The External Validity of Experimental Social Preference Games

Sarah Xu

April 2018

Abstract

The external validity of experimental games is a growing topic of interest. This topic is important because researchers commonly use such games to study individuals' social preferences. Many papers have conducted within-subjects experiments, comparing subjects' in-lab decisions to their real-world decisions. However, the current accumulated evidence yield mixed conclusions: some papers found statistically significant lab-field correlations, while other papers did not. This research provides a systematic approach to exploring this matter: I compared behavior in various social preference games and self-reported measures regarding past social behavior to a natural, far-removed field measure. My results show social preference games do a poor job explaining the field behavior. Furthermore, self-reported measures perform just as well, if not better, than the games in both explaining and predicting the field measure.

Contents

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 well-being 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.

When studying social preferences, a common 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[1]. Below are common games used to study social preferences:

- Dictator Game: The player (the dictator) receives an endowment and chooses how to split the amount between themselves and a second player (the receiver).

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

- Trust Game: Two-player game where the first player (the proposer) receives an endowment and proposes how to divide the endowment between themselves and the second mover. The offered amount is multiplied by a factor, and the second player (the responder) decides how much of the multiplied endowment to return to the proposer.

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

---

[1]Many papers have identified various aspects of social preferences such as altruism, social welfare, inequality aversion, and reciprocity (e.g., Charness and Rabin (2002); Fehr and Schmidt (1999); Andreoni and Miller (2002); Rabin (1993); Fisman et al. (2014)).

sum of money and simultaneously decides how much to put into a public
fund. The total amount in the public fund is multiplied by a factor, and
divided evenly amongst the players.

A big question is the external validity of these social preference games since
there are numerous factors that can lead to differences in behaviors in a lab setting
versus real-world setting. There are many papers that explore this question, com-
paring subjects' lab behavior to the same subjects' real-world behavior. Galizzi
and Navarro-Martinez (2017) summarized that about 40% of reported lab-field
correlations and regressions found statistically significant association between lab
and field behaviors[2]. Therefore there is room for further exploration. Similar to
previous lab-field experiments, I examined whether in-lab social behavior predicts
real-life social behavior. Adopting the experimental design in the study by Gal-
izzi and Navarro-Martinez to provide a systematic approach to the question, I
recruited Wesleyan University seniors and recent alumni to play various experi-
mental games. I also had the participants answer self-reported questions regarding
their past social behavior as an additional layer to evaluate the predictive ability
of the games, and I used donations to Wesleyan University as the field measure.

My results show that social preference games explain little to none of the
donations behavior. In fact, the self-reported measures seem to be a better tool
than the games. In addition, self-reported measures play a more important role
when predicting the field measure. However, further systematic research should
be done in order to draw a definite conclusion regarding the external validity of
social preference games.

---

[2]See Galizzi and Navarro-Martinez for their full systematic review and meta-analysis of lit-
erature on within-subjects lab-field studies.

2 Literature Review

There are many papers that study social preferences using experimental games. For example, Forsythe et al. (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 in which both proposers and responders do not send any money, most players in fact did share proportions of their endowments. Berg et al. (1995) recruited undergraduate students from the University of Minnesota to participate in trust games. They found that people showed high levels of trust and reciprocity: 94% of the proposers sent money, and over one-third of the responders returned an amount greater than what was given to them. Hermann et al. (2008) conducted public goods experiments across 16 participant pools, and found that that varying opportunities for punishment led to varying cooperation levels. These papers are only a few examples from an abundance of literature (with many papers using social preference games not mentioned in the previous section), making it clear that using social preference games has become one of the building blocks of experimental and behavioral economics.

However, because laboratory experiments are abstract and remote from realistic situations, an important question is the extent to which experimental social preference games can be generalized to real world behavior. 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 concluded 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 framing of the situation. For example, in a two-person trust game, Burnham et al. (2000) switched between calling the responder as "partner" or "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. Some studies also found that varying the level of stakes may lead to significantly different behaviors. For example, Carpenter et al. (2005) discovered 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 et al. (2002) also found that higher stakes leads to significantly different results than with lower stakes. Interestingly, however, Cherry et al. (2002) found no differences in offers when increasing the stakes from $10 to $40. Of course, scrutiny, context, and stakes are only a few factors that can lead to deviations in behavior[3].

The fact that 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 explain field behavior. Baran et al. (2010) compared MBA alumni donations to their university with their reciprocity behavior when playing the trust game. The authors found that responder behavior predicted university donations. Franzen and Pointner (2012) 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 proso-

---

[3]See Levitt and List (2007) for a full discussion of factors, with supporting literature, that may lead to deviations in behavior between laboratory and field settings.

cial behavior in the lab returned the misdirected letters more often than subjects who were selfish in the lab. Englmaier and Gebhardt (2011) conducted field experiments where they compared university students' free riding behavior at the library to free riding behavior in a public goods game, and found statistically significant correlation between field and lab measures. There are also studies that used non-student subjects: Fehr and Leibbrandt (2011) conducted public goods games with Brazilian fishermen, and found that those who were 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 found lab behavior had no explanatory 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 CO_2 emissions. The authors found that behaviors in both tasks were uncorrelated. Hill and Gurven (2004) carried out ultimatum and public goods games on Paraguay Ache Indians. They compared the game behaviors to observed food production and sharing patterns to individuals outside the nuclear family, and found no significant relationship 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 concluded 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 found no meaningful correlation in behavior between the lab and field measures.

With other studies finding statistically significant, not statistically significant,

and mixed results, it is clear that evidence for the external validity of experimental games is weak. Galizzi and Navarro-Martinez noted, however, that the previous studies compared 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 proceeded 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 was that behavior when playing the experimental games does a poor job predicting both the survey questions and the field behaviors. They mentioned, however, that more systematic studies are needed in order to draw a definite conclusion.

Similar to Galizzi and Navarro-Martinez, my research provides another systematic approach to this issue by having participants play multiple social preference games that all are theoretically mapped to my field measure. Furthermore, it is critical to think about why some studies found correlations between lab and field measures and some studies did not, in order to further inform my experiment design. While playing the experimental games for the study by Hill and Gurven, 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. In the studies by Gurven and Winking and Voors et al., participants were also subject to high scrutiny and low anonymity since it was easy for community members to find out the choices each subject made. Therefore an important aspect in my research design is that participants played the games in their own time and location. This set-up ensured participants were ensured minimal scrutiny and higher anonymity.

In addition, a criticism of the experimental design by Galizzi and Navarro-Martinez is how the authors executed their field measures. Since the field situations occured as the participants were 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 in fashion. This may have led to actions that participants would not have done if they were unaware of the scrutiny. This is called the experimenter demand effect, which is a common problem in experimental design: if participants know their actions are being scrutinized, they may change their behavior towards what constitutes appropriate behavior. Therefore it is better to use a natural, far-removed situation, which is why I used donations to Wesleyan University as my field measure.

3   Methods

Each participant was presented with two sets of tasks: (i) incentivized social preference games, and (ii) self-reported questions regarding past social behavior. The field measure used to compare to participants' lab results is their donations to Wesleyan University. In addition, I examined which tool, the social preference games, the self-reported measures, or both, is better predict social preferences exhibited in real-world situations.

Wesleyan University seniors and recent alumni received an invitation email

9

briefly explaining my research and asking for participation in my study (2,004 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 allowed participants to make decisions anonymously and with minimal scrutiny, reducing 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.

3.1   Incentivized social preference games

Those who chose to participate in the study received a unique identifier assigned by Qualtrics. Participants first played four social preference games (in random order): 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 had the opportunity to play the roles of both proposer and responder. At the end of the participation deadline, all participants were randomly paired, and ticket payoffs were calculated.

Below are the games each participant played, along with an example for each

game:

- Generalized Dictator Game[4]: Each participant, playing as the dictator, played nine different rounds of the generalized dictator game. In each round, participants were given different endowments and prices of giving, and were asked how they would like to divide the endowment with an anonymous, random Player 2. Therefore the budget constraint is given by $\pi_s + p\pi_o = m$, where π_s is how many tickets were kept, π_o is how many tickets were given, p is the relative price of giving, and m is the endowment. The endowment in each round was either 10, 12, or 15. Every ticket the participant kept was worth either 1, 2, 3, or 4 tickets (hold price), and every ticket given to Player 2 was also worth either 1, 2, 3, or 4 tickets (pass price). Below are the nine sets of endowments and prices of giving used in the study:

Budget	Ticket Endowment	Hold Price	Pass Price	Relative Price of Giving
1	15	1	2	$\frac{1}{2}$
2	10	1	3	$\frac{1}{3}$
3	15	2	1	2
4	12	1	2	$\frac{1}{2}$
5	10	3	1	3
6	15	1	1	1
7	12	2	1	2
8	10	4	1	4
9	10	1	4	$\frac{1}{4}$

Participants were given the endowment and hold/pass price for each round, and were presented an interactive slider bar that displayed the different

---

[4]The design is based off of the study by Andreoni and Miller.

ticket amounts each player could earn. The slider bar ensured choices fulfilled the budget constraint, and eliminated the need to perform calculations by directly showing how many tickets each player would receive.

Example: Divide 10 tickets: Hold @ 1 ticket each, Pass @ 3 tickets each. In this example, the endowment is 10 tickets, and the relative price of giving is $\frac{1}{3}$. That is, however many tickets the player decides to keep is multiplied by 1 (the hold price), and however many tickets given to Player 2 is multiplied by 3 (the pass price). The player is presented a slider bar, ranging from giving 0 tickets to giving 10 tickets, and with each possible offer amount, the player was told the total ticket amounts received by both players in the form (total tickets received, total tickets given). For example, if the player decides to give 4 tickets to Player 2, they would see (6, 12) which means they would receive 6 tickets (10 - 4 = 6) and Player 2 would receive 12 tickets (4 * 3 = 12).

- Ultimatum Game:

  - Player 1: Each participant was endowed with 10 tickets, and were told to decide how much of their endowment to send to an anonymous, random responder (Player 2) so that $\pi_s + \pi_o = 10$. They were also informed that the responder may or may not reject the proposed allocation: if the allocation is accepted, then the proposal is implemented (Player 1 will receive π_s and Player 2 will receive π_o), and if the allocation is rejected, neither player receives any tickets.

    Example: Suppose the participant chooses to give 3 tickets to Player 2. If Player 2 accepts the proposal, then the participant will receive 7 tickets (10 - 3 = 7) and Player 2 will receive 3 tickets. However, if Player 2 does not like the proposal and rejects the proposition, then

both players will receive 0 tickets.

- – Player 2: Each participant was informed that the proposer (Player 1) was given an endowment of 10 tickets. They were presented a list from 0 tickets to 10 tickets (in 1-ticket increments), which represents the different amounts that Player 1 could choose to give them. The participants were instructed to indicate whether they accept or reject each hypothetical proposal, and were told that accepting means they agree to receiving the offered amount, and rejecting means they do not agree and instead both players will receive 0 tickets.

  Example: Suppose the participant chooses to reject all offered amounts below 5 tickets, and accept all offered amounts equal to or greater than 5 tickets. If Player 1 chose to offer 3 tickets, then both players will receive 0 tickets. However, if Player 1 chose to offer 6 tickets, then Player 1 will receive 4 tickets (10 - 6 = 4) and the participant will receive 6 tickets.

- Trust Game:

  - – Player 1: Each participant was endowed with 10 tickets. They were prompted to decide how much of their endowment to send to an anonymous, random responder (Player 2) so that $\pi_s + \pi_o = 10$. The participants were told that the amount sent over would be multiplied by 3, i.e. Player 2 would receive $3\pi_o$, and that Player 2 would then decide how many tickets, $r \leq 3\pi_o$, they would like to return. Overall, Player 1 received $\pi_s + r$ tickets and Player 2 received $3\pi_o$ - r tickets.

    Example: If the participant chooses to send 5 tickets, then Player 2 will receive 15 tickets (5 * 3 = 15). Player 2 will then decide how many tickets to return. If Player 2 chooses to 3 tickets, then overall Player 1

will earn 8 tickets ($5 + 3 = 8$) and Player 2 will earn 12 tickets ($15 - 3 = 12$).

– Player 2: Each participant was informed that the proposer (Player 1) received an endowment of 10 tickets. They were then given a list of all ten possible multiplied amounts that Player 1 could have chosen to send, ranging from 3 tickets to 30 tickets in 3-ticket increments. For each possible offered amount, the participants were prompted to enter the number of tickets they would like to return back to the proposer. They were given clear instructions that their input could not exceed what was given to them. For example, if Player 1 gave them a multiplied amount of 15 tickets, then the maximum number of tickets the participant could return is 15 tickets.

Example: Suppose the participant's return scheme is as below:

Possible Offer Amounts (Multiplied)	Tickets Returned
3	0
6	1
9	3
12	3
15	5
18	5
21	5
24	5
27	5
30	5

where the first column is a list of all possible multiplied amounts that

14

Player 1 could choose to send, and the second column indicates how many tickets the participant would like to return. If Player 1 decided to give 7 tickets to the participant, then the participant will receive a multiplied amount of 21 tickets (7 * 3 = 21). As indicated by the participant's return scheme, the participant chose to return 5 tickets. Thus Player 1 receives 8 tickets ($3 + 5 = 8$), and the participant receives 16 tickets (21 - 5 = 16).

- Public Goods Game: Each participant was given 10 tickets, and told that they would be randomly matched with one other player. They were then prompted to decide how much of their endowment to contribute to a group fund, and that the other player was prompted to do the same task. The total tickets in the group fund was multiplied by 2, and divided evenly between the two players. Therefore each player received the remaining amount of their endowment plus the divided amount from the public fund. The payoff function is given by $P_i = 10 - g_i + \sum_{n=1}^{2} g_n$ where g_i is the amount that player i donated to the fund, and $\sum_{n=1}^{2} g_n$ is the sum of both players' donations to the public fund[5].

    Example: Suppose the participant chooses to contribute 3 tickets, and the randomly assigned second player chooses to contribute 5 tickets. The total donated amount, $\sum_{n=1}^{N} g_n$ is 8 tickets. Therefore the participant receives:

    $P_1 = 10 - g_1 + \sum_{n=1}^{2} g_n = 10 - 3 + 8 = 15$ tickets

    and the second player receives:

    $P_2 = 10 - g_2 + \sum_{n=1}^{2} g_n = 10 - 5 + 8 = 13$ tickets

---

[5]Note that the players directly receive the total amount donated to the public fund because the multiplied factor, 2, is cancelled by the even split between the two players.

Appendix C contains all materials (images of each game along with their respective instructions and examples) from the online study.

3.1.1 Strategy Method

Since the participants were 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 after the fact. 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 trust game, and provides participants' minimum accepted amount in the ultimatum game.

3.1.2 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

16

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. Winners were emailed instructions on how to receive the prize through direct deposit.

3.2   Self-reported measures of past social behaviors

Participants then reported on social behaviors exhibited in the past. The questions were adapted from the Self-Report Altruism (SRA) scale introduced by Rushton et al. (1981). The survey was comprised of 10 items, and participants reported how frequently they have done each item in the past. Participants rated each statement as either "Never", "Once", "More than once", "Often", or "Very often". Examples included: "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 from the online study is displayed in Appendix D.

3.3   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 the online study was released between solicitation cycles for both seniors and recent alumni (early January 2018), and donations information was provided as of January 2018, participants' donations behavior was not influenced by the experimenter demand effect.

There are several reasons why I chose to use donations as my field measure. Many behavioral economics papers exploring the relationship between in-lab and

17

field behavior have used donations[6]. Falk et al. (2013) justified that donations are an accurate field measure because they do not rely on self-reported responses but on actual decisions. Furthermore, donations are made in private and never made public, and students/alumni are unaware that their actions will be analyzed in a future research study. The lack of scrutiny therefore indicates donations should reflect the donors' genuine altruism. The authors then pointed out the most important reason for their decision to use donations: all students and alumni are solicited, so everybody has to make the decision about donating.

In addition, donations are a more accurate field measure than creating a scenario at the end of the online session. I previously mentioned that the experimental design by Galizzi and Navarro-Martinez is unsatisfactory since their participants were subject to the experimenter demand effect: because participants encountered the field experiments shortly after completing the lab experiments, it would not be difficult for participants to link the two situations together. Therefore, participants could have acted more pro-socially than they would usually under normal circumstances. In this research, participants' donations are a far-removed situation, and thus are not affected by their participation in the lab.

3.3.1 Theoretical Correspondences Between Games and Donations

All of the social preference games I chose for my study tap into different types of pro-social behaviors that are also related to donating. First, the dictators in the generalized dictator game, proposers in the ultimatum and trust games, responders in the trust game, and players in the public goods game have the option to keep their endowment to themselves, but may also choose to share their endowment with another person. Therefore the actions in these games can be explained by altruism: those who choose to keep their endowment exhibit lower

---

[6]See, e.g., Falk, Meier, and Zehnder; Benz and Meier (2006); Baran et al.

levels of altruism, and those who choose to share their endowment demonstrate higher levels of altruism. Likewise, altruism is the most common representation for donating since the donor chooses to give their income to the university to benefit future students.

Proposers' behaviors in the trust and public goods games indicate their trust levels. In the trust game, since the proposers' offerings get multiplied by a factor, the proposers may exhibit trust by sharing their endowment with the hopes of the responders sending back an amount larger than what was given. Similarly, in the public goods game, the total amount donated into the public fund gets multiplied by a factor. Thus, players who donate some of their endowment into the public fund may do so with the hopes that the other players will do the same, and they will receive a net amount larger than their initial endowment. Correspondingly, donating may also represent trust since donors are trusting that the University will use their donations to benefit future students.

Responders' actions in the trust game can also be explained by reciprocity: if the proposers decided to share their endowment, the responders then has the opportunity to give back to the proposers. The responders display higher levels of reciprocity if they return some of what was first given to them, and display lower levels of reciprocity if they give back little or nothing. The responders' actions can also be represented by trustworthiness: the proposers may contribute some of their endowment in hopes of the responders returning a larger amount than what was offered. Thus responders who return a larger amount show higher levels of trustworthiness. Equivalently, donations may symbolize reciprocity and trustworthiness since seniors and alumni may donate as thanks for providing them invaluable resources and opportunities.

Players' responses in all of the games can also correspond with fairness preferences, more specifically inequality aversion. Dictators in the generalized dictator

game, proposers in the ultimatum and trust games, responders in the trust game, and players in the public goods game may choose to share their endowment so that the other player will not be left with nothing. Similarly, responders in the ultimatum game may exhibit fairness preferences by rejecting offer amounts of zero (and those who exhibit extreme fairness preferences may reject offer amounts lower than half the endowment). Donations can also symbolize inequality aversion: people may donate because they believe their donations will help students receive opportunities they would not be able to obtain otherwise. In addition, donors are able to target what area their donations can go towards: for example, if the donor targets their donations to Financial Aid, then they are displaying inequality averse preferences.

Lastly, players' behaviors in the public goods game can also be explained by their cooperation levels. Players may choose to contribute a part of their endowment because they believe the other players are doing the same thing. In the case for donations, people may also exhibit cooperation by donating because the University solicits for donations (i.e. they are cooperating with what the University is asking for), or donating because they believe other seniors and/or alumni are also donating.

3.4   Participants and Sessions

Wesleyan University Relations provided a random sample of 2,004 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. 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.

The deadline for participation was mid-February 2018, and a total of 397 people completed the online study. Shortly after the participation deadline, all participants were randomly paired, ticket payoffs were calculated, and 10 winners were randomly selected. The winners were emailed in the beginning of March 2018 with instructions on how to receive their prizes.

4 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. Appendix B contains the variables definition table, and Appendix A contains all of the figures, correlation tables, and regression tables from this section.

4.1 Social Preference Games

Since the generalized dictator game consisted of 9 rounds with varying sets of endowments and prices of giving, similar to Andreoni and Miller and Fisman et al. (2007), I assumed each participant's giving preferences was a member of the constant elasticity of substitution (CES) utility function. The CES utility function is written as:

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

The first parameter, α, measures the relative weight on the payoff for self. Holding ρ constant, when α=1, $U_s = \pi_s$, i.e. the individual's utility depends only

on the amount they keep for themselves. When $\alpha=0$, $U_s = \pi_o$, i.e. the individual's utility depends only on the amount the other person receives. Therefore as α approaches 1, the individual exhibits selfish preferences, and as α approaches 0, the individual exhibits selfless preferences.

The second parameter, ρ, indicates the willingness to trade off payoffs to themselves and the other person in response to price changes. Holding α constant, $U_s = \alpha\pi_s + (1-\alpha)\pi_o$ when $\rho=1$, i.e. the individual's utility depends on the sum of the players' payoffs. This means the individual exhibits perfect substitutes preferences for giving: they will prefer to give their entire endowment to the other player when the price of giving is cheap (p < 1), and they will prefer to keep their entire endowment when the price of giving is expensive (p > 1). These preferences are also called efficiency-minded preferences. As $\rho \to -\infty$, $U_s = min(\alpha\pi_s, (1-\alpha)\pi_o)$. That is, the individual's utility equals the minimum payoff between both places, meaning the individual exhibits inequality-averse preferences. These preferences are called Leontief preferences: the individual prefers to split the endowment equally. Lastly, when $\rho = 0$, $U_s = A\pi_s^{\alpha}\pi_o^{1-\alpha}$. In this case, the individual has Cobb-Douglas preferences.[7]

Maximizing utility subject to the budget constraint ($p_s\pi_s + p_o\pi_o = m$) yields the CES demand function given by:

$$\pi_s(p, m) = \frac{[\alpha/(1-\alpha)]^{1/(1-\rho)}}{\rho^{-\rho/(\rho-1)}+[\alpha/(1-\alpha)]^{(1/(1-\rho))}}m$$

$$= \frac{A}{p^r+A}m$$

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

---

[7]See Arrow et al. (1961) on how Leontief preferences and Cobb-Douglas preferences are derived from the CES production function.

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

where i represents each participant, t represents each independent decision-problems in the generalized dictator game, and ϵ_i^t is assumed to be distributed normally with mean zero and variance σ_i^2. For each participant i, I used the 9 combinations of π_s, p, and m and generated the estimates $\hat{A}_i$ and $\hat{r}_i$ using non-linear least squares. From the estimates I retrieved each participant's efficiency level, $\hat{\rho}_i$, which then gave me each participant's sharing propensity, $\hat{\alpha}_i$.

Results for the proposers in the ultimatum game are also represented by the proposers's pass rate, with higher pass rates indicating higher altruism and/or fairness preferences. For example, if the proposer sent 6 tickets to the responder, the pass rate is 0.6. Results for the proposers in the trust game are also represented by the proposer's pass rate (the percentage of the initial endowment passed to the other responder). Higher pass rates indicate higher levels of altruism, fairness preferences, and/or trust. Similarly, outcomes for the players in the public goods games are represented by the proposer's pass rate into the public fund. Thus the larger the amount the player contributes into the public fund signals the higher levels of cooperation the player has.

Results from the responders in the ultimatum game are represented as the minimum pass rate the responder chooses to accept. Since responders were presented an ascending list of all possible amounts the proposer could choose to send, minimum pass rates were obtained using the switch point where the responders changed from rejecting an offer amount to accepting an offer amount. Lower minimum accepted pass rates indicate lower levels of selfishness and negative reciprocity, whereas higher minimum accepted pass rates indicate higher levels of selfishness and negative reciprocity.

The trust game asks for the responder's return amount given the 10 possible offer amounts from the proposer. For each responder, I regressed their return amount on the offer amount, and retrieved the estimated slope. Thus the results from the responders in the trust game are represented by the estimated slope, which measures the participant's reciprocity, or trustworthiness level. Values closer to one represent more reciprocal behavior, and lower levels closer to zero represent more selfish behavior.[8]

Figure 1 consists of two panels (Panels A and B) that shows the distribution of parameters, α and ρ, derived from responses in the generalized dictator game. The parameter estimates vary dramatically across subjects, implying that preferences for giving are very heterogeneous. Panel A displays a high peak of 21% at $\alpha=1$: a considerable amount of participants displayed extremely selfish preferences. There is a smaller peak of 12% at $\alpha=0.5$, and there are more α levels above 0.5 than below 0.5. These results show that participants tended to have more selfish preferences. 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 ($\rho \approx 1$): these 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, i.e. efficiency-minded preferences. A little over 20% of participants demonstrated Leontief preferences (ρ far below 0), i.e. they prefer splitting the endowment equally, and roughly 15% of subjects possess Cobb-Douglas preferences ($\rho \approx 0$). 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 ($-0.9 \leq \rho \leq 0.9$).

---

[8]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 2 consists of four panels (Panels C, D, E, and F) which shows the distribution of responses in the ultimatum, trust, and public goods games. Panel C reports a maximum of 42% of proposers in the ultimatum game who gave half of their endowment to the responders, and there is slightly more emphasis on those who gave 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 0.25-0.50. Correspondingly, Panel D shows the distribution of the minimum offers that the responder chose to accept. With the exception of one 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 the endowment. Surprisingly, most participants (27%) accepted an offer amount of 0 tickets.

Panel E shows the distribution of contributions the proposers made in the trust game. Pass rates were scattered all across the range from offering none of their endowment to offering all of their endowment. There were two maxima, both around 22%, at giving half the endowment or the entire endowment. Other contributions that have more than 10% are proposers who gave 0.3 or 0.4 of the given endowment. This is also broadly in line with typical findings that report average transfers of roughly half of the endowment. Panel F shows that almost half the participants in the public goods game sent their entire endowment into the public fund. The next most popular choice (20%) was to send half their endowment to the public pool, and only 4% of participants sent nothing into the public pool. Again, this broadly matches usual results in literature.

Figure 3 contains two panels (Panels G and H) that displays the distribution of reciprocity levels and average repayment rate, respectively, which are derived from responders' responses in the trust game. Panel G shows that, with the

25

exception of two participants who demonstrated negative reciprocity (perhaps they did not understand the game), reciprocity levels were very heterogenous between 0 and 1. There is a maximum (27%) at reciprocity levels around 0.5, and a local maximum (14%) around 0.35. There is also a small local maximum (10%) at 0, which represents those who did not show any reciprocity. Panel H shows a strong peak of repayment rates of around 0.45, and participants showed 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 half of the transfer.

Table 1 shows the pairwise correlations (Spearman's ρ) between the different game outcomes. A large majority of the correlations are statistically significant at the 5% level (12 out of 15). All of the statistically significant negative correlations involve the α levels from GDG and the behavior in UG2. The negative correlations for α reflect that participants who are more selfish ($\alpha \to 1$) are more likely to make smaller contributions in the other games, and the negative correlations for UG2 reflect that participants who accept smaller contributions are more likely to make larger contributions in the other game decisions. Otherwise, decisions in UG1, TG1, TG2, and PGG have positive correlations with one another. These results show that participants demonstrated consistent behavior in all the games.

4.2   Self-Reported Measures of Past Social Behaviors

Total SRA scores were obtained by summing across each participant's responses for the 10 items in the SRA Scale ("Never" = 0, "Once" = 1, "More than once" = 2, "Often" = 3, "Very often" = 4). A higher SRA score indicates higher pro-social behavior. Figure 4 shows the distribution of total scores. There is a wide variety in the total SRA scores obtained, ranging between 20 and 48.

Scores are centered around 33 and the shape is symmetric. Panel J displays the distribution of monetary SRA scores, which were obtained by summing across each participant's responses for only the items related to money (items 2, 3, 4, and 7 in Appendix B). A large majority of the scores are between 9 and 12.

Table 2 contains pairwise correlations (Spearman's ρ) between the game responses and the SRA scores. I included both total SRA scores ("SRA") and monetary SRA scores ("SRAmoney"). None of the game results are significantly correlated with total SRA scores, and only the results from the responders in the trust game and players in the public goods game are significantly correlated with the monetary SRA scores. Overall, there is a very weak relationship between the social preference games and self-reported measures.

4.3 Donations Behavior

Figure 5 shows the distribution of Wesleyan donations. Panel I shows that 64% of the participants have donated to Wesleyan University before, and 36% have not. Panel J shows the distribution of total donation amounts below $100, with donations above $100 aggregated into the rightmost bar. Most of the donation amounts are below $20 (about 36%). There are 30 participants who each donated a total amount greater than $100. All of these larger donations amounts are between $100 and $625, with the exception of two very large donations.

4.4 External Validity of Social Preference Games

4.4.1 Do the Games Explain Donations Behavior?

I now turn to the main question of whether the game decisions explain the field behavior. There are two ways that I represented donations behavior: (1) whether the participant has ever donated or not, and (2) the total amount the participant

has donated. I included the α and ρ parameters from the generalized dictator game, proposers' pass rates in the ultimatum and trust games ("ultimatum1" and "trust1"), responders' minimum accepted pass rates from the ultimatum game ("ultimatum2"), responders' reciprocity levels from the trust game ("trust2"), and players' pass rates in the public goods game ("cooperation") as the explanatory game variables. I also included total and monetary SRA scores ("SRAtotal" and "SRAmoney") as the explanatory self-reported variables. Appendix B contains the table of variable definitions.

Table 3 presents twelve logistic regression models using the binary donations representation as the response variable. The first seven columns present each of the game outcomes on their own, and the eighth column includes all of the game variables. The ninth and tenth columns contain SRAtotal and SRAmoney scores on their own, and the last two columns include all of the game outcomes along with the total and monetary SRA scores, respectively. None of the games or SRA results are statistically significant on their own. When both the games and SRA results are used in the regression model, still none of the variables are statistically significant.

I then used ordinary least squares (OLS) to regress the same explanatory variables on the log of the amount donated. Table 4 presents the twelve linear regression models, presented in the same format as Table 3. Similar to the logistic regression models, none of the game outcomes or SRA scores are statistically significant.

It is also useful to look at the R^2 statistic, which measures the proportion of variance explained by the regressors. For the logistic regressions, I calculated each model's McFadden's pseudo-R^2 values which are displayed at the bottom of the table[9]. The highest value belongs to the model that includes all of the game

---

[9] $R^2_{McFadden} = 1 - \frac{log(L_c)}{log(L_{null})}$ where L_c denotes the maximized likelihood for the current fitted model, and L_{null} denotes the maximized likelihood the model with no predictors.

outcomes and monetary SRA score. Interestingly, the pseudo-R^2 value for the model with all of the game variables is lower than the pseudo-R^2 values for the models that include only the SRA scores. This suggests that self-reported measures explain the field behavior better than even using multiple social preference games.

The R^2 values for the linear regression models are displayed at the bottom of Table 4. The values are very low, and again, the highest value belongs to the model that includes both the games and monetary SRA score. In contrast to the logistic regression models, this time the R^2 value for the model with all of the game variables is larger than the R^2 values for the models that includes only the total and monetary SRA scores.

Altogether, these results tell us that self-reported measures are a better tool than even using multiple social preference games when explaining whether people donate or not, and using multiple social preference games is better than using self-reported measures when explaining how much people donate. The best tool overall, though, is to use both social preference games and self-reported measures. Nevertheless, all of the models' R^2 values are very low – practically zero – so even the "best" model explains very little of the variability of donations behavior.

Tables 3 and 4 therefore provide evidence that the social preference games, even when combined with self-reported measures, have little to no power to explain donations behavior. However, this may be because the right combination of variables are not included in the models. I performed best subset selection on both the logistic and linear regression models[10]. Since best subset selection fits all possible models to find the best candidate, I included more explanatory variables:

---

[10]Best subset selection identifies all of the possible regression models derived from all of the possible combinations of the candidate predictors, and determines the model that does the best at meeting some well-defined criteria. In this case, the criteria I used is Mallows' C_p-statistic, which assesses assess fits when models with different numbers of parameters are being compared. See Mallows (1973).

I added the responders' average return outcome from the trust game ("average return"), and replaced SRAtotal and SRAmoney with each of the ten individual SRA items ("SRA1" - "SRA10"). I first performed best subset selection for the logistic model, where Table 5 displays the five best subset logistic regression models. Notably four SRA items (items 3, 4, 8, and 10) appear in all five of the models, where they are statistically significant at the 5% level. Each model's pseudo-R^2 values are considerably larger than the previous logistic models in Table 3 (almost 8 times larger, with the highest value belonging to the model that includes α and five SRA items (items 3, 4, 7, 8, and 10). However, the R^2 value is still pretty close to zero.

Table 6 reports the five best subset linear regression models. Two game variables (ultimatum1 and trust2) as well as four SRA items (items 3, 4, 5, and 6) appear regularly in all of the models. Ultimatum1 and trust2 are both statistically significant at the 5% level, SRA items 3 and 6 are statistically significant at the 10% level, and SRA items 4 and 5 are statistically significant at the 1% level. Again, all of the the R^2 statistics are significantly larger than the values in Table 4. The highest R^2 value belongs to the model that contains the mentioned variables along with cooperation and trust1, but the model still explains only about 17% of the variability in donations behavior.

Overall, self-reported measures seem to perform just as well as social preference games in explaining the field measure. This is seen in Table 5 and 6 where the same SRA items consistently appear in each of the models. However, each of the estimated models' still explain only a little of the variability of donations behavior. Moreover, some of the game variables in the models as seen in Tables 3-6 have signs that are not expected. First, it makes sense that α should be negative since higher levels of alpha indicate more selfish behavior. Ultimatum1, trust1, and cooperation should have positive relationships with donations since higher

pass rates signal more pro-social preferences. Ultimatum2 is expected to have a negative relationship with donations since higher minimum accepted pass rates signal higher negative reciprocity, and trust2 should have a positive relationship with donations. However, in the logistic models, α, ultimatum1, and ultimatum2 have opposite signs than expected, and in the linear models, α, trust1, and trust2 have opposite signs than expected. Therefore both the extremely low R^2 values and unexpected direction of game coefficients suggest that the external validity of social preference games is pretty weak.

4.4.2   Do the Games Predict Donations Behavior?

The poor results from the previous section lead to another interesting topic: perhaps the social preference games and self-reported measures can still be a useful tool to predict donations behavior. That is, if Wesleyan University wanted to predict donations from incoming alumni, they can possibly use social preference games and/or self-reported measures from older alumni to help anticipate donations behavior.

In predictive modeling, a common approach is to split the data into a training set and testing set. The training set is used to build and train the model. Once the model is ready, the model is tested on the testing set to determine its accuracy and performance. I used the data from the older alumni participants (those who graduated in 2013, 2014, 2015, and 2016) to create the training set, and used the data from the most recent alumni (those who graduated in 2017) to create the testing set[11]. Similar as before, I estimated models for both the binary and continuous representations of donations behavior. The models I estimated include: the logistic/linear regression model, the best-subset logistic/linear regression model, and the least absolute shrinkage and selection operator (lasso) model[12]. To com-

---

[11]I decided not to use current seniors as the testing setWhy not use current seniors.

[12]The least absolute shrinkage and selection operator (lasso) is a useful regression analysis

pare how well the models performed, I tested each of the models on the testing set and looked at the mean-squared prediction error (MSPE)[13], where the smallest MSPE value indicates the best predictive model.

First I estimated logistic regression models on the training set. Table 7 reports each of the models. Each of the estimated models were then tested on the testing set, and the bottom row reports the corresponding MSPE values. The lowest MSPE belongs to the model that includes only the monetary SRA score. Next I performed best subset selection. Table 8 reports the best five models that were chosen, and again, the MSPE values are displayed at the bottom. The model that includes α, ultimatum2, and a few of the SRA items (items 3, 4, and 10) has the lowest MSPE. Lastly I used the lasso method, and all of the variables except for α, ultimatum2, and two of the SRA items (items 3 and 4) were shrunk to zero. Table 9 shows the lasso penalized logistic model[14] and its corresponding MSPE value.

I repeated the same steps for the linear models using the log of donation amounts. Table 10 reports the linear regression models, where again the lowest MSPE corresponds to the model that includes only the monetary SRA score. I ran best subset selection, and Table 11 reports the top five models where the lowest MSPE belongs to the model that contains ρ, trust2, ultimatum1, and four of the SRA items (items 1, 4, 5, and 6) . Finally, the lasso method truncated most of

---

method that performs both variable selection and regularization in order to enhance prediction accuracy. The lasso estimate is defined by

$$\beta^{lasso} = \underset{\beta}{\operatorname{argmin}}\{\tfrac{1}{2}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|\}$$

where λ is a free parameter set to minimize the out of sample error. Certain coefficients are truncated to be set to zero to effectively choose a simpler model that does not include those coefficients. In other words, lasso regression works like a feature selector that picks out the most important coefficients, i.e. those that are most predictive (and have the lowest p-values). See Tibshirani (1996) for a full introduction on the lasso method.

[13]$MSPE = E[(g(x_i) - \hat{g}(x_i))^2]$, i.e. the expected value of the squared difference between the fitted values implied by the predictive model $\hat{g}$ and the values of the (unobservable) model g

[14]Statistical significance tests were based off of the proposed test by Lockhart et al. (2014). See Kyung et al. (2010) and Tibshirani for a full discussion on standard errors for lasso predictions.

the variables to zero, leaving out ρ, average return, cooperation, and three SRA items (items 4, 5, and 6). Table 12 shows the lasso penalized linear model with its MSPE value.

Below are tables that report each predictive model's MSPE values:

Logistic model	Best subset logistic model	Lasso penalized logistic model
donated = SRAmoney	donated = α + ultimatum2+ SRA3 + SRA4 + SRA10	donated = α + ultimatum2 + SRA3 + SRA4
0.24649	0.26193	0.25177

Linear model	Best subset linear model	Lasso penalized linear model
log(donations) = SRAmoney	log(donations) = ρ + trust2 + ultimatum1 + PGG + SRA1 + SRA4 + SRA5 + SRA6	log(donations) = ρ + avgreturn + cooperation + SRA4 + SRA5 + SRA6
6.98972	6.77793	5.25229

The first table shows that the model containing only the monetary SRA score is the most accurate model for predicting whether people will donate or not. The second table shows that the lasso penalized model containing three of the game outcomes and three SRA items is the best-performing model for predicting donation amounts. Something to note, however, is that the difference between the MSPE values in all of the estimated models is very small: whether all of the social preference games are used, only a couple games are used, or only the self-reported measures are used, the prediction performances are all similar.

5   Discussion and Conclusion

Currently, the accumulated literature exploring the external validity of social preference games has mixed results. My thesis provides a systematic approach to this topic, where I elicited decisions in four experimental social preference games along with self-reported social behaviors performed in the past. Most importantly, I used a natural field measure that is far-removed from my study.

The overall conclusion is that the social preference games do a poor job explaining the field behavior. In fact, the self-reported measures explain donations

behavior similarly, if not better, to the games. These poor results are still very interesting, however, because it suggests that social preference games may not generalize as much as it is anticipated to. Social preference games are thought to at least generally indicate individuals'pro-social behavior. With the results from my research, however, this doesn't seem to be the case.

Self-reported measures also seem to play an equal role with the games in predicting donations behavior. As seen in Tables 7-12, models that included all of the game variables, some of the game variables, only self-reported measures, or combinations of the two tools all reported similar MSPE values. It can even be argued that the self-reported measures are a more important tool. The same SRA items were consistently included in the best subsets and lasso-penalized models (items 4, 5, and 6 for the logistic models, and items 3 and 4 for the linear models).

Altogether, self-reported measures perform just as well (if not better) than social preference games in both explaining and predicting donations behavior. This conclusion has a potential policy implication from a research perspective. Experimental games require a lot of resources: not only is it expensive paying participants for attending the lab session and providing money aligned with the payoffs of the games in order to elicit honest actions, but it also takes a lot of time to program the experimental games. Therefore perhaps eliminating experimental games in favor of survey questions is more practical.

However, this is just the beginning of a systematic approach to uncovering the external validity of experimental social preference games. There are a few things that further research can include. First, both Galizzi and Navarro-Martinez and my study used university (recent) students who self-selected into the experiments. It would be beneficial to use a different participant pool, since our subjects could be inherently different than the general population[15]. Further research should

---

[15]See Levitt and List for a discussion on student participants.

34

also use more field measures. Galizzi and Navarro-Martinez created five field situations. However, their subjects were likely influenced by the experimenter demand effect. On the other hand, my study included only one field measure, but the measure was not influenced by the experimenter demand effect. Therefore further research should incorporate more field measures that can be theoretically mapped to various behavioral constructs, but also far removed from the study itself. Lastly, perhaps using other social preference games and/or exploring repeated games could provide further insight into the topic.

Finally, this research may 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 would be intriguing to adopt a systemic approach to explore whether in-lab behavior is correlated to field measures.

References

G. Charness and M. Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, Jan 2002.

E. Fehr and K. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, 1999.

J. Andreoni and J. Miller. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

M. Rabin. Incorporating fairness into game theory and economics: Comment. *American Economic Review*, 83:1281–1302, 1993.

R. Fisman, P. Jakiela, and S. Kariv. The distributional preferences of Americans. 2014.

M. Galizzi and D. Navarro-Martinez. On the external validity of social preference games: A systematic lab-field study. *Management Science*, 2017.

R. Forsythe, J. Horowitz, N. Savin, and M. Sefton. Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347–369, 1994.

J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 1995.

B. Hermann, C. Thoni, and S. Gachter. Anti-social punishment across societies. *Science*, 319:1362–1367, 2008.

E. Hoffman, K. Mccabe, K. Shachat, and V. Smith. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3): 346–380, 1994.

J. List. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, 114 (51):1–37, 2006.

U. Gneezy, E. Haruvy, and H. Yafe. The inefficiency of splitting the bill: A lesson in institutional design. *The Economic Journal*, 114(495):265–280, 2004.

T. Burnham, K. McCabe, and V. Smith. Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior and Organization*, 43(1):57–73, 2000.

L Ross and A. Ward. Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and Knowledge*, pages 103–135, 1996.

J. Carpenter, E. Verhoogen, and S. Burks. The effect of stakes in distribution experiments. *Economics Letters*, 86(3):393–398, 2005.

R. Slonim and A. Roth. Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica*, 66(3):569–596, 1998.

J. Parco, A. Rapoport, and W. Stein. Effects of financial incentives on the breakdown of mutual trust. *Psychological Science*, 13(3):292–297, 2002.

T. Cherry, P. Frykblom, and J. Shogren. Hardnose the dictator. *American Economic Review*, 92(4):1218–1221, 2002.

S. Levitt and J. List. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174, 2007.

N. Baran, P. Sapienza, and L. Zingales. Can we infer social preferences from the lab? Evidence from the trust game. *NBER Working Paper*, 15654, 2010.

A. Franzen and S. Pointner. The external validity of giving in the dictator game. *Experimental Economics*, 16(2):155–169, 2012.

F. Englmaier and G. Gebhardt. Free-riding in the lab and in the field. *SSRN Electronic Journal*, 2011.

E. Fehr and A. Leibbrandt. A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics*, 95(9-10):1144–1155, 2011.

D. Karlan. Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95(5):1688–1699, 2005.

T. Goeschl, S. Kettner, J. Lohse, and C. Schwieren. What do we learn from public good games about voluntary climate action? Evidence from an artefactual field experiment. *SSRN Electronic Journal*, 2015.

K. Hill and M. Gurven. Economic experiments to examine fairness and cooperation among the Ache Indians of Paraguay. *Foundations of Human Sociality*, pages 382–412, 2004.

M. Gurven and J. Winking. Collective action in action: Prosocial behavior in and out of the laboratory. *American Anthropologist*, 110(2):179–190, 2008.

M. Voors, T. Turley, A. Kontoleon, E. Bulte, and J. List. Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone. *Economics Letters*, 114(3):308–311, 2012.

J. Rushton, R. Chrisjohn, and G. Fekken. The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, 2:293–302, 1981.

M. Benz and S. Meier. Do people behave in experiments as in the field? Evidence from donations. *SSRN Electronic Journal*, 2006.

A. Falk, S. Meier, and C. Zehnder. Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, 11(4):839–852, 2013.

R. Fisman, S. Kariv, and D. Markovits. Individual preferences for giving. *American Economic Review*, 97(5):1858–1876, 2007.

K. Arrow, H. Chenery, B. Minhas, and R. Solow. Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics*, 43(3):225–250, 1961.

C. Mallows. Some comments on cp. *Technometrics*, 15(4):661–675, 1973.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

J. Lockhart, J. Taylor, R.J. Tibshirani, and R. Tibshirani. A significance test for the lasso. *Annals of Statistics*, 42(2):413–468, 2014.

M. Kyung, M. Ghosh, and G. Casella. Penalized regression, standard errors and bayesian lassos. *Bayesian Analysis*, 5(2):369–412, 2010.

Appendix A   Computerized Social Preference Games

Generalized Dictator Game (GDD)

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 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)**.

For example, say you are given 10 tickets, hold @ 1 ticket each and pass @ 2 tickets each. If you decided to pass 3 tickets, then you keep (10-3)*1 = 7 tickets and Player 2 receives 3*2 = 6 tickets.
Instead, if hold @ 2 tickets each and pass @ 1 ticket each, you keep (10-3)*2= 14 tickets, and Player 2 receives 3*1= 3 tickets.

Please choose how you would like to divide the tickets.

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

(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 send?

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

(15, 0)	(14, 1)	(13, 2)	(12, 3)	(11, 4)	(10, 5)	(9, 6)	(8, 7)	(7, 8)	(6, 9)	(5, 10)	(4, 11)	(3, 12)	(2, 13)	(1, 14)	(0, 15)
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 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

Ultimatum Game Player 1 (UG1)

You are Player 1.

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 many of the 10 tickets you will give to Player 2:

Next

Ultimatum Game Player 2 (UG2)

Now you are Player 2.

Player 1 was endowed with 10 tickets 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 tickets.**

For example, say Player 1 offers you 4 tickets (i.e. Player 1 keeps 6 tickets). If you agree to the split, then click "Accept". If you do not like the split, then you can click "Reject" and you both will get 0 tickets.

Now for each possible offer, please state whether you accept or reject.

	Reject	Accept
Player 1 offers you 0 tickets (they keep 10 tickets)	○	○
Player 1 offers you 1 ticket (they keep 9 tickets)	○	○
Player 1 offers you 2 tickets (they keep 8 tickets)	○	○
Player 1 offers you 3 tickets (they keep 7 tickets)	○	○
Player 1 offers you 4 tickets (they keep 6 tickets)	○	○
Player 1 offers you 5 tickets (they keep 5 tickets)	○	○
Player 1 offers you 6 tickets (they keep 4 tickets)	○	○
Player 1 offers you 7 tickets (they keep 3 tickets)	○	○
Player 1 offers you 8 tickets (they keep 2 tickets)	○	○
Player 1 offers you 9 tickets (they keep 1 tickets)	○	○
Player 1 offers you 10 tickets (they keep 0 tickets)	○	○

Next

Trust Game Player 1 (TG1)

You are Player 1.

You are endowed with 10 tickets. Please decide how many 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.**

For example, if you decide to offer 2 tickets, then Player 2 will receive 2*3 = 6 tickets. Player 2 will then decide if they would like to send back any of their 6 tickets to you.

Now please decide how many of the 10 tickets you would like to offer:

[   ]

Next

Trust Game Player 2 (TG2)

Now you are Player 2.

Player 1 was endowed with 10 tickets, 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 tickets back to Player 1.**

For example, if Player 1 gives you 3 tickets, then you receive 3*3 = 9 tickets. You can then decide whether you would like to send any of the 9 tickets back to player 1.

Now for each possible offer, please state how many tickets you would like to send back to Player 1.

	How many tickets you would like to return:
Player 1 passes 1 ticket, i.e. you receive 3 tickets	
Player 1 passes 2 tickets, i.e. you receive 6 tickets	
Player 1 passes 3 tickets, i.e. you receive 9 tickets	
Player 1 passes 4 tickets, i.e. you receive 12 tickets	
Player 1 passes 5 tickets, i.e. you receive 15 tickets	
Player 1 passes 6 tickets, i.e. you receive 18 tickets	
Player 1 passes 7 tickets, i.e. you receive 21 tickets	
Player 1 passes 8 tickets, i.e. you receive 24 tickets	
Player 1 passes 9 tickets, i.e. you receive 27 tickets	
Player 1 passes 10 tickets, i.e. you receive 30 tickets	

Next

Public Goods Game (PGG)

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.**

For example, if you contribute 6 tickets and Player 2 contributes 4 tickets, the total tickets contributed is (6+4)*2= 20 tickets. You and Player 2 will then both receive 20/2 = 10 tickets. Therefore you will earn your remaining tickets (10-6 = 4) plus the extra 10 tickets from the pool for a total of 14 tickets.

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

Next

Appendix B Self-Report Altruism (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 to a charity.*
5. I have done volunteer work for a charity/organization.
6. I have delayed an elevator/held door open for stranger(s).
7. I have pointed out a clerk's error (at a supermarket, restaurant) in undercharging me.*
8. 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).
9. I have offered my seat on a bus/train to a stranger who was standing.
10. I have helped an acquaintance with moving in/ moving out of their dorm/apartment/house.

Note: * indicates SRA item is related to money.

Appendix C   Variables Definition Table

Alpha	Player's α level derived from the generalized dictator game
Rho	Player's ρ level derived from the generalized dictator game
Ultimatum1	Player 1's pass rate in ultimatum game
Ultimatum2	Player 2's minimum pass rate accepted in ultimatum game
Trust1	Player 1's pass rate in trust game
Trust2	Player 2's reciprocity level in trust game
Cooperation	Player's pass rate into public pool in public goods game
SRAmoney	Player's monetary SRA score
SRAtotal	Player's total SRA score

Appendix D   Figures and Tables

Figure 1: Distribution of CES parameters from GDG
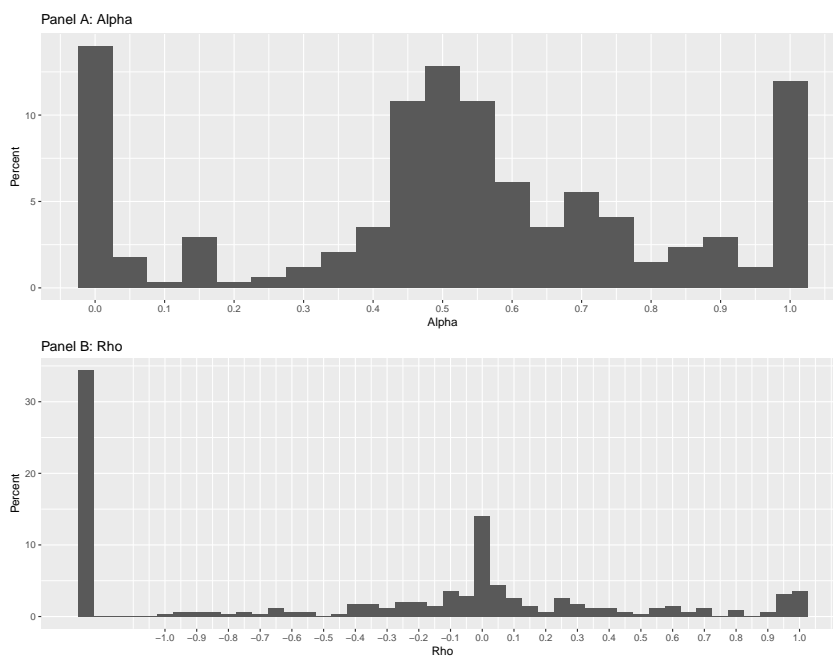
Figure 2: Distribution of responses in UG1, UG2, TG1, and PGG
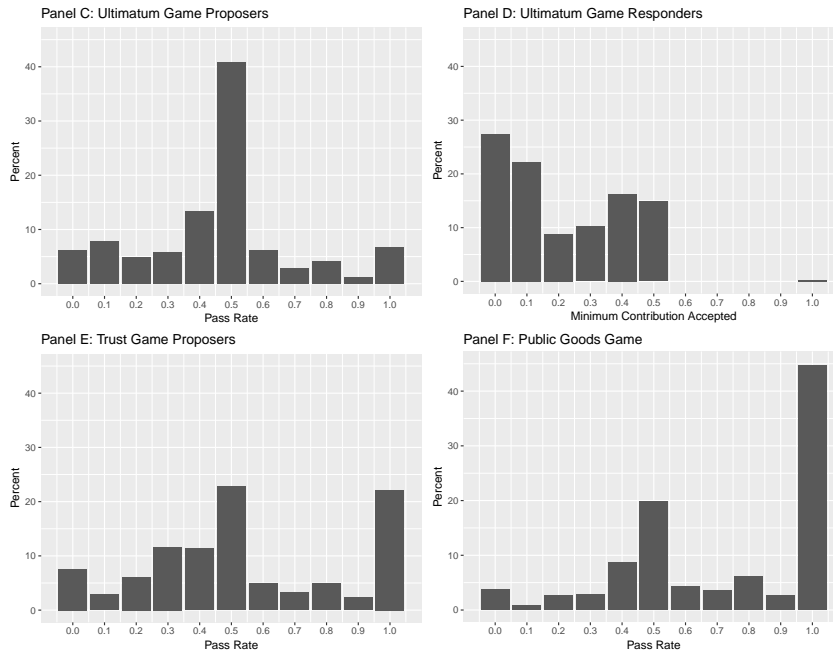
Figure 3: Distribution of responses from TG2
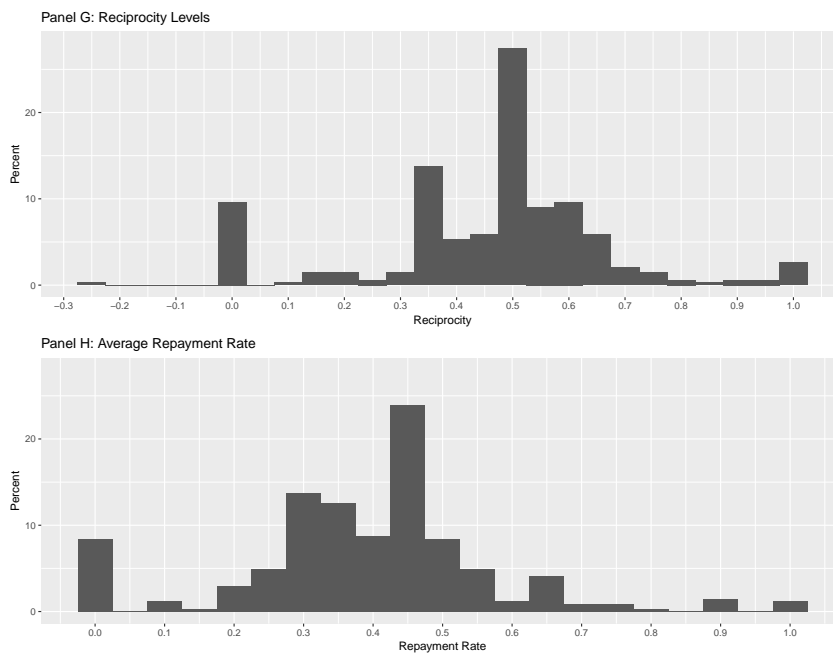
Figure 4: Total SRA scores
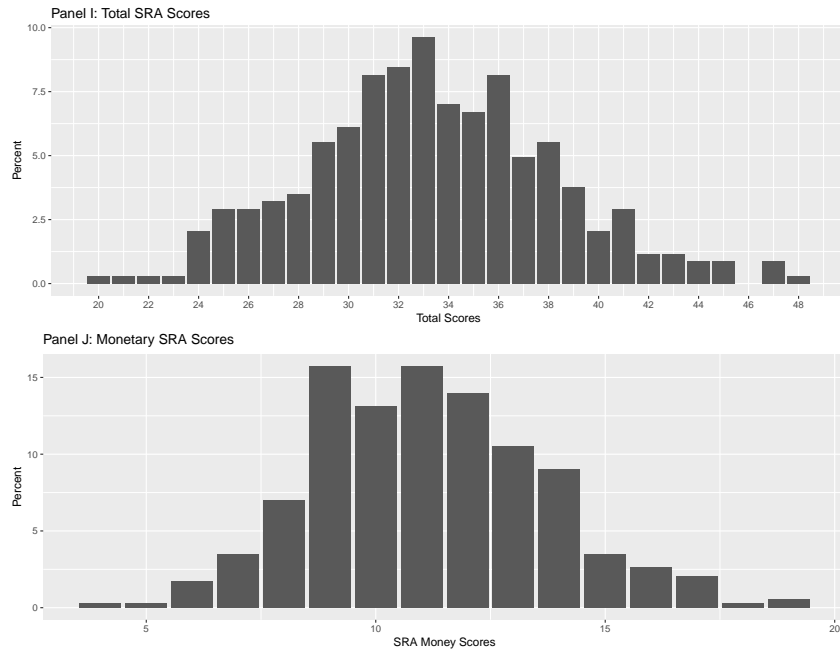
Figure 5: Donations behavior

Figure 6: Donations behavior by class year

Panel M: Average Donation Amount by Class Year

Panel N: Percent Donated by Class Year
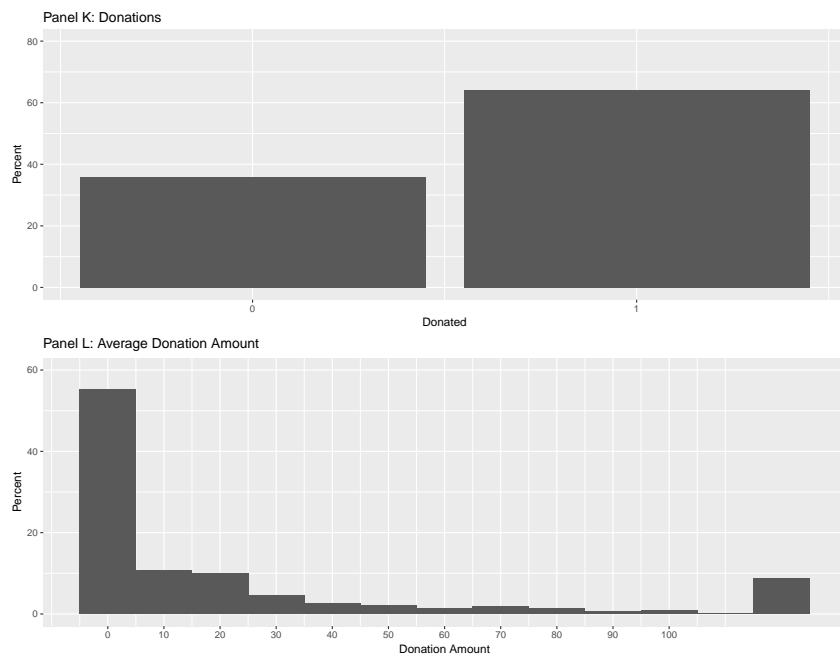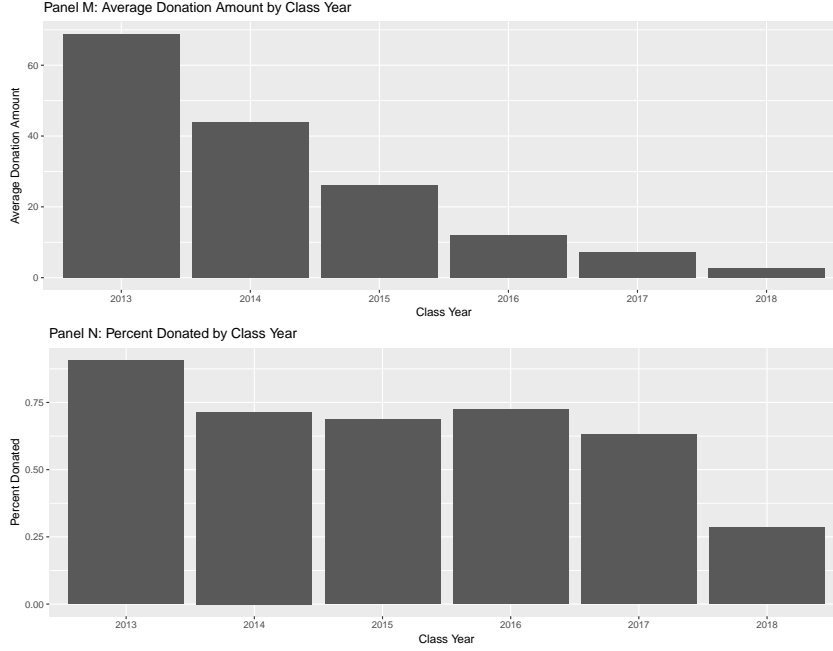
Table 1: Pairwise correlations between game results (Spearman's ρ)

	alpha	UG1	UG2	TG1	TG2	PGG
alpha	1.00	-0.28***	0.04	-0.26***	-0.30***	-0.16***
UG1	-0.28***	1.00	0.09	0.41***	0.31***	0.30***
UG2	0.04	0.09	1.00	-0.10*	-0.08	-0.16***
TG1	-0.26***	0.41***	-0.10*	1.00	0.45***	0.44***
TG2	-0.30***	0.31***	-0.08	0.45***	1.00	0.21***
PGG	-0.16***	0.30***	-0.16***	0.44***	0.21***	1.00
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 2: Correlations between game results and SRA scores (Spearman's ρ)

	SRAtotal	SRAmoney
alpha	-0.01	-0.04
UG1	0.02	0.07
UG2	-0.05	-0.05
TG1	0.07	0.08
TG2	0.04	0.11**
PGG	0.06	0.10*
SRAtotal	1.00	0.81***
SRAmoney	0.81***	1.00
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: Logistic regression models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Dependent variable:						
						Donated						
Alpha	0.052 (0.086)							0.102 (0.111)			0.095 (0.111)	0.100 (0.111)
Rho		−0.00000 (0.00001)						−0.00001 (0.00001)			−0.00000 (0.00001)	−0.00001 (0.00001)
UG1			0.016 (0.107)					−0.003 (0.125)			−0.001 (0.125)	−0.008 (0.125)
UG2				0.120 (0.137)				0.125 (0.142)			0.132 (0.142)	0.133 (0.142)
TG1					0.011 (0.083)			0.019 (0.108)			0.011 (0.108)	0.020 (0.108)
Reciprocity						0.025 (0.120)		0.046 (0.140)			0.043 (0.140)	0.030 (0.140)
PGG							0.024 (0.086)	0.035 (0.099)			0.031 (0.099)	0.024 (0.099)
SRAtotal									0.007 (0.005)		0.007 (0.005)	
SRAmoney										0.016 (0.010)		0.016 (0.010)
Constant	0.614*** (0.052)	0.640*** (0.028)	0.634*** (0.056)	0.616*** (0.039)	0.635*** (0.052)	0.630*** (0.061)	0.624*** (0.067)	0.499*** (0.127)	0.400** (0.176)	0.461*** (0.117)	0.270 (0.212)	0.333** (0.165)
Observations	343	343	343	343	343	343	343	343	343	343	343	343
Log Likelihood	−235.468	−235.635	−235.640	−235.264	−235.642	−235.629	−235.611	−234.711	−234.689	−234.392	−233.776	−233.449
Akaike Inf. Crit.	474.937	475.270	475.280	474.528	475.284	475.258	475.222	485.423	473.378	472.785	485.551	484.898
Psuedo-R^2	0.00078	0.000068	0.000048	0.00165	0.000039	0.000095	0.00017	0.00401	0.00410	0.00536	0.00799	0.00939

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Linear regression models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Dependent variable:						
						Log(Donations)						
Alpha	0.080 (0.383)							0.414 (0.492)			0.403 (0.493)	0.429 (0.491)
Rho		−0.00003 (0.00003)						−0.0001 (0.00004)			−0.00005 (0.00004)	−0.0001 (0.00004)
UG1			0.488 (0.454)					0.707 (0.506)			0.725 (0.506)	0.695 (0.504)
UG2				−0.270 (0.590)				−0.350 (0.598)			−0.325 (0.599)	−0.322 (0.597)
TG1					−0.210 (0.365)			−0.325 (0.464)			−0.358 (0.466)	−0.320 (0.463)
Reciprocity						−0.713 (0.516)		−0.787 (0.600)			−0.799 (0.600)	−0.851 (0.600)
PGG							0.635* (0.383)	0.685 (0.421)			0.691 (0.422)	0.652 (0.421)
SRAtotal									0.017 (0.022)		0.018 (0.023)	
SRAmoney										0.064 (0.043)		0.064 (0.043)
Constant	2.903*** (0.232)	2.896*** (0.121)	2.721*** (0.238)	3.005*** (0.171)	3.060*** (0.229)	3.272*** (0.262)	2.487*** (0.298)	2.447*** (0.565)	2.377*** (0.763)	2.221*** (0.502)	1.839** (0.932)	1.768** (0.729)
Observations	220	220	220	220	220	220	220	220	220	220	220	220
R^2	0.0002	0.005	0.005	0.001	0.002	0.009	0.012	0.043	0.003	0.010	0.046	0.053
Adjusted R^2	−0.004	0.001	0.001	−0.004	−0.003	0.004	0.008	0.011	−0.002	0.005	0.010	0.017
Residual Std. Error	1.668 (df = 218)	1.663 (df = 218)	1.663 (df = 218)	1.667 (df = 218)	1.667 (df = 218)	1.661 (df = 218)	1.657 (df = 218)	1.654 (df = 218)	1.666 (df = 218)	1.660 (df = 218)	1.656 (df = 218)	1.650 (df = 218)
F Statistic	0.044 (df = 1; 218)	1.149 (df = 1; 218)	1.155 (df = 1; 218)	0.210 (df = 1; 218)	0.331 (df = 1; 218)	1.907 (df = 1; 218)	2.749* (df = 1; 218)	1.363 (df = 1; 218)	0.568 (df = 1; 218)	2.190 (df = 1; 218)	1.275 (df = 1; 218)	1.469 (df = 1; 218)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Best-subset logistic regressions

	Dependent variable:				
	Donated				
	(1)	(2)	(3)	(4)	(5)
Alpha	0.090	0.097			
	(0.083)	(0.083)			
Ultimatum2			0.128		
			(0.133)		
SRA3	−0.097***	−0.101***	−0.094***	−0.096***	−0.100***
	(0.028)	(0.029)	(0.028)	(0.028)	(0.029)
SRA4	0.124***	0.120***	0.123***	0.123***	0.119***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
SRA7		0.028			0.026
		(0.021)			(0.021)
SRA8	0.077**	0.075**	0.077**	0.075**	0.073**
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
SRA10	−0.055**	−0.058**	−0.053**	−0.052**	−0.055**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Constant	0.340**	0.316*	0.349**	0.386**	0.367**
	(0.161)	(0.161)	(0.160)	(0.155)	(0.156)
Observations	343	343	343	343	343
Log Likelihood	−219.787	−218.865	−219.909	−220.385	−219.562
Akaike Inf. Crit.	451.574	451.731	451.817	450.770	451.124
Psuedo-R^2	0.06761	0.07154	0.06709	0.06506	0.06857

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Best-subset linear regressions

	Dependent variable:				
	Log(Donations)				
	(1)	(2)	(3)	(4)	(5)
Ultimatum1	1.035**	0.831*	0.992**	0.960**	0.958**
	(0.449)	(0.456)	(0.470)	(0.446)	(0.446)
Ultimatum2	−0.772				
	(0.557)				
Trust1			−0.590		
			(0.427)		
Trust2	−1.473***	−1.416***	−1.102**	−1.387***	−1.337***
	(0.513)	(0.510)	(0.557)	(0.511)	(0.511)
Cooperation		0.487	0.664*		
		(0.371)	(0.392)		
SRA1					−0.214
					(0.162)
SRA3	−0.235*	−0.228*	−0.237*	−0.220*	−0.192
	(0.122)	(0.122)	(0.121)	(0.122)	(0.123)
SRA4	0.758***	0.732***	0.734***	0.744***	0.755***
	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)
SRA5	−0.433***	−0.404***	−0.408***	−0.414***	−0.383***
	(0.131)	(0.131)	(0.130)	(0.131)	(0.133)
SRA6	0.260*	0.258*	0.281*	0.254*	0.300**
	(0.147)	(0.147)	(0.148)	(0.147)	(0.151)
Constant	1.805**	1.325	1.233	1.592*	1.860**
	(0.850)	(0.861)	(0.861)	(0.838)	(0.861)
Observations	220	220	220	220	220
R^2	0.164	0.164	0.171	0.157	0.164
Adjusted R^2	0.137	0.136	0.140	0.133	0.136
Residual Std. Error	1.546	1.547	1.543	1.549	1.547
	(df = 212)	(df = 212)	(df = 211)	(df = 213)	(df = 212)
F Statistic	5.959***	5.925***	5.445***	6.604***	5.927***
	(df = 7; 212)	(df = 7; 212)	(df = 8; 211)	(df = 6; 213)	(df = 7; 212)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Logistic predictive models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Dependent variable:						
						Donated						
Alpha	0.212** (0.094)							0.214* (0.119)			0.218* (0.119)	0.222* (0.119)
Rho		0.00001 (0.00001)						0.00000 (0.00001)			0.00000 (0.00001)	0.00000 (0.00001)
Ultimatum1			−0.049 (0.121)					−0.054 (0.138)			−0.054 (0.138)	−0.059 (0.138)
Ultimatum2				0.276* (0.150)				0.291* (0.154)			0.301* (0.154)	0.308** (0.154)
Trust1					−0.033 (0.093)			−0.008 (0.116)			−0.016 (0.116)	−0.005 (0.116)
Trust2						0.028 (0.129)		0.160 (0.151)			0.172 (0.152)	0.161 (0.151)
Cooperation							−0.037 (0.099)	0.005 (0.109)			0.001 (0.109)	−0.007 (0.109)
SRAtotal									0.003 (0.006)		0.005 (0.006)	
SRAmoney										0.011 (0.012)		0.016 (0.012)
Constant	0.650*** (0.056)	0.776*** (0.032)	0.783*** (0.064)	0.699*** (0.044)	0.777*** (0.058)	0.746*** (0.066)	0.786*** (0.077)	0.543*** (0.138)	0.662*** (0.199)	0.635*** (0.140)	0.359 (0.243)	0.367* (0.194)
Observations	216	216	216	216	216	216	216	216	216	216	216	216
Log Likelihood	−121.414	−123.035	−123.866	−122.244	−123.885	−123.927	−123.879	−119.047	−123.827	−123.536	−118.605	−118.183
Akaike Inf. Crit.	246.828	250.069	251.733	248.488	251.770	251.853	251.758	254.095	251.655	251.072	255.211	254.365
MSPE	0.26410	0.25548	0.24911	0.24984	0.24999	0.24837	0.24903	0.26641	0.24726	0.24649	0.26513	0.26493

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Best subsets logistic predictive models

	Dependent variable:				
	Donated				
	(1)	(2)	(3)	(4)	(5)
Alpha	0.283***	0.281***	0.244***	0.246***	0.249***
	(0.093)	(0.094)	(0.090)	(0.089)	(0.090)
Average Return	0.203	0.177			
	(0.151)	(0.151)			
Ultimatum2	0.225		0.208	0.203	
	(0.144)		(0.143)	(0.143)	
SRA3	−0.086***	−0.094***	−0.091***	−0.091***	−0.098***
	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)
SRA4	0.137***	0.140***	0.143***	0.139***	0.141***
	(0.032)	(0.032)	(0.033)	(0.032)	(0.032)
SR10			−0.022		
			(0.025)		
Constant	0.284*	0.361**	0.450***	0.399***	0.455***
	(0.170)	(0.163)	(0.158)	(0.147)	(0.142)
Observations	216	216	216	216	216
Log Likelihood	−107.722	−108.970	−108.264	−108.649	−109.672
Akaike Inf. Crit.	227.444	227.940	228.528	227.298	227.345
MSPE	0.26557	0.26473	0.26193	0.26468	0.26419

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Lasso penalized logistic predictive model

	Dependent variable:
	Donated
Alpha	0.13728
Ultimatum2	0.07993
SRA3	−0.05163*
SRA4	0.09403***
Constant	0.51105
Observations	216
MSPE	0.25177
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Linear predictive models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Dependent variable: Log(Donations)						
Alpha	−0.279 (0.401)							0.170 (0.488)			0.192 (0.488)	0.250 (0.487)
Rho		−0.0001* (0.00003)						−0.0001* (0.00004)			−0.0001 (0.00004)	−0.0001* (0.00004)
Ultimatum1			0.603 (0.476)					0.486 (0.523)			0.487 (0.523)	0.456 (0.520)
Ultimatum2				0.134 (0.608)				0.017 (0.612)			0.083 (0.616)	0.128 (0.612)
Trust1					0.195 (0.388)			0.120 (0.468)			0.070 (0.471)	0.121 (0.465)
Trust2						−0.526 (0.528)		−0.978 (0.607)			−0.939 (0.608)	−0.979 (0.603)
Cooperation							1.015** (0.402)	0.964** (0.429)			0.974** (0.429)	0.925** (0.427)
SRAtotal									0.030 (0.024)		0.024 (0.024)	
SRAmoney										0.093* (0.048)		0.083* (0.048)
Constant	3.426*** (0.248)	3.194*** (0.126)	2.991*** (0.253)	3.244*** (0.184)	3.169*** (0.241)	3.516*** (0.270)	2.542*** (0.313)	2.544*** (0.581)	2.258*** (0.822)	2.203*** (0.563)	1.712* (1.022)	1.569* (0.810)
Observations	164	164	164	164	164	164	164	164	164	164	164	164
R²	0.003	0.019	0.010	0.0003	0.002	0.006	0.038	0.076	0.010	0.023	0.082	0.093
Adjusted R²	−0.003	0.013	0.004	−0.006	−0.005	−0.00005	0.032	0.035	0.003	0.017	0.034	0.047
Residual Std. Error	1.513 (df = 162)	1.501 (df = 162)	1.508 (df = 162)	1.515 (df = 162)	1.514 (df = 162)	1.510 (df = 162)	1.486 (df = 162)	1.484 (df = 156)	1.508 (df = 162)	1.498 (df = 162)	1.484 (df = 155)	1.475 (df = 155)
F Statistic	0.484 (df = 1; 162)	3.073* (df = 1; 162)	1.608 (df = 1; 162)	0.049 (df = 1; 162)	0.254 (df = 1; 162)	0.992 (df = 1; 162)	6.380** (df = 1; 162)	1.834* (df = 7; 156)	1.559 (df = 1; 162)	3.790* (df = 1; 162)	1.727* (df = 8; 155)	1.994* (df = 8; 155)
MSPE	7.29724	7.31552	7.21467	7.20326	7.24405	7.26221	7.47346	7.63702	7.2354	6.98972	7.60423	7.36116
Note:											*p<0.1; **p<0.05; ***p<0.01	

Table 11: Best subset linear predictive models

	Dependent variable:				
	Log(Donations)				
	(1)	(2)	(3)	(4)	(5)
rho	-0.0001^{*}	-0.0001^{*}	-0.0001^{*}	-0.0001^{*}	-0.0001^{*}
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
TG2	-1.165^{**}	-1.157^{**}	-1.006^{**}	-0.988^{**}	-0.999^{**}
	(0.508)	(0.509)	(0.494)	(0.494)	(0.495)
UG1	0.598	0.596			
	(0.465)	(0.466)			
PGG	0.666^{*}	0.684^{*}	0.765^{**}	0.744^{*}	0.783^{**}
	(0.385)	(0.385)	(0.378)	(0.379)	(0.378)
SRA1	-0.211		-0.210		
	(0.173)		(0.173)		
SRA2				-0.137	
				(0.118)	
SRA4	0.640^{***}	0.626^{***}	0.645^{***}	0.667^{***}	0.631^{***}
	(0.147)	(0.147)	(0.147)	(0.150)	(0.147)
SRA5	-0.356^{***}	-0.386^{***}	-0.353^{***}	-0.361^{***}	-0.383^{***}
	(0.131)	(0.129)	(0.132)	(0.131)	(0.130)
SRA6	0.384^{**}	0.333^{**}	0.360^{**}	0.333^{**}	0.310^{**}
	(0.156)	(0.151)	(0.155)	(0.151)	(0.150)
Constant	1.021	0.669	1.227	0.930	0.875
	(0.920)	(0.875)	(0.908)	(0.862)	(0.862)
Observations	164	164	164	164	164
R^2	0.217	0.210	0.209	0.208	0.201
Adjusted R^2	0.177	0.174	0.173	0.173	0.171
Residual Std. Error	1.370 (df = 155)	1.373 (df = 156)	1.373 (df = 156)	1.374 (df = 156)	1.375 (df = 157)
F Statistic	5.371^{***} (df = 8; 155)	5.907^{***} (df = 7; 156)	5.878^{***} (df = 7; 156)	5.857^{***} (df = 7; 156)	6.593^{***} (df = 6; 157)
MSPE	6.77793	6.80894	6.82706	6.86435	6.86567

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Lasso penalized linear predictive model

	Dependent variable:
	Log(Donations)
Rho	−0.000015
Average Return	−0.38387
Cooperation	0.42147
SRA4	0.38572
SRA5	−0.12091
SRA6	0.15950
Constant	1.47608
Observations	164
MSPE	5.25229
Note:	*p<0.1; **p<0.05; ***p<0.01