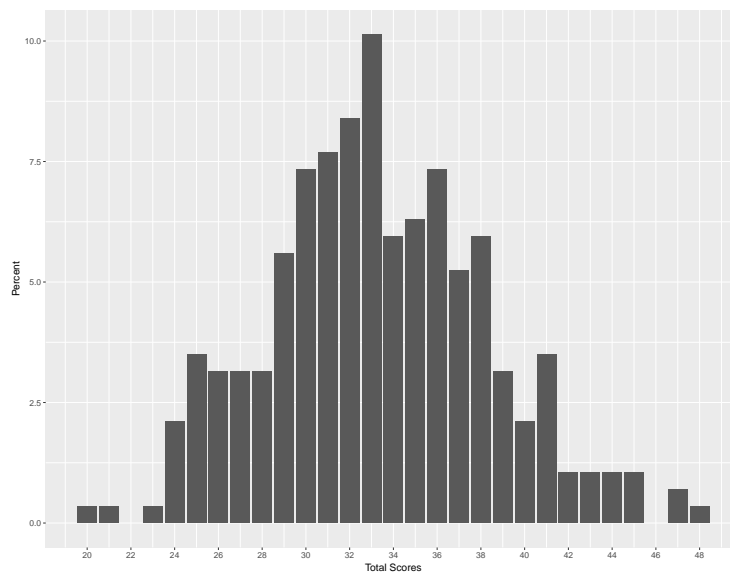


Figure 4: Donations

