Gender and Economic Norms

Mayo, Robert

2017

Gender and Economic Norms

Robert L. Mayo[†]

February 16, 2017

Abstract

I conduct an on-line experiment to test the hypothesis that the divergence in behaver between men and women found in Mayo (2016) was caused by men and women following different economic norms. The definition of norms is based on Bicchieri (2005), and includes three elements: First, that men and women exhibit different behavior. This was shown in Mayo (2016) and is confirmed here. Second, that men and women have different beliefs about what other men and women do and advocate be done when presented with the same options. This was shown in the results of an incentive compatible compensated online survey. Third, that providing messages about the behavior and statements of other men and women would cause a change in behavior. This was tested with an online controlled experiment. The results of this experiment were suggestive of the third element of the hypothesis, but did not prove causation. Further testing with higher statistical power is suggested.

JEL Classification: C99, D31, D64
Keywords: Charitable giving; Altruism; Gender

[†] Department of Economics, George Mason University, rmayo3@gmu.edu. I wish to acknowledge, with gratitude, Dr. Kevin McCabe for his advice and guidance. All errors or omissions are the sole responsibility of the author.

1. Introduction

In this paper, I use experimental data from a dictator game conducted over the internet to examine whether following different social norms is a cause of the significant difference in behavior between men and women found in Mayo (2016). In that study, a modified dictator game was used to decompose giving into its altruistic and egoistic components in a within subject design, thus allowing a measurement of the warm glow effect (Andreoni 1989). The experiment presented subjects selected to participate as dictators with a list of multiplication factors used to modify any amount they would send. Multiplication factors ranged from 0.2 to 5.0 and subjects were asked how much of their endowment they wished to send to an anonymously paired subject for each value. After indicating their choices, one multiplication factor was randomly selected and the amount the subject had entered for that value was deducted from their endowment, modified by the selected multiplication factor, and send to the anonymously paired subject.

That design held constant the price of the act of giving while varying the price of benefit to the recipient, thus allowing an estimation of the relative contributions of pure altruism and the warm glow of the act of giving as motivations for amounts sent. It was found that women sent roughly the same amount for all values of the multiplier, while men sent significantly less than women at low multiplier values and significantly more than women at high multiplier values. The conclusion of that paper was that women are motivated primarily by the warm glow of the act of giving while men are primarily motivated by pure altruism.

I consider here an alternate explanation for the difference in behavior. The pattern of choices made in the previous study are consistent with men and women following different behavioral norms. Sending a constant positive amount that is insensitive to the varying price of benefit to the recipient is consistent with adherence to a sharing norm. Sending a varying amount that is little or none when the multiplier is low and a large amount when the multiplier is high is consistent with following a norm of maximizing group wealth.

Two bodies of literature are particularly relevant to this issue; studies of the difference in behavior based on gender, and studies on the role of norms in regulating economic choice. That men and women make systematically different decisions in some economic contexts is neither a controversial nor new insight. This topic has been extensively examined in the psychology, sociology, and economic literature (Andreoni and Vesterlund 2001), (Swope, K. J., Schmitt, and Shupp 2008), (Mansbridge 1991), (Deaux 1976). Volume, however, has not equated to clarity. In prisoners' dilemma games, some studies have found women to be more cooperative than men (Meux 1973), (Aranoff and Tedeschi 1968), (Jones et al. 1968). Others have found women to be less cooperative than men, (Mack, Auburn, and Knight 1971), (Kahn, Hottes, and Davis 1971), (Rapoport and Chammah 1965). And a third group have found no gender difference, (Orbell, Dawes, and Schwartz-Shea 1994), (Stockard, Van de Kragt, and Dodge 1988), (Dana, Cain, and Dawes 2006), (Dawes, McTavish, and Shaklee 1977). In addition, Gilligan (1982) claims that men and women follow entirely different conceptions of morality, with women focusing on relationship maintenance while men focus on rule adherence.

Studies of gender in dictator games have also produced conflicting results. Cox and Deck (2006) find that women are more sensitive to the price of giving, while (Eckel et al. 2017) find that women are less price sensitive. Andreoni and Vesterlund (2001) find that women give

more than men when the price is high but men give more when the price is low, suggesting that women are more price sensitive to the cost of giving. They also find that men are more likely to give an extreme amount (i.e. all or nothing) while women rarely do so.

The body of research on the influence of social norms on economic choice is less conflicted. A wide variety of papers have documented economic behavior that is modified by the knowledge of the behavior of others in the same situation, absent any strategic or reputational interaction. This phenomenon has been found in dictator games (Krupka and Weber 2009), (Cason and Mui 1998); in gift exchange (Gächter, Nosenzo, and Sefton 2012), (Thöni and Gächter 2012), and in a trust game (Mittone and Ploner 2011).

To test whether men and women follow different economic norms, a clear definition of norms is required. For the purposes of this analysis, I adopt the definition of norms presented in Bicchieri (2005). By this definition, a norm has three components. 1.) Empirical expectation: a set of beliefs about what others do in a given circumstance. 2.) Normative expectation: a set of beliefs about what others advocate be done in that circumstance. 3.) Conditional preference: a strategy choice in that circumstance that is predicated on the person's empirical and/or normative expectation of others. All of these three necessary and sufficient conditions for a norm to exist are relative to a reference group of persons. For example, if I believe that other men in my office wear ties, and I believe that other men in my office believe ties should be worn in the office, and I prefer to wear a tie in the office because of those beliefs, then a norm exists relative to the reference group of men in my office. My empirical or normative expectations about people not in my reference group, such as women in my office or men in another office, are not relevant to the norm. It is important to also understand the difference between normative expectations and personal normative beliefs. My normative expectations, in the office example, are that other

men in my office believe ties should be worn. This is not the same as my personal belief that ties are useless and uncomfortable, and so should not be worn.

Inserting this definition of norm into the statement that men and women follow different economic norms, makes the hypothesis testable. The remainder of this paper is organized as follows. Section two describes the experimental design. Section three reviews results. Section four is discussion, and section five concludes.

2. Methods

2.1. Experiment Design

2.1.1. Overview

To conclude that men and women follow different economic norms in the variable price dictator game used in Mayo (2016), it is necessary to show all three components of the norm as defined in Bicchieri (2005), specifically that in that game:

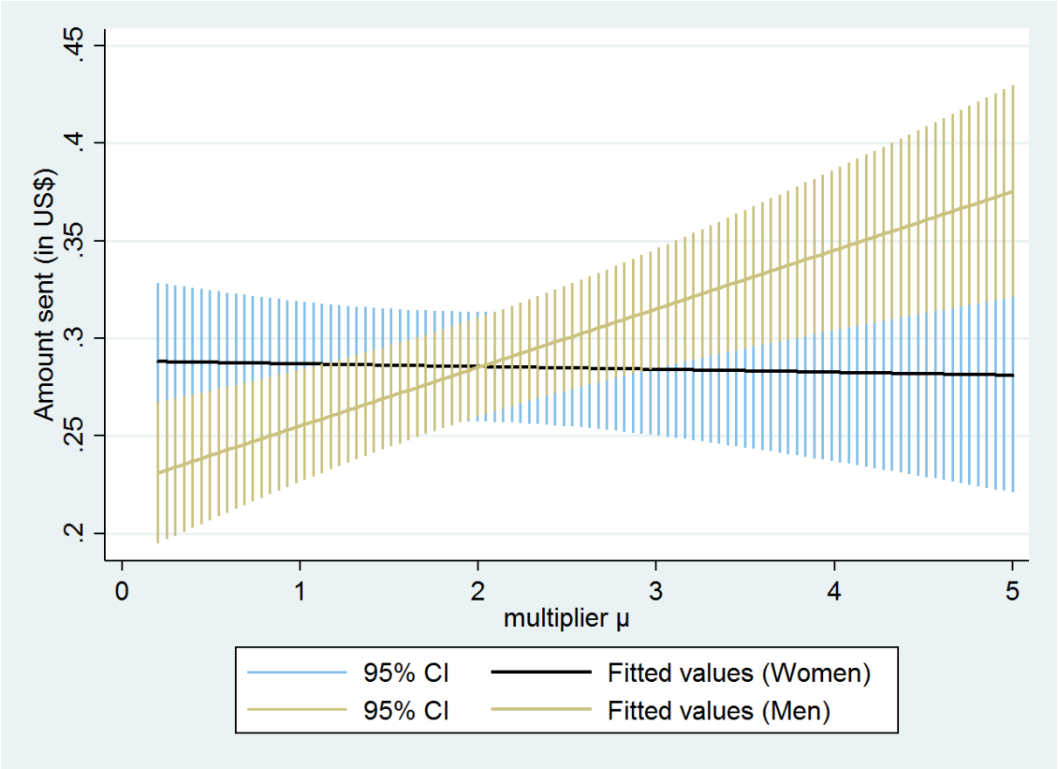
- Men and women make different choices.
- Men and women have different empirical and/or normative expectations.
- Men and women's different choices are caused by their different expectations.

That men and women make different choices was established in Mayo (2016). A brief recitation of the relevant results is included below. The second element, that men and women have different expectations about what others in their reference group do (empirical) and say should be done (normative), will be shown using an incentive compatible compensated survey. The third element, showing that the difference in beliefs causes the difference in behavior will be shown by a controlled experiment providing information to change beliefs and then observing choices.

2.1.2. Men and women make different choices.

The core result from Mayo (2016) is shown in figure 1. In a dictator game where the amount sent was multiplied by the experimenter prior to reaching the recipient, the multiplier had no significant effect on the amounts sent by women, but had a significant, positive effect on the amounts sent by men. A full discussion of the results is found in that paper.

Figure 1. Amounts sent by multiplier for men and women



In Mayo (2016), subjects were presented with multipliers $\mu \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5\}$. For this paper, I examine only the behavior of men and women at multiplier $\mu = 0.2$. I choose this value because it makes the clearest predictions from following a sharing vs. efficiency norm. At $\mu > 1.0$, a norm to share and a norm to maximize social total both induce sending a positive amount. At $\mu < 1.0$, a sharing norm would still induce sending a positive amount, but a norm to maximize social total would induce sending zero. Assuming subjects do

not follow a single norm exclusively or perfectly, a difference in behavior due to following different norms should be most apparent at the lowest value of μ, in Mayo (2016) that was 0.2. The mean amounts sent at $\mu = 0.2$ in that study were 20.4¢ for men and 28.2¢ for women.

2.1.3. Men and women have different empirical and/or normative expectations.

Empirical and normative expectations were found using a compensated survey. The survey was run online using Amazon's Mechanical Turk micro-employment website for recruitment and payment. A discussion of the technology used is in section 2.2 and the experiment interface as seen by subjects is included in appendix A. Subjects received a $0.25 base payment, equivalent to a show-up fee in an in-person laboratory experiment. After reading and accepting an IRB disclosure, but prior to starting the experiment, prospective subjects were asked to answer this question: *"You have a basket containing five apples. You eat one apple and sell two apples. How many apples are now in your basket?"* This question serves to prevent automated scripts from entering the experiment, similar in function to CAPTCHA[1] codes seen on websites where automated spam is a concern. Subjects were not allowed to enter the experiment without entering the correct answer. Upon entering the experiment, subjects were told that they would see the instructions from a prior study, and would then be asked questions about the choices they think participants in that study made. To encourage subjects to give accurate responses, they were told that the person whose answer was closest to what people in the other study actually did, would be paid $20, in addition to their base participation fee. In case of a tie, the bonus payment was split evenly.

After reading the instructions from the prior study, subjects were randomly assigned to be asked either "In the previous study you have just read about: What was the average amgount sent

[1] Examples of the use of CAPTCHA systems is available at https://www.google.com/recaptcha/intro/index.html.

by men?" or "In the previous study you have just read about: What was the average amgount sent by women?" Subjects were only askeed about one gender to eliminate possible consistency and sequence effects. Responses to these questions gives me the beliefs of men and women about what both men and women do; their empirical expectations. On the following page, subjects were asked "In the previous study you have just read about: How much do you think should be sent?" To elicit second order beliefs rather than first order beliefs (e.g. what I think you want, not what I want) subjects were told that the person whose answer was closest to the average answer would receive an additional payment of $20. Answers to this question give me subjects' normative expectations. Finally, subjects were asked their gender which allowed the prior two answers to classified by gender.

2.1.4. Men and women's different choices are caused by their different expectations.

To show causation, I use an experiment design adapted from Bicchieri and Xiao (2009). In that study, subjects were shown information derived from a prior study (Xiao and Houser 2007) of both what other subjects did and what they said should be done. Data from different sessions was used to construct multiple different yet truthful statements[2]. These statements were "60% of subjects in a prior session…" followed by a statement that the subjects either made a fair choice (FC), a selfish choice (SC), stated a fair belief (FB), stated a selfish belief (SB), FC and SB, or SC and FB, followed by an opportunity to make a choice themselves. The purpose being to alter the normative or empirical expectations of the subjects then observe any resulting change in choices.

[2] The issue of experimental deception was addressed in that paper to the effect that so long as subjects are informed that the information presented is only a part of the data and not claimed to be representative of the entire data set, no deception has occurred.

I adapt this design by presenting subjects with the same instructions as were seen in the incentivized survey, and are given the stated initial endowment of $1.00. On the next screen subjects are shown two things: information about what subjects in a prior study did and said, and a question of how much they want to send to an anonymous, randomly paired person in another group. In Bicchieri and Xiao (2009), two of the six information treatments (FC + SB, FB + SC) pushed empirical and normative beliefs in opposite directions and so eliminated, within reason, the possibility of experimenter demand effects. In the other four treatments, however, this was not the case. The authors state that the double-blind procedure should protect against this possibility as the subjects knew that the experimenters could not know who made which choice. While this may be sufficient, I guarded against this possibility by providing information that (if effective) would push choices in opposite directions in every treatment. Data was selectively drawn from responses in Mayo (2016) to construct the following four truthful statements:

> *"Before you make your decision, here is information about select groups from a previous study. These responses may not be representative of everyone in the previous study."* plus one of the following four statements:
>
> *The most frequent amount that MEN:*
>
> | *ACTUALLY sent* | *SAID should be sent* |
> |---|---|
> | *$0.50* | *$0.00* |
>
> *The most frequent amount that MEN:*
>
> | *ACTUALLY sent* | *SAID should be sent* |
> |---|---|
> | *$0.00* | *$0.50* |
>
> *The most frequent amount that WOMEN:*
>
> | *ACTUALLY sent* | *SAID should be sent* |
> |---|---|
> | *$0.50* | *$0.00* |

The most frequent amount that WOMEN:

ACTUALLY sent	*SAID should be sent*
$0.00	*$0.50*

Subjects were randomly assigned to see one of the four statements. In addition, whether "actually" or "said" appeared on the left or right of the screen was also randomized. Below this information was a request to choose how much to send. On the next page, subjects were asked their gender and exited the experiment.

If subjects' preferences were not conditional on their empirical or normative expectations, then providing them with information about what others did or said should be done, and thus modifying those expectations, would have no effect on their choices. However, if subjects' preferences were conditional on those expectations, and if the new information changed them, then a change in choices may be observed. Since the amounts sent and said should be sent were always opposite ($0.00/$0.50 or $0.50/$0.00), the direction of change would indicate which expectation, empirical or normative, had the causal effect. For reasons discussed in the next section, subjects in the recipient group are recruited in a separate asynchronous session.

2.2. Tools

The design described in the previous section was implemented online using oTree and Mechanical Turk.

2.2.1. oTree

oTree (Chen, Schonger, and Wickens 2016) is a software package for running multi-player decision experiments over the internet. oTree experiments are coded in Python using the

Django web framework.[3] Experiments are developed on a local machine, and then iploaded to a remote server for internet access. The experiment software was hosted on Heroku[4]. Since the experiment is available online, the only technology required for subjects' computers is a normal browser using any operating system. The experiment can be accessed through a desktop computer, laptop, tablet, or even a phone. Django automatically re-scales the pages to best fit each screen size.

2.2.2. Mechanical Turk

Amazon's Mechanical Turk micro-employment website was used for subject recruitment and payment. The experiment was posted to Mechanical Turk using IRB approved advertisement language. Workers who chose to participate are shown the experiment interface inside a window on the Mechanical Turk Website. To guard against automated scripts generating random responses to generate payments, the script filtering question discussed in the previous section was used before both the incentivized questionnaire and the controlled experiment.

Interaction between dictators and recipients are asynchronous. All dictators make their choices of how much to send, then later a second group of Mechanical Turk workers is recruited as passive recipients. Participants in the dictator group are ineligible to participate in the recipient group. Since the information flow is only from dictator to recipient, the small time delay this method imposes between money sent and money received should not bias dictator choices. No Mechanical Turk worker was allowed to participate in any part of the study more

[3] Technical details of Django can be found at https://www.djangoproject.com/.
[4] Technical details of Heroku can be found at https://www.heroku.com/.

than once. This was enforced through assigning a "worker qualification" to participants and excluding workers who already possessed the qualification.

2.3. Security and randomization

No personally identifying information is passed from worker to experimenter other than a Mechanical Turk worker identification number. Amazon securely stores workers' personal information including financial information required to process payments. This simplifies experimenter record keeping and ensures subject privacy. Through the oTree API, a worker qualification can be set. This tells Mechanical Turk that the requester only wants the task to be visible to workers who meet certain criteria. The criteria used in this experiment were that the worker must reside in the United States, not have previously participated in the experiment, must have previously completed at least 100 work assignments with a satisfactory rating of at least 90%. The latter two conditions were imposed to exclude persons who have not learned how to use the Mechanical Turk website or who tend to enter random responses.

The only information workers see prior to accepting the assignment is an IRB disclosure and a generic description of the task as an academic experiment in decision making. Therefore, self-selection into the experiment should not occur. Methodological questions about the external validity of experiments conducted on-line without control of subjects' environments have been examined and found to not be a significant problem (Horton, Rand, and Zeckhauser 2011). Payments in on-line experiments using Mechanical Turk can be much lower than the same experiment in an in-person laboratory. Amir, Rand, and Gal (2012) studied both issues by replicating a series of classic economic experiments in an on-line environment and found no significant difference between the behavior of on-line and in-person subjects.

3. Results

3.1. Compensated survey

The compensated survey was run over three consecutive weekday evenings between the hours of 4PM and 10PM Eastern time. 163 Mechanical Turk workers participated in the experiment. There was rough gender balance with 91 (56%) male subjects and 72 (44%) female subjects. Since each subject was asked only about a single randomly chosen gender, there were four sub-groups. Amounts subjects believed were sent in the prior study are shown in table 1. Amounts subjects said should be sent are shown in table 2.

Table 1. Summary statistics of the amounts subjects believed were sent in the prior study.

	N	Mean	Median	St. Dev.
Men asked about men	50	0.331	0.250	0.307
Men asked about women	41	0.476	0.500	0.359
Women asked about men	27	0.299	0.250	0.264
Women asked about women	45	0.481	0.500	0.338

Table 2. Summary statistics of the amounts subjects said should be sent.

	N	Mean	Median	St. Dev.
Men	91	0.353	0.325	0.338
Women	72	0.390	0.500	0.400

As shown in figure 2, the beliefs of men and women are similar. Both genders believe that men send less than women. Results of a one-way between subjects ANOVA to compare mean amounts in the six groups is shown in table 3. The p-value of 0.024 provides significant evidence of difference. A post-hoc test of pairwise comparison of means with Duncan correction for multiple comparisons is shown in table 4.

Figure 2. Amounts subjects believe were sent and should be sent.

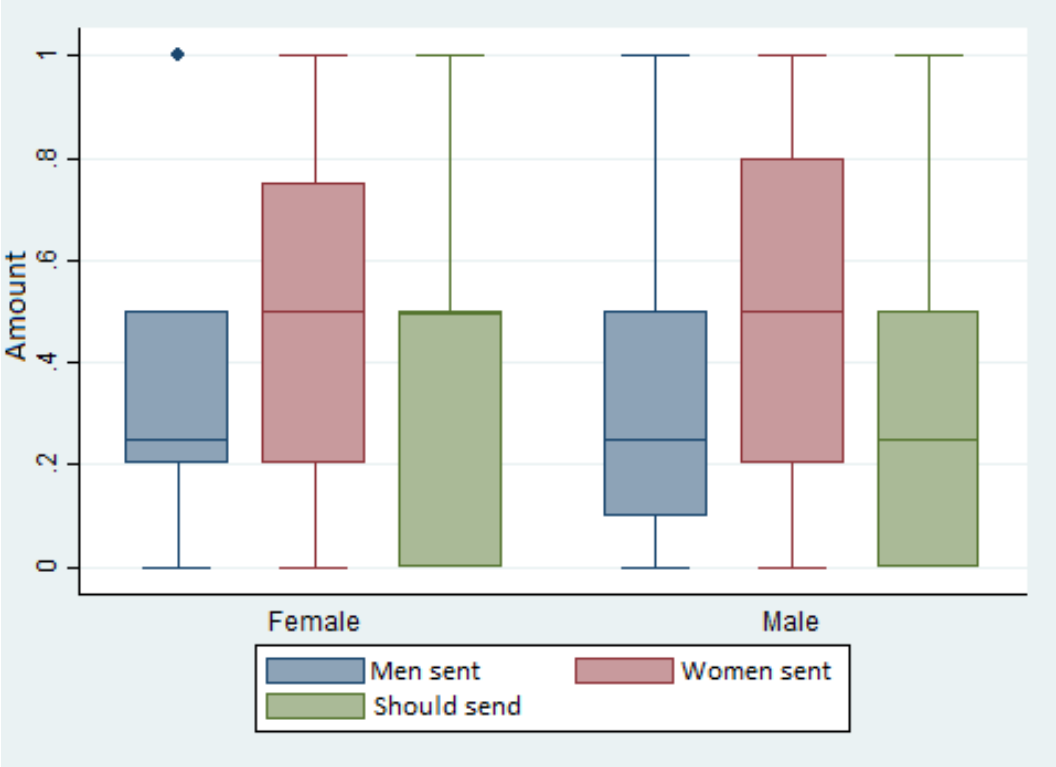


Table 3. One-way ANOVA of amounts by group.

Source	Sum of squares	Degrees of freedom	Mean square	F	P-value
Between groups	1.048	3	0.349	3.23	0.024
Within groups	17.165	159	0.108		
Total	18.213	162	0.112		

Table 4. Pairwise comparisons of means

Comparison of means	Diff.	Std. Err.	t-statistic	p-value
Men believe women send vs. men believe men send	0.146	0.069	2.11	0.037**
Women believe men send vs. men believe men send	-0.0.3	0.078	-0.40	0.690
Women believe women send vs. men believe men send	0.155	0.068	2.23	0.035**
Women believe men send vs. men believe women send	-0.177	0.081	-2.17	0.040**
Women believe women send vs. men believe women send	0.004	0.071	0.07	0.946
Women believe women send vs. Women believe men send	0.182	0.080	2.27	0.038**

p-values are adjusted with Duncan correction for multiple comparisons.
*** Significant at 1%
** Significant at 5%
* Significant at 10%

The difference of beliefs within gender and similarity across gender is visually apparent in figure 2. Pairwise comparisons in table 4 confirm the visual observation.

Result 1: Men believe that women send more than men and the difference is statistically significant.

Result 2: Women believe that women send more than men and the difference is statistically significant.

For the amounts, they believed should be sent (personal normative beliefs rather than normative expectations), the mean for men was $0.35 and the mean for women was $0.41. This is consistent with the prior result, however as shown in table 5 the difference was not statistically significant.

Table 5. Two sample t-test assuming equal variance of amounts men and women believe should be sent.

	Men	Women
Mean	0.346	0.409
Variance	0.137	0.135
Observations	91	72
t-statistic	-1.080	
p-value (two tails)	0.282	

*** Significant at 1%
** Significant at 5%
* Significant at 10%

Result 3: The amount that women believe should be sent is greater than the amount men believe should be sent, but the difference is not statistically significant.

3.2. Controlled experiment

241 Mechanical Turk subjects participated in the controlled experiment. There was rough gender balance with 123 (51%) men and 118 (49%) women. Given the 2 x 2 x 2 design, amounts sent are in 8 groups, i.e. Men/women x told that men/women x sent $0.50 and said $0.00/sent $0.00 and said $0.50 as shown below with brief names in parentheses.

1.) Men who were told men sent $0.50 and said $0.00 should be sent. (men men hi)

2.) Men who were told men sent $0.00 and said $0.50 should be sent. (men men lo)

3.) Women who were told women sent $0.50 and said $0.00 should be sent. (women women hi)

4.) Women who were told women sent $0.00 and said $0.50 should be sent. (women women lo)

5.) Men who were told women sent $0.50 and said $0.00 should be sent. (men women hi)

6.) Men who were told women sent $0.00 and said $0.50 should be sent. (men women lo)

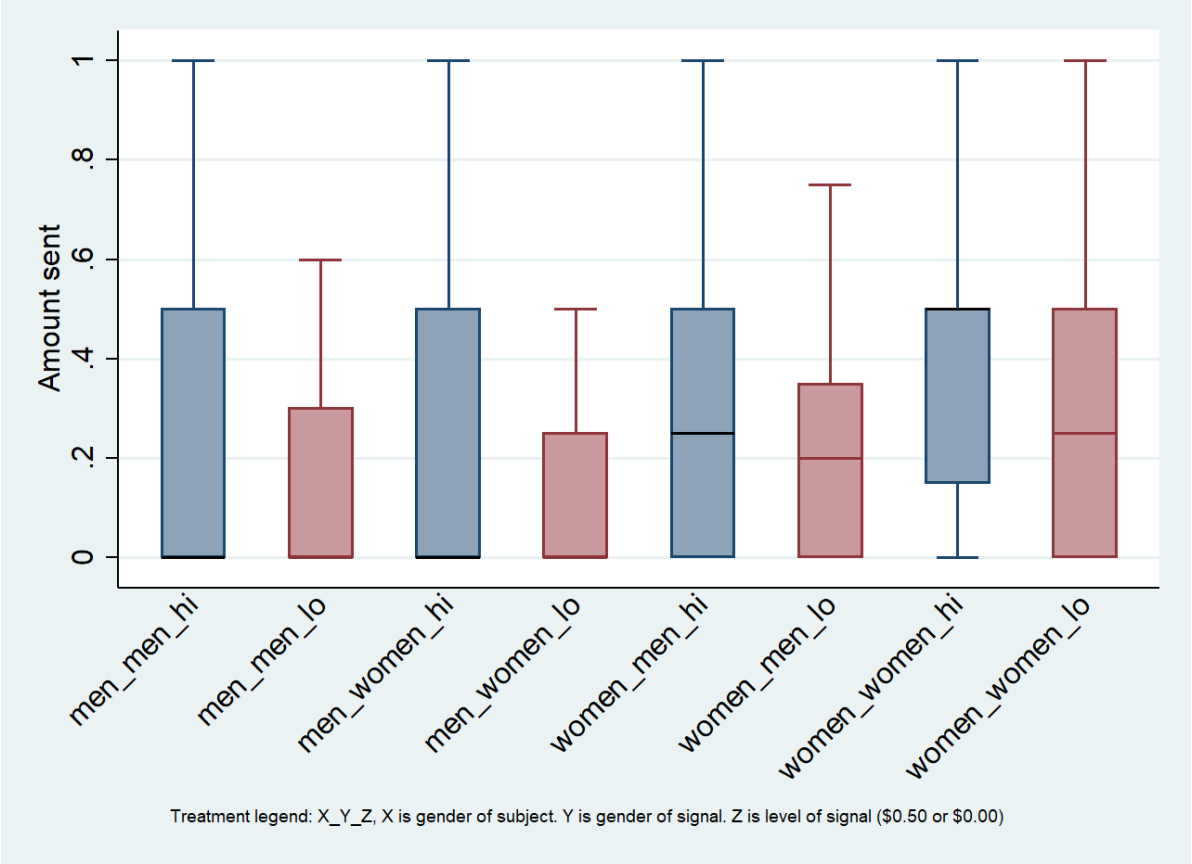
7.) Women who were told men sent $0.50 and said $0.00 should be sent. (women men hi)

8.) Women who were told men sent $0.00 and said $0.50 should be sent. (women men lo)

Groups 1 through 4 show the effects of information about the same gender, while groups 5 through 8 show the effects of information about the opposite gender. Summary statistics for the amounts each group sent are shown in table 6. Amounts sent as box plots are shown in figure 3.

Table 6. Summary statistics of amount sent for each treatment group.

Group	N	Mean	Median	St. Dev.
Men men hi	28	0.22	0.00	0.28
Men men lo	39	0.15	0.00	0.21
Women women hi	32	0.36	0.50	0.26
Women women lo	30	0.26	0.50	0.26
Men women hi	33	0.24	0.00	0.33
Men women lo	23	0.12	0.00	0.19
Women men hi	27	0.27	0.25	0.30
Women men lo	23	0.12	0.20	0.21



It is visually apparent that both genders sent more after receiving information that either gender sent $0.50 compared to the amounts sent after receiving information that either gender sent $0.00. The median amount sent by men in all treatments was $0.00, while the median amount sent by women in all treatments was between $0.20 and $0.50. The results of two sample unpaired t-tests of the difference between amounts sent after receiving information that the same gender sent a high ($0.50) and low ($0.00) amount are not statistically significant, as shown in tables 7 and 8.

Table 7. Two sample t-test comparing amounts sent by men told that other men sent $0.50 vs. $0.00

	$0.50 message	$0.00 message
Mean	0.221	0.153
Variance	0.080	0.045

Observations	28	39
t-statistic	1.139	
p-value (two tails)	0.259	

*** Significant at 1%
** Significant at 5%
* Significant at 10%

Result 4: Men sent less after receiving information that other men sent nothing, but the result is not statistically significant.

Table 8. Two sample t-test comparing amounts sent by women told that other women sent $0.50 vs. $0.00

	$0.50 message	$0.00 message
Mean	0.358	0.262
Variance	0.070	0.072
Observations	32	30
t-statistic	1.418	
p-value (two tails)	0.161	

*** Significant at 1%
** Significant at 5%
* Significant at 10%

Result 5: Women sent less after receiving information that other women sent nothing, but the result is not statistically significant.

As a comparison, the model shown in equation 1 was estimated using quantile regression, which at the median is the same as least absolute deviation. Given that subjects' choices of amount to send were constrained to be between $0.00 and $1.00, and both values did occur in the data, the data may[5] be censored and so OLS regression could not be relied upon to produce unbiased estimates. Quantile regression was selected as a substitute over Tobit estimation since

[5] The fact that responses above/below a threshold will be coded at the threshold, and a substantial number of observations are at the threshold is necessary but not sufficient evidence of censoring. A simple visual test to confirm censoring is the presence of a substantial number of observations near, but not at, the threshold. In the absence of such observations the more likely explanation is that the observed distribution is the same as the latent distribution. This may be caused by the data having multiple data generating processes that have not been separately modeled. Histograms and kernel density plots are included in the appendix for evaluation of this issue. For a comprehensive discussion of these issues, see Greene (2008) pp 875-881.

unlike quantile regression, Tobit estimation is both parametric and highly sensitive to assumption violations (Arabmazar and Schmidt 1982), (Arabmazar and Schmidt 1981), (Maddala and Nelson 1975). Regression results are shown in table 9.

$$s = \beta_1 gen + \beta_2 hi + \beta_3 same + \beta_4 same * hi + \epsilon \qquad \text{eq.1}$$

Where:

$$s = amount\ sent$$

$$gen = \begin{Bmatrix} 1\ if\ subject\ is\ female \\ 0\ if\ subject\ is\ male \end{Bmatrix}$$

$$hi = \begin{Bmatrix} 1\ if\ message\ is\ \$0.50 \\ 0\ if\ message\ is\ \$0.00 \end{Bmatrix}$$

$$same = \begin{Bmatrix} 1\ if\ message\ is\ about\ the\ same\ gender \\ 0\ if\ message\ is\ about\ the\ opposite\ gender \end{Bmatrix}$$

Table 9. Quantile regression results, N = 241, Pseudo R2 = 0.0987.

Variable	Estimate
D_Gender, 1 = Female	0.250 (0.000)
D_Sent high, 1 = Subject told others sent $0.50	0.00 (1.000)
D_Same gender, 1 = Information is about the same gender	0.00 (1.000)
D_Sent high*D_same gender	0.200 (0.165)
Constant	0.000 (1.000)

D indicates a dummy variable
p-value in parentheses
*** Significant at 1%
** Significant at 5%
* Significant at 10%

The coefficient for the dummy for gender is positive and highly significant, indicating that women sent more than men.

Result 6: Women send more than men at μ = 0.2. The result is statistically highly significant. This result in Mayo (2016) is confirmed.

The interaction term between receiving a message that others sent $0.50 (D_Sent high) and that message being about the same gender as the receiver of the information (D_Same gender) is positive, but not significant.

Result 7: Subjects who received information that the same gender sent $0.50, sent more, but the result is not statistically significant.

4. Discussion

The results reported in the previous section show significant evidence in support of two of the three elements necessary to prove that in the context of a dictator game with a multiplier of 0.2, men and women follow different economic norms. The first element, that men and women make different decisions was shown in Mayo (2016) and is confirmed in this study. The second element, that men and women have different empirical and/or normative expectations is also shown. Both men and women believe that women send more in this game than men. This is evidence of different empirical expectations. The evidence of different normative expectations is, however, inconclusive but not necessary to prove the proposition. The remaining necessary element is to show that the preferences of men and women are conditional on the behavior and/or advocacy of those in their reference groups. Do men and women define their reference groups as same gender persons or all persons? The evidence that both men and women believe that women send more than men is consistent with same gender reference groups. If reference groups are

same gender, then we would expect that receiving information about the behavior of persons of the same gender would have a larger effect on behavior than receiving information about the behavior of persons of the opposite gender. This effect is seen in the sample, as men who were told that men sent a high amount sent more than did men who were told that women sent a high amount. The same effect was observed among women, i.e. women changed their behavior more when given information about other women than when given information about men. Although this is consistent with the hypothesis that men and women follow different economic norms, the effect is not statistically significant. The p-value of the interaction of message and same gender in the regression is 0.165, which is suggestive but nothing more.

What are the possible explanations for this? The first possible explanation is simple, that the hypothesis is wrong; men and women do not follow different economic norms. Failure to reject the null hypothesis does not imply that the null is correct, so consideration of other possible explanations is warranted. As discussed previously, in Bicchieri and Xiao (2009) the majority of treatments used a message containing a single piece of information, e.g. that men made a fair choice, or that women had selfish beliefs. This introduces the possibility of experimenter demand effects magnifying the effect size. Although their argument as to why this did not occur is reasonable, I chose to employ an additional safeguard by always presenting subjects with two, opposite pieces of information, e.g. men sent high *and* said you should send low, or women sent low *and* said you should send high. Assuming both pieces of information had some effect on choices, then the observed effect would that of the dominant effect minus the lesser effect. If the effects of changing normative and empirical expectations were roughly the same, then no aggregate effect would be seen. The purpose of Bicchieri and Xiao (2009) was to determine which factor was dominant and they found that empirical expectations, but not

normative altered subjects behavior. This is consistent with the finding here that behavior did shift towards the empirical signal, but cannot be ruled out as an explanation.

Another possible explanation is a lack of statistical power. A power analysis conducted on the results shown in table 7, testing whether men who received a message that other men sent a high amount and said a low amount should be sent, subsequently sent a different amount than men receiving the opposite message. Power analysis results are shown in table 10.

Table 10. Post-hoc power analysis: unpaired t-test.

Parameter	Value
Tails	2
Effect size d	0.312
α-level	0.05
df	65
N high	28
N low	39
Mean sent high	0.221
Mean sent low	0.153
S.D. high	0.283
S.D. low	0.121
Power	0.237

Any of the previously discussed explanations for the results are possible, but it is obvious that my tests were underpowered. A power of 0.237 implies a probability of 0.763 of type II error. Assuming the effect size remained constant, a n increase in sample size from 67 to 550 would be needed for significance at $\alpha = 0.5$. Given the low cost of experimentation using Mechanical Turk, this is not an unreasonable sample size. Unfortunately, the budget for this study did not allow for the larger samples that may have provided a conclusive answer.

5. Conclusion

This paper reports the results of an on-line experiment designed to test the hypothesis that the divergence in behaver between men and women found in Mayo (2016) was caused by men and women following different economic norms. The definition of norms used included three elements: First, that men and women exhibit different behavior. This was shown in Mayo (2016) and is confirmed here. Second, that men and women have different beliefs about what other men and women do and advocate be done when presented with the same options. This was shown in the results of an incentive compatible compensated online survey. Third, that providing messages about the behavior and statements of other men and women would cause a change in behavior. This was tested with an online controlled experiment. The results of this experiment were suggestive, but did not prove the third element of the hypothesis, possibly caused by low statistical power.

References

Amir, Ofra, David G Rand, and Ya'akov Kobi Gal. 2012. "Economic Games on the Internet: The Effect of $1 Stakes." *PLoS ONE* 7 (2): 1–4. doi:10.1371/journal.pone.0031461.

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97 (6): 1447. doi:10.1086/261662.

Andreoni, James, and Lise Vesterlund. 2001. "Which Is the Fair Sex? Gender Differences in Altruism." *The Quarterly Journal of Economics* 116 (1): 293–312. doi:10.1162/003355301556419.

Arabmazar, Abbas, and Peter Schmidt. 1981. "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity." *Journal of Econometrics* 17 (2): 253–58. doi:10.1016/0304-4076(81)90029-4.

———. 1982. "An Investigation of the Robustness of the Tobit Estimator to Non-Normality." *Econometrica* 50 (4): 1055–63. http://www.jstor.org/stable/1912776.

Aranoff, Daniel, and James T Tedeschi. 1968. "Original Stakes and Behavior in the Prisoner's Dilemma Game." *Psychonomic Science* 12 (2): 79–80. doi:10.3758/BF03331202.

Bicchieri, Christina. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511616037.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only If Others Do so." *Journal of Behavioral Decision Making* 22 (2): 191–208. doi:10.1002/bdm.621.

Cason, Tn, and Vl Mui. 1998. "Social Influence in the Sequential Dictator Game." *Journal of Mathematical Psychology*. doi:doi:10.1006/jmps.1998.1213.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9. Elsevier B.V.: 88–97. doi:10.1016/j.jbef.2015.12.001.

Cox, James C, and Cary A Deck. 2006. "When Are Women More Generous than Men?" *Economic Inquiry*. doi:10.1093/ei/cbj042.

Dana, Jason, Daylian M Cain, and Robyn M Dawes. 2006. "What You Don Õ T Know Won Õ T Hurt Me : Costly (but Quiet) Exit in Dictator Games" 100: 193–201. doi:10.1016/j.obhdp.2005.10.001.

Dawes, Robyn M., Jeanne McTavish, and Harriet Shaklee. 1977. "Behavior, Communication, and Assumptions About Other People's Behavior in a Commons Dilemma Situation." *Journal of Personality and Social Psychology* 35 (1): 1–11. doi:10.1037/0022-3514.35.1.1.

Deaux, Kay. 1976. *The Behavior of Women and Men*. Monterey, CA: Brooks Publishing.

Eckel, Catherine C, Philip J Grossman, Catherine C Eckel, and Philipj Grossman. 2017. "Are Women Less Selfish Than Men ?: Evidence from Dictator Experiments Published by : Wiley on Behalf of the Royal Economic Society Stable URL : http://www.jstor.org/stable/2565789 JSTOR Is a Not-for-Profit Service That Helps Scholars , Researchers , and S" 108 (448): 726–35.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2012. "The Impact of Social Comparisons on Reciprocity." *Scandinavian Journal of Economics*. doi:10.1111/j.1467-9442.2012.01730.x.

Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*.

Harvard University Press.

Greene, William H. 2008. *Econometric Analysis*. 7th ed. Upper Saddle River, N.J.: Prentice Hall.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics* 14 (3): 399–425. doi:10.1007/s10683-011-9273-9.

Jones, B, M Steele, J Gahagan, and J Tedeschi. 1968. "Matrix Values and Cooperative Behavior in the Prisoner's Dilemma Game." *Journal of Personality and Social Psychology* 8 (2): 148–53.

Kahn, Arnold, Joe Hottes, and William L Davis. 1971. "Cooperation and Optimal Responding in the Prisoner's Dilemma Game: Effects of Sex and Physical Attractiveness." *Journal of Personality and Social Psychology* 17 (3): 267–79. doi:10.1037/h0030597.

Krupka, Erin, and Roberto A. Weber. 2009. "The Focusing and Informational Effects of Norms on pro-Social Behavior." *Journal of Economic Psychology*. doi:10.1016/j.joep.2008.11.005.

Mack, David, Paula N Auburn, and George P Knight. 1971. "Sex Role Identification and Behavior in a Reiterated Prisoner's Dilemma Game." *Psychonomic Science* 24: 280–81. doi:10.3758/BF03329005.

Maddala, G.S., and Forrest Nelson. 1975. "Specification Errors in Limited Dependent Variable Models." 96. National Bureau of Economic Research Working Paper Series. Cambridge, MA. doi:10.3386/w0096.

Mansbridge, Jane. 1991. *Beyond Self-Interest*. Chicago: University of Chicago Press.

Mayo, Robert L. 2016. "Decomposing Altruism : An Experiment to Measure Motivations for

Giving by Demographic Group." *Under Submission*. Arlington, VA.

Meux, Eleanor P. 1973. "Concern for the Common Good in an N-Person Game." *Journal of Personality and Social Psychology* 28 (3): 414–18.

Mittone, Luigi, and Matteo Ploner. 2011. "Peer Pressure, Social Spillovers, and Reciprocity: An Experimental Analysis." *Experimental Economics*. doi:10.1007/s10683-010-9263-3.

Orbell, John, Robyn Dawes, and Peregrine Schwartz-Shea. 1994. "Trust, Social Categories, and Individuals: The Case of Gender." *Motivation and Emotion* 18 (2): 109–28. doi:10.1007/BF02249396.

Rapoport, A, and a M Chammah. 1965. "Sex Differences in Factors Contributing to the Level of Cooperation in the Prisoner's Dilemma Game." *Journal of Personality and Social Psychology* 2 (6): 831–38. doi:10.1037/h0022678.

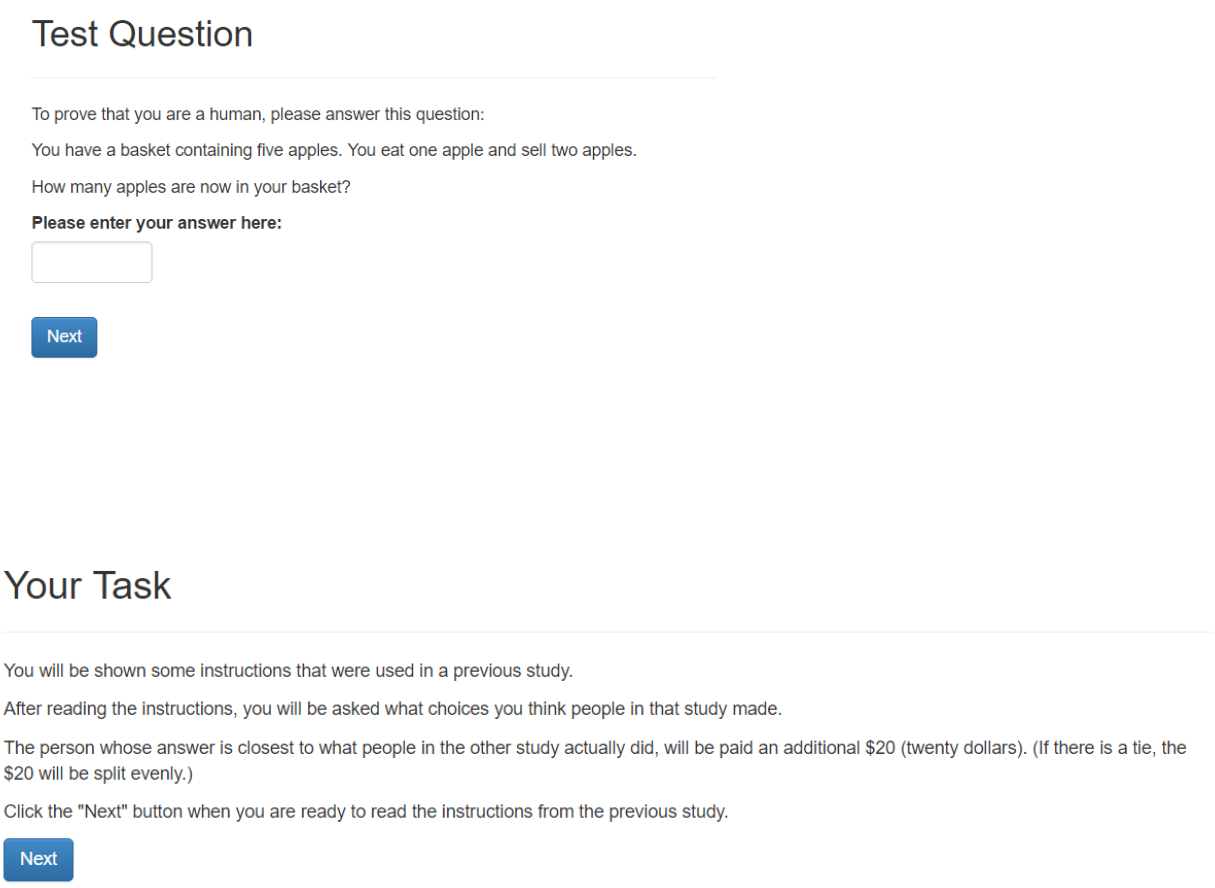
Stockard, Jean, Alphons J. C. Van de Kragt, and Patricia J. Dodge. 1988. "Gender Roles and Behavior in Social Dilemmas: Are There Sex Differences in Cooperation and in Its Justification?" *Social Psychology Quarterly* 51 (2): 154–63. doi:10.2307/2786837.

Swope, K. J., Cadigan, J. Schmitt, and R. Shupp. 2008. "Personality Preferences in Laboratory Economics Experiments." *Journal of Socio-Economics* 37 (3): 998–1009.

Thöni, Christian, and Simon Gächter. 2012. "Peer Effects and Social Preferences in Voluntary Cooperation." *IZA Discussion Paper Series*.

Xiao, Erte, and Daniel Houser. 2007. "Emotion Expression and Fairness in Economic Exchange." Interdisciplinary Center for Economic Science. Fairfax, VA.

Appendix A. Compensated survey interface.



Introduction

Instructions

Two People
You will be randomly paired with a person in another group. No one will learn the identity of the person she/he is paired with.

Starting Amounts
You will start with $1 and the paired person will start with $0.

Your Decision
You will decide how to divide your $1 between how much to send to the paired person and how much to keep. Any amount you send to the paired person will be multiplied by 0.2 by the experimenter. This means that the paired person will receive 20% of the amount sent. The following table shows how this works.

If You Send	The Experimenter multiplies the amount by 0.2	The Paired Person Receives
$0.00	0.2 x $0.00	$0.00
$0.25	0.2 x $0.25	$0.05
$0.50	0.2 x $0.50	$0.10
$0.75	0.2 x $0.75	$0.15
$1.00	0.2 x $1.00	$0.20

Complete Privacy
This experiment is structured so that the other subjects will not know the personal decision of anyone in the experiment.

Next

Question

In the previous study you have just read about:

What was the average amount sent by men?

$

Next

Question

In the previous study you have just read about...

How much do you think should be sent?

$

What is your gender?

- **Male**
- **Female**
- **Prefer not to state**

Next

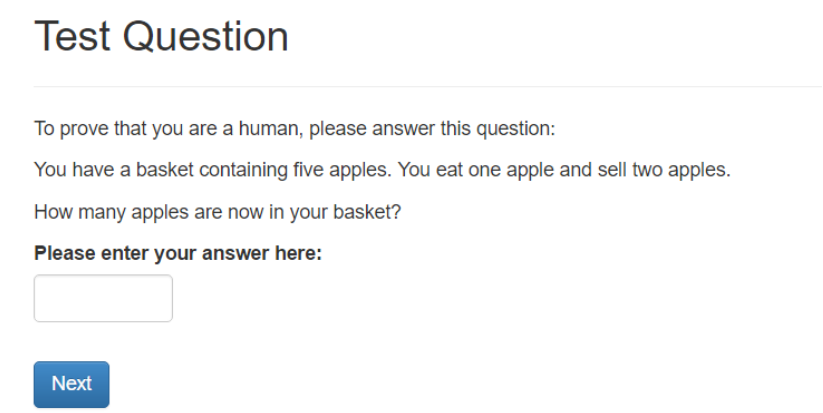
Thank you

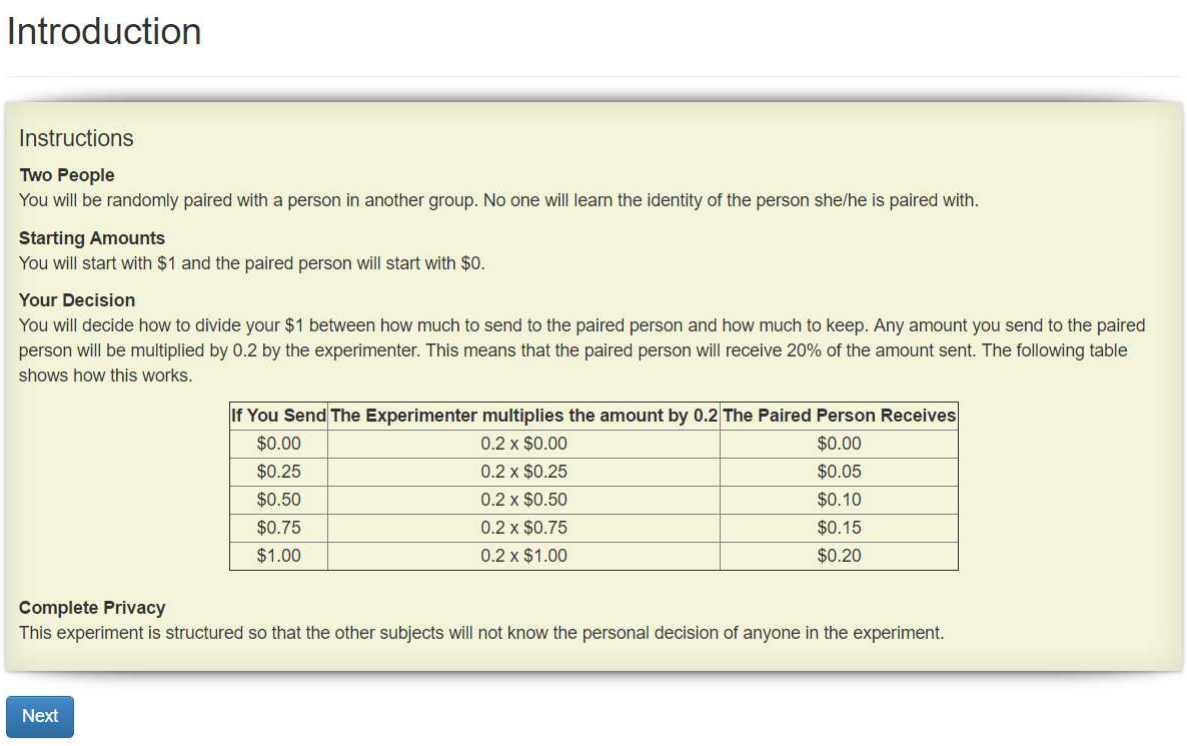
Payments will be calculated and paid within 72 hours.

Thank you for participating!

Next

Appendix B. Controlled experiment interface.





Your Choice

Before you make your decision, here is information about select groups from a previous study. These responses may not be representative of everyone in the previous study.

The most frequent amount that MEN:

ACTUALLY sent	SAID should be sent
$0.50	$0.00

You have $1. Please decide how much to send. Recall that the paired person will receive 20% of the amount sent.

I will send (from 0 to $1):

$

Next

Instructions

Two People
You will be randomly paired with a person in another group. No one will learn the identity of the person she/he is paired with.

Starting Amounts
You will start with $1 and the paired person will start with $0.

Your Decision
You will decide how to divide your $1 between how much to send to the paired person and how much to keep. Any amount you send to the paired person will be multiplied by 0.2 by the experimenter. This means that the paired person will receive 20% of the amount sent. The following table shows how this works.

If You Send	The Experimenter multiplies the amount by 0.2	The Paired Person Receives
$0.00	0.2 x $0.00	$0.00
$0.25	0.2 x $0.25	$0.05
$0.50	0.2 x $0.50	$0.10
$0.75	0.2 x $0.75	$0.15
$1.00	0.2 x $1.00	$0.20

Complete Privacy
This experiment is structured so that the other subjects will not know the personal decision of anyone in the experiment.

Question

What is your gender?

- ○ **Male**
- ○ **Female**
- ○ **Prefer not to state**

Next

Thank you

Payments will be calculated and paid within 72 hours.

Thank you for participating!

Next