External Validity of Experimental Social Preference Games

Sarah Xu

16 December 2017

1 Introduction and Literature Review

The standard economic model assumes actions are motivated purely by self-interest. However, it's clear from simple observation that people actually care about the wellbeing of others: social programs exist, people volunteer, and people donate. 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 social preferences, with numerous research studies providing evidence of social preferences. Papers have identified different 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, and receive monetary incentives that are aligned with the payoffs of the games. The experimental games allow researchers to target the different aspects of social behavior, and the laboratory setting strips the games from contextual features. Below are common games used to study social preferences:

- Dictator Game: Two-player game where the first player (the dictator) receives an endowment, m. The dictator chooses how to split the amount between themselves and the 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.

- Ultimatum Game: Two-player game where two players bargain over a fixed endowment, m. The first mover divides the sum, so that $\pi_s + \pi_o = m$. The second mover 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 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.

- 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.

There is an abundance of literature that study social preferences using experimental games. For example, Fehr and Schmidt (1999) used ultimatum and dictator games to identify whether people have inequality aversion, Berg et al (1995) employed trust games to provide evidence of reciprocity, and Hermann, Thoni, and Gachter (2008) used public goods game to provide evidence of punishment and reciprocity. This methodology has become one of the building blocks of experimental and behavioral economics.

However, lab experiments also have its disadvantages. 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 lab setting. One issue is that in a lab 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) under a normal dictator game. However, the authors then implemented a "double-blind" treatment where both experimenter and other subjects couldn't observe the dictator's actions, and 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 (List, 2006; Gneezy et al., 2004). People are also influenced heavily on the context of the situation. Ross and Ward (1996) found that simply calling a prisoner's dilemma game 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 (Burnham, McCabe, and Smith, 2000; and Gintis, 2001). Lastly, studies 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, though, 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 that can lead to deviations in behavior between laboratory and field setting.

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. Baran et al (2010) compared MBA alumni donations to their university with their reciprocity behavior when playing a trust game in the lab. They find that that the in-lab behavior predicts university donations. Franzen and Pointner (2013) compare behaviors from university students participating in standard dictator games and their actions when receiving a misdirected letter containing money. The authors find 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 with university students, comparing free riding behavior at the university library with free riding behavior when participating in a public goods experiment. The authors find statistically significant correlation between field and lab measures. There are also studies using non-student subjects. Fehr and Leibbrandt (2011) conducted public goods games with Brazilian fishermen, and find 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 behaviors to their repayment behavior of micro loans, and finds correlation between game results and repayment behavior.

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) used Tsimane forager-horticulturalists in Bolivia as their subjects, 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. This is clear evidence that justifies further research is needed.

While the literature on comparing the behavior of a single individual in both the lab and field, Galizzi and Navarro-Martinez (2017) argue that the 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 games to the field measure, and so it is crucial to have a more systematic approach. The authors proceed to carry out their own study, comparing behavior in various games with behaviors elicited in various field situa-

tions. Participants answered a questions about social behaviors exhibited in the past, played various f social preference games, and encountered naturalistic field situations related to social preferences. Examples of 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 note, however, that more systematic studies are needed to draw a definite conclusion. For example, future studies could look at other field situations or other game structures.

It is critical to think about why some experiments found correlations and some did not, in order to inform the design of our 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.

Similar to Galizzi and Navarro-Martinez (2017), I will recruit Wesleyan University seniors and recent alumni to play various experimental games and non-incentivized social preference questions. The different social preference games include the generalized dictator game, ultimatum game, trust game, and public good game. All the games tap into different types of pro-social behaviors related to giving money and helping others.

I will use Wesleyan University donations behavior as the field measure. A major criticism of Galizzi and Navarro-Martinez's design is how the authors executed their field measures: since the setup occurs 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. Therefore it is better to use a natural, far-removed situation. For my paper, I will compare Wesleyan University seniors and recent alumni's game behavior to their donations behavior.

WHY USE DONATIONS? USED IN A LOT OF PAPERS - AND Donors can choose to target where their donations will go towards, for example financial aid, academic programming and faculty support, or student activities. These different donation options can be related to some behavioral constructs. For example, donating may represent altruism and reciprocity, and donating to financial aid may represent inequality aversion.

The relationship between the games and field situation is furthered discussed in the next section. The self-report measures of past social behaviors provide an additional layer to evaluate the explanatory ability of the games.

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

Ultimately, my paper will add to the growing literature on the external validity of experimental social preference games. By comparing my results with those of past studies, I can contribute to the discussion on whether experimental games are a useful tool to examine social preferences. In addition, my paper can perhaps trigger interest into future studies on the external validity of using lab experiments for other behavioral economics topics.

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

2 Methods

The goal of this study is to evaluate the external validity of experimental social preference games to behaviors in the field, and if perhaps another tool (in this case, survey questions) or both tools can better predict social preferences exhibited in the field. The participants were presented with two sets of tasks: (1) incentivized social preference games, and (2) survey questions regarding past social behavior. The field measure used to compare to participants'lab results is their Wesleyan donations.

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 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.

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. At the end of the participation deadline (February 2018), all participants were randomly paired, and ticket payoffs were calculated. 10 lottery tickets were then chosen at random, and each ticket owner won $100. I will explain the payment process in further detail in Section 2.4.1.

Those who chose to participate in the study received a unique identifier, as well as detailed instructions about the tasks they will complete in the session. They were also informed that upon completion of the study, their major, class year, and Wesleyan donations information would be released to me by University Relations.

The entire study was computerized, and they were programmed and implemented using Qualtrics. Images of the computerized games and questions are in Figure 1.

2.1 Incentivized social preference games

Participants first played various social preference games. I used four games that are widely used in behavioral economics to study social preferences: generalized dictator game, ultimatum game, trust game, and public goods game. Since the ultimatum and trust games both require player 1 and player 2 strategies, For the ultimatum and trust games, each participant will have the opportunity to play the role of the both first mover and second mover.

Participants received detailed instructions for each game. All instructions given included examples to illustrate how each game works. Below are the different games each participant played:

- Generalized Dictator Game (GDG): Each participant plays as 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. The design is based off of Andreoni and Miller (2002).

- Ultimatum Game:

  - Player 1 (UG1): 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 Player 2. They are told that Player 2 may or may not reject the allocation. If the allocation is rejected, neither player receives any tickets.

  - Player 2 (UG2): Each participant is now the responder. They are given a list from 0 tickets to 10 tickets (in 1-ticketincrements), and are asked whether they accept or reject 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 Player 2. The amount sent over is multiplied by three. Player 2 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 Player 1 could have chosen to send. For each amount, they are prompted to enter the number of tickets they would like to return.
- Public Goods Game (PGG): Each participant is endowed with 10 tickets and have to 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.

These experimental games address many of the main behavioral constructs that explain social preferences. Some constructs include: altruism (dictator game, public goods game); positive reciprocity (trust game responder); negative reciprocity (ultimatum game responder); trust (trust game proposer); cooperation (public goods game); and inequality aversion (all games).

Since the participants are completing the games at their own availability, they will not be randomly matched with other participants. Instead, if the game requires two players, the second player will see a list of all possible options, and are prompted for their choice for each option listed (further discussion about the strategy method is in Section 2.2.1).

2.1.1 Strategy Method

Each participant has the opportunity to be both proposer (Player 1) and responder (Player 2) in each game. When the participant is the responder, the strategy method is used: the participant has to make decisions for all possible situations. For example, as the responder in the trust game, the participant has to decide how much money they will send back for all possible amounts that were given to them by the proposer.

There are several reasons why the strategy method is employed. First, as mentioned previously, participants complete the experimental games in their own time, so participants will not be paired up while they are playing. Instead, they will be randomly paired at a later period. Therefore it is useful to have all of Player 2's potential choices so that payoffs can be determined. Most importantly, the strategy method gives more information. Using the strategy method will provide all returned amounts for all possible donated amounts in the ultimatum game. In the trust game, the minimum donated amount that the responder will accept would be provided.

2.1.2 Payment

For each game, participants will be randomly paired, and payoffs will be determined. For each pair in the ultimatum game and trust game, one participant will be randomly assigned to their Player 1 choices, and the other participant will be assigned to their Player 2 choices. If the generalized dictator game is chosen, one player will be randomly assigned as the dictator, and the other player will be the receiver. However many tickets each participant earns in the end will correspond with the number of tickets they will have in the lottery. There will be ten lottery winners, and each winner will receive $100. Winners will be contacted through email.

Below are descriptions on how payoffs will be determined for each game:

- Generalized dictator game: Participants will be paired up randomly. In each pair, one player will be randomly selected as the dictator, and the other will be the receiver. Since the generalized dictator game consists 11 different sets of endowment and prices of giving, one set will be randomly picked for payoffs.

- Ultimatum game: Participants will be randomly matched. For each pair, one participant will be randomly assigned to their choice when playing Player 1, and the other participant will be assigned to their choices when playing as Player 2. For example, if Player 1 chooses to give 4 tickets but Player 2 chooses not to accept that amount, then both players will receive nothing. But if Player 1 does accept that amount, then Player 1 will receive 6 tickets and Player 2 will receive 4 tickets.

- Trust game: Participants will be randomly matched. For each pair, one participant will be randomly assigned to their choice when playing Player 1, and the other participant will be assigned to their choices when playing as Player 2, and payoffs will be determined. For example, if Player 1 chose to give 2 tickets to Player 2, then the multiplied amount is 6 tickets. Given being offered 6 tickets, if Player 2 chose to give back 1 ticket, then ultimately Player 1 will earn 9 tickets and Player 2 will earn 5 tickets.

- Public goods game: Participants will be randomly paired up. The overall tickets in the public fund will be multiplied by 2 and divided evenly. Thus each participant will receive their remaining tickets plus the divided amount from the public fund.

2.2 Self-reported measures of past social behaviors

Participants then reported on social behaviors exhibited in the past. The questions are loosely based off of 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 one each item in the past. Participants rate each statement on a scale from 1 ("never") to 5 ("very 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

Each participant's Wesleyan donations information will be accessed upon their informed consent and completion of the survey and games.

Wesleyan donations can represent different aspects of social preferences. Whether a participant donates or not can represent altruism. Donating can also represent reciprocity, since seniors/recent alumni are giving back to the university that provided them four years of education and opportunities, as well as represent cooperation and trust. In addition, donors are able to target what area their donations can go towards. For example, donating towards Financial Aid can represent inequality aversion, since Financial Aid makes the Wesleyan experience possible for talented, low-income students who might not otherwise be able to attend.

A main feature of my design is that the donations information is a naturalistic environment that is far-removed from the experiment, which provides a more accurate field measure. 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. I believe that it is not too difficult to link the two situations together. In my design, participants' donations behavior is not affected by their participation in the lab at the time.

I chose to use the generalized dictator game, ultimatum game, trust game, and public goods game because each game can be mapped to donations behavior. First, the very nature of donating represents altruism - therefore, the generalized dictator game and proposers for the ultimatum and trust games are expected to be correlated with donating. In addition, inequality aversion can also explain why seniors and alumni donate: they want future students, perhaps especially those on financial aid, to be provided with the best opportunities. Dictators in the generalized dictator game, proposers in the ultimatum and trust games, and players in the public goods game can also represent inequality aversion. Donating to one's university may also represent reciprocity, which is represented by the responders in the ultimatum and trust games. Lastly, the public goods game can also be mapped to donating because it represents cooperation: seniors and alumni will give if they believe that others are donating as well.

2.4 Participants and Sessions

Email invitations were sent in early January 2018, and the deadline for participation was mid-February 2018. A total of 397 people completed the online study. The participants were volunteers who open the email 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. Wesleyan University Relations provided me a random sample of 2,004 Wesleyan students and recent alumni (334 emails for each class year from 2013 to 2018). Besides soliciting seniors and recent alumni to participate, I used no other eligibility or exclusion criteria to select participants.

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 donation behavior. 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 different sets of endowments and prices of giving, similar to 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 to self (i.e. propensity to share, or equity-mindedness), and ρ indicates the willingness to trade off payoffs to self and other in response to price changes (i.e. utilitarian preferences, or efficiency-mindedness). Using the Lagrange function maximize utility subject to a budget constraint ($p_s\pi_s + p_o\pi_o = m$), the CES demand function is given by:

$$\pi_s(p_s, p_o, m) = [\tfrac{g}{(p_o/p_s)^r + g}]\tfrac{m}{p_s}$$

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

$$\frac{p_{s,i}\pi^t_{s,i}}{m^t_i} = \frac{g_i}{(p^t_{o,i}/p^t_{s,i})^{r_i}+g_i} + \epsilon^t_i$$

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

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.

The 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. Therefore the outcome for TG is represented by the participant's reciprocity, or trustworthiness level. Reciprocity levels closer to 1 represent more reciprocal behavior, and lower levels closer to 0 represent selfishness. 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, a large majority follow a linear slope.

Figure 2a consists of 4 panels (Panels A, B, C, and D), which shows the distribution of responses in UG1, UG2, TG1, and PGG, respectively. Panel A shows that people give a wide range of contribution amounts in UG1 (from giving 0 tickets to giving their entire endowment, with slightly more emphasis on giving smaller contributions). There is a high 42% spike at giving half of their endowment to Player 2, which typically interpreted as an anticipation of negative reciprocity in Player 2 if offered an amount smaller than half the given endowment. Correspondingly, Panel B shows the distribution of the minimum offer that Player 2 chooses to accept in UG2. With the exception of 1 participant, everyone accepted an amount less than or equal to half of Player 1's endowment. Surprisingly, most participants accepted an offer amount of 0 or 1 tickets. Panel C shows contributions scattered all across the range from offering 0% to 100% of their given endowment in TG1. The highest bars, with each having around 22% of the participants choosing these contributions, are at giving 50% of their endowment or giving their entire endowment. Other contributions that have more than 10% participants are giving 30% or 40% of the given endowment. Contributions above 0% in PGG are generally interpreted as the player's trust in Player 2. Panel D shows that a majority of participants (45%) send their entire endowment into the public pool. The next most popular choice (20%) is to send half their endowment to the public pool. Only 4% of participants sent 0% into the public pool, which shows that a large majority of participants have cooperative behavior.

Figure 2b, consisting of 3 panels (Panels E, F, and G), displays the the distribution of parameters from responses in GDG (α and ρ) and TG2 (reciprocity).

Panel E and Panel F

Panel G displays the distribution of reciprocity levels from TG2. With the exception of two participants who showed negative reciprocity (perhaps they did not understand the game), everybody else had reciprocity levels between 0 and 1. There is a high spike (27%) of reciprocity levels around 0.5, a lower spike (14%) around 0.35, and a small spike (10%) at 0, which suggests that while people do exhibit reciprocity, levels are on the lower end.

3.2 Self-Report Measures of Past Social Behaviors

Figure 3 shows the distribution of total scores obtained by the participants on the SRA Scale. A higher SRA score means the participant is more altruistic. As Figure 2 shows, there was a wide variety in the total SRA scores obtained (scores ranging between 20 and 48). Scores are centered around the center (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.

3.4 External Validity of Social Preference Games

4 Analysis

After gathering all experimental, survey, and donations data, I will be running two regression models:

$$donations = \beta_0 + \beta_1 GDG_{equity} + \beta_2 GDG_{efficiency} + \beta_3 UG1 + \beta_4 UG2 + \beta_5 TG1 + \beta_6 TG2 + \beta_7 PGG \quad (1)$$

$$donations = \beta_0 + \beta_1 SRA \quad (2)$$

It makes sense that all βs, except for the β for UG2, will be positive. If the participant has a high sharing propensity, i.e. they place a high weight on giving, and a high efficiency level, i.e. they care a lot about total social welfare, then donations should be larger. As the participant's pass rate for the ultimatum and trust game (coefficients for UG1 and TG1) increases, it indicates they have higher levels altruism, inequality

aversion, or both, so donations should also increase. If the participant has higher levels of reciprocity, indicated by the coefficient on TG2, then donations should increase - the very nature of giving back to one's university is represented by reciprocity, so there should be a high, positive correlation between TG2 and measured donations. A higher pass rate into the public goods fund indicates higher levels of cooperation and altruism, so the coefficient on PGG should also be positive. Lastly, if the minimum pass rate accepted by the responder in the ultimatum game is high, this indicates that the responder is selfish - they will punish the proposer if the offer is not to their liking. Therefore it makes sense that the coefficient on UG2 is negative.

First, I will run each model and examine how well each instrument predicts donations. With Model (1), I want to see which game best predicts donations behavior. For example, if I see that the dictator game and responder ultimatum game is statistically significant but all other games are not statistically significant, then perhaps further research could exclude using the other games when predicting charitable giving. This would certainly save researchers the time cost in programming the games, as well as make the experiment shorter (and therefore more attractive to participants).

However, the overall goal is to determine which instrument is better at predicting the field measure. Along with Model (1) and Model (2), I will also be running:

$$donations = \beta_0 + \beta_1 GDG_{equity} + \beta_2 GDG_{efficiency} + \beta_3 UG1 + \beta_4 UG2 + \beta_5 TG1 + \beta_6 TG2 + \beta_7 PGG + \beta_8 SRA \tag{3}$$

I will examine each model's prediction errors, using mean squared errors (MSE). The MSE measures the average of the squares of the errors - that is, the difference between the estimator and what is estimated. MSE for Models (1), (2), and (3) is computed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(donations_i - \hat{donations}_i)^2$$

Comparing Model (1) and Model (2)'s MSE will reveal whether incentivized experimental games or non-incentivized self-reported measures can better predict field measures on social preferences. Including Model (3) will tell us if, perhaps, having both instruments is better than only including one instrument.

5 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) *The external validity of giving in the dictator gam: A field experiment using the misdirected letter technique.* Governance and the efficiency of social systems (GESY), CESifo Working Paper No. 344.

[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 gam: 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] Hermann, B., Thoni, C., and Gachter, S. (2008). *Anti-social punishment across societies.* Science, 319, 1362-67.

[21] 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.

[22] 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.

[23] Karlan, D.S. (2005). *Using experimental economics to measure social capital and predict financial decisions.* American Economic Review, 95, 1688-99.

[24] 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.

[25] 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.

[26] Parco, J., Rapoport, A., and Stein, W. (2002). *Effects of financial incentives on the breakdown of mutual trust.* Psychological Science.

[27] Rabin, M. (1993). *Incorporating fairness into game theory and economics.* The American Economic Review, 83, 1281-1302.

[28] Ross, L., and Ward, A. (1996). *Naive realism in everyday life: Implications for social conflict and misunderstanding.* Values and Knowledge, 103-35.

[29] 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.

[30] Slonim, R., and Roth, A. (1998). *Learning in high stakes ultimatum games: An experiment in the Slovak Republic.* Econometrica, 66, 569-96.

[31] 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 under-charging 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

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)	○	○
Player 1 passes 1 point (keeps 9 points)	○	○
Player 1 passes 2 points (keeps 8 points)	○	○
Player 1 passes 3 points (keeps 7 points)	○	○
Player 1 passes 4 points (keeps 6 points)	○	○
Player 1 passes 5 points (keeps 5 points)	○	○
Player 1 passes 6 points (keeps 4 points)	○	○
Player 1 passes 7 points (keeps 3 points)	○	○
Player 1 passes 8 points (keeps 2 points)	○	○
Player 1 passes 9 points (keeps 1 point)	○	○
Player 1 passes 10 points (keeps 0 points)	○	○

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	
Player 1 passes 2 points, i.e. you receive 6 points	
Player 1 passes 3 points, i.e. you receive 9 points	
Player 1 passes 4 points, i.e. you receive 12 points	
Player 1 passes 5 points, i.e. you receive 15 points	
Player 1 passes 6 points, i.e. you receive 18 points	
Player 1 passes 7 points, i.e. you receive 21 points	
Player 1 passes 8 points, i.e. you receive 24 points	
Player 1 passes 9 points, i.e. you receive 27 points	
Player 1 passes 10 points, i.e. you receive 30 points	

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	○	○	○	○	○
I have donated money at the cash register when buying groceries	○	○	○	○	○
I have given money to a stranger (or lent to an acquaintance I don't know too well) in need	○	○	○	○	○
I have donated instead of sold my clothes/used items	○	○	○	○	○
I have donated to a charity	○	○	○	○	○
I have donated blood	○	○	○	○	○
I have done volunteer work for a charity/organization	○	○	○	○	○
I have delayed an elevator/held door open for a stranger	○	○	○	○	○
I have lent an acquaintance that I don't know too well with something of value to me (clothes, tools, etc)	○	○	○	○	○
I have pointed out a clerk's error (at a supermarket, restaurant) in undercharging me	○	○	○	○	○
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)	○	○	○	○	○
I have helped carry a stranger's belongings (e.g. groceries)	○	○	○	○	○
I have offered my seat on a bus/train to a stranger who was standing	○	○	○	○	○
I have helped an acquaintance with moving in/ moving out of their dorm/apartment/house	○	○	○	○	○
I have donated money/coins to the Salvation army bell-ringers	○	○	○	○	○

Figure 2a: Distribution of responses in ultimatum game, trust game, and public goods game

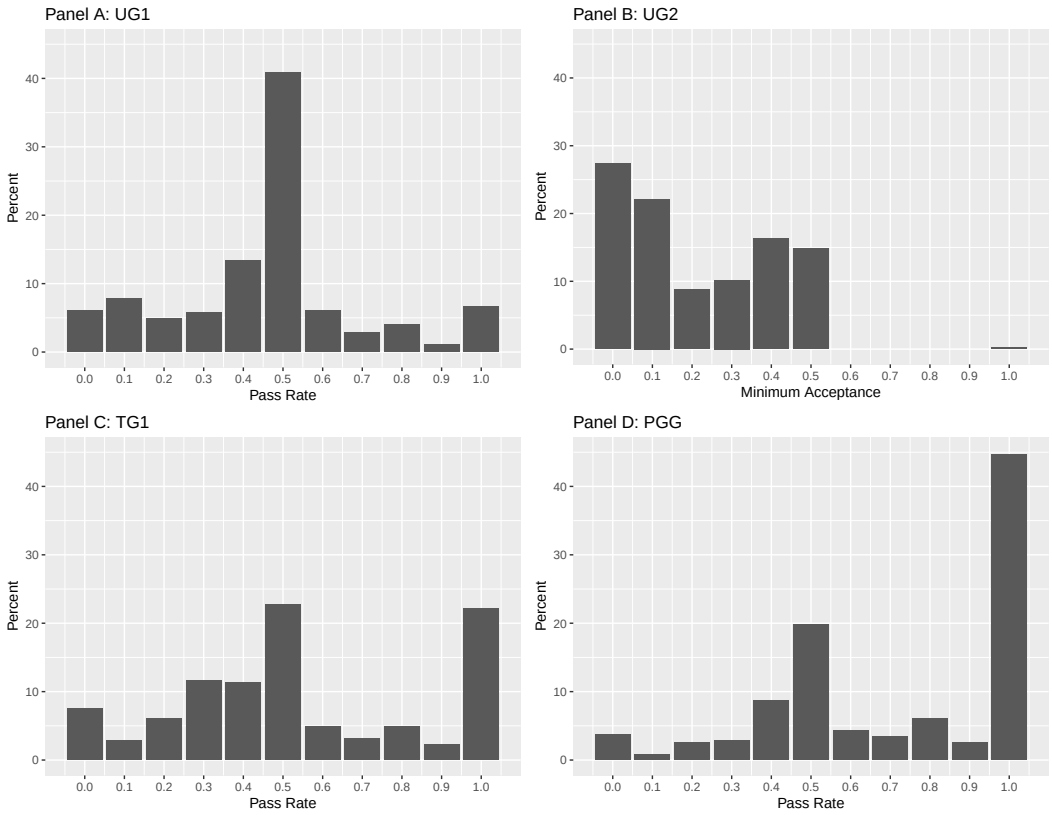


Figure 3: Total SRA scores

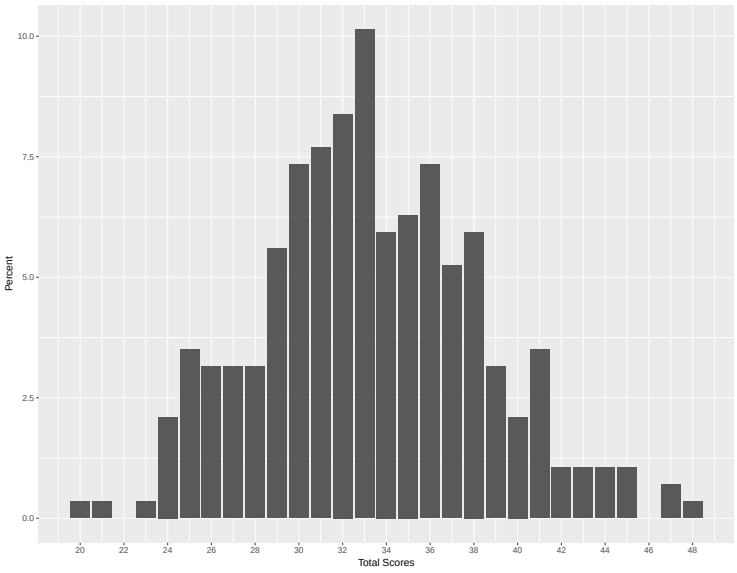
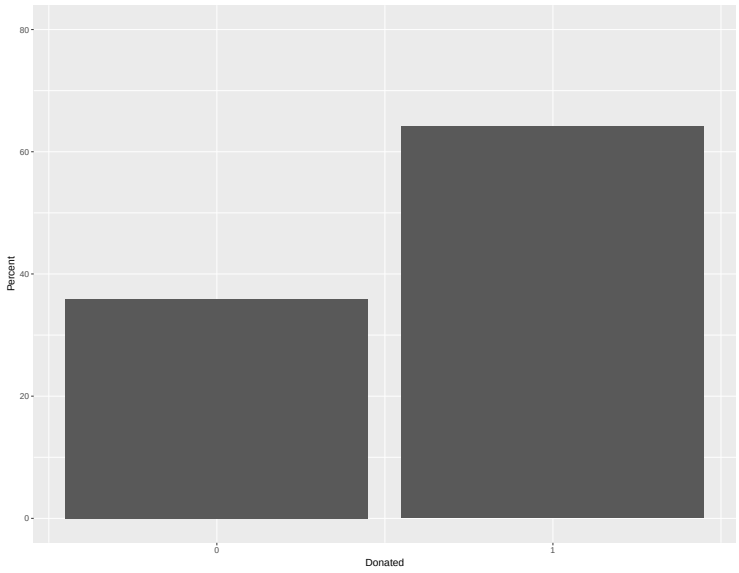
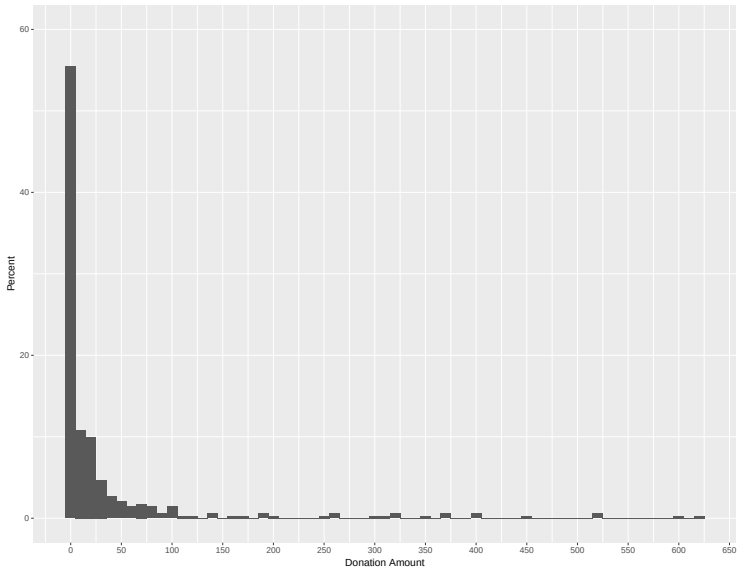


Figure 4: Donations


80
60
40
20
0
Percent
0
1
Donated


60
40
20
0
Percent
0
50
100
150
200
250
300
350
400
450
500
550
600
650
Donation Amount