Social Norms and Competitiveness: My Willingness to Compete Depends on Who I am (supposed to be)

Zhang, Peilu and Zhang, Yinjunjie and Palma, Marco

Texas AM University, Australian National University, Texas AM University

5 October 2018

Social Norms and Competitiveness: My Willingness to Compete Depends on Who I am (supposed to be)*

Peilu Zhang Yinjunjie Zhang Marco A. Palma†

Working Paper, October 2018

Abstract

Women often respond less favorably to competition than men. In this paper, we test for the effects of social norms on willingness to compete. Subjects compete in two-person teams. In the treatment, one team member is randomly assigned the role of "breadwinner", and the other person is randomly assigned as the "supporter". There are no real differences between the roles in our experiment, except for the framing. These two roles have opposite social norms for competitiveness, reminiscent of gender roles in western society. In the baseline, subjects compete in two-person teams without role assignment. We find women's willingness to compete significantly increases when they are assigned as breadwinners compared to women in the baseline or female supporters. We also find that there is no gender gap in willingness to compete between female breadwinners and males in the baseline. The increase in willingness to compete is mainly contributed by high-ability women. Males are also affected by the role assignment; male supporters are less likely to enter the tournament than male breadwinners. We argue that the changes in willingness to compete are mainly driven by the social norms implied by the two roles.

Keywords: gender gap, gender stereotypes, tournament entry

JEL Codes: A14, B54, C90

*This study was approved by the IRB. IRB Number: IRB2017-0371. Approval date: 06/14/2017. Expiration Date: 06/13/2022.

†Department of Agricultural Economics, Texas A&M University

"By a peculiar weakness of human nature, people generally think too much about the opinion which others form of them." Arthur Schopenhauer[1]

I Introduction

Niederle and Vesterlund (2007) first documented large gender differences in competition entry in a controlled laboratory setting using two different payment schemes: piece rate or tournament. Following Niederle and Vesterlund (2007), laboratory and field experiments largely confirm this finding (for a review see Niederle and Vesterlund (2011)). Economists are increasingly interested in investigating whether such gender differences in competitiveness may be useful for explaining persistent labor market differences. If women are reluctant to compete, then they may also be less likely to seek job promotions or choose more lucrative and competitive fields (Ibarra et al., 2010). Selecting out of certain labor markets is costly for society, especially when competent women are reluctant to compete for positions for which they are the best suited candidate.

In order to understand why the gender gap in competition entry persists, prior experimental research focuses on three aspects: beliefs, risk attitudes and other-regarding preferences. First, while men are typically more confident than women, this only explains part of the gender gap. After controlling for beliefs, a significant gender difference still remains (e.g. Balafoutas and Sutter (2010); Niederle and Vesterlund (2010); Shurchkov (2012). Second, women are more risk averse than men (Eckel and Grossman, 2008). However, gender differences in risk attitudes do not contribute to gender differences in tournament entry (Niederle and Vesterlund, 2007; Große and Riener, 2010). Third, there is little evidence that other-regarding preferences play a big role in explaining the gender gap (e.g. Dreber and Johannesson (2008); Bartling et al. (2009); Dohmen and Falk (2011)). Even after simultaneously controlling for all three factors, a significant gender gap in tournament entry still persists (Kamas and Preston, 2010; Niederle and Vesterlund, 2011).

In this paper we ask whether social identity or social norms play an important role in individual's willingness to compete (hereafter, WTC). Akerlof and Kranton (2000) introduced identity concepts from psychology and sociology into economic behavior. They pose that identity is a sense of belonging to a social category, and it prescribes how people within a category should behave. One

[1]Schopenhauer, A., Saunders, T.B. and Bax, E.B., 1901. *The Wisdom of Life: And Other Essays.* Washington, Dunne.

important application of the social identity model is gender, where man and woman constitute two separate social categories. More importantly, man and woman are associated with different behavioral prescriptions. For example, "men should not do domestic work at home and men should earn more money than their wives."[2] Social identity can influence economic outcomes because deviating from socially prescribed behavior generates a social cost. In psychology, this social cost is known as the "Backlash effect". Rudman (1998) defines *backlash effects* as social and economic reprisals for behaving counter-stereotypically. In this paper, we argue that one reason why women are less competitive is due to prescribed social expectations or social norms. Specifically, women are expected to conform to a supporting wife role and to maternal duties. Men, on the other hand, are expected to compete in the workplace (Blackstone, 2003; Oakley, 2016). Deviating from the established social norms or misaligning with prescriptive stereotypes results in social costs. For example, Bertrand et al. (2015) show that marriage satisfaction is lower in households where the wife's income exceeds the husband's income; in fact, if this is the case, couples are more likely to divorce.[3] Fisman et al. (2006) show that men in the dating market prefer women who are less professionally ambitious. Men also tend to avoid highly educated female partners (Brown and Lewis, 2004; Greitemeyer, 2007; Hitsch et al., 2010). Job promotions increase the likelihood of divorce for women, but not for men (Folke et al., 2016). Bursztyn et al. (2017) argue that social identity costs make single female students demonstrate less ambition in the presence of single males. Although men are also subject to gender-role constraints, the prescriptions of female stereotypes are particularly problematic for women in competitive environments. As noted in the role incongruity theory, men's social roles promote competitiveness, while women's roles penalize it (Eagly and Karau, 2002).

A direct way to close the gender gap in competition is to find ways to encourage women to compete, especially women with high abilities. In previous research, one approach is to assume preferences for competition are malleable. In this regard, Gneezy et al. (2009) find that the gender gap in tournament entry reverses in a matrilineal society compared to patriarchal societies. Single-sex education has been proposed as an alternative to increase the competitiveness of women (Fryer Jr and Levitt, 2010). Booth and Nolen (2009) find that girls in selective single-sex schools

[2]Hochschild and Machung (2012) suggest these prescriptions exist within many households.

[3]Cochard et al. (2018) uses a lab experiment with couples who live together and suggests that there are no differences in preference for work-division in couples.

are more likely to enter competitions against boys than girls from mixed-sex schools. Another approach is to take the gender gap in competitiveness as given and to implement institutional changes designed to encourage women to enter competitive environments. Institutional changes previously studied in the literature include changing the gender composition of competitions (Niederle et al., 2013; Sutter and Rützler, 2010; Datta Gupta et al., 2013), and changing the competitors (Große and Riener, 2010; Apicella et al., 2017), developing a sponsorship program of competition entry (Baldiga and Coffman, 2016), among others.

We conduct a laboratory experiment to investigate competitiveness by changing the institution. Specifically, by randomly assigning different roles or titles, we assign each individual a different social norm for competitiveness than the socially prescribed gender norm. To study how social norms impact WTC, we set up a competition in pairs instead of individual competitions. We introduce in our design WTC scales (0-5) instead of a binary choice for entering the competition.[4] In the treatment, before subjects proceed to choose their WTC level, we exogenously and randomly assign the role of "breadwinner" or "supporter" to each group member. The control has no such role assignment, and subjects directly choose their WTC. The two roles in the treatment were selected based on traditional gender roles in western society. The labels used to describe the roles —"breadwinner" and "supporter"— were selected based on the results of surveys within the same study population.[5] There are no real differences between the two roles in terms of payment, power, responsibility and so on, except for the framing. The two selected roles (framing) have opposite social norms for competitiveness, analogous to gender roles in competitiveness for men and women. We expect these two roles to reduce or raise the social cost of entering the competition, and subsequently, we also expect the WTC of men and women to change depending on their randomly assigned roles.

Consistent with previous literature, our results show that the gender gap in WTC exists in the control group; however there is no gender difference in WTC in the treatment. When women are randomly assigned as breadwinners, their WTC significantly increases compared to women in the baseline or women who are randomly assigned as supporters. More interestingly, the gender gap

[4]This design gives us more information about decision makers' preferences for competition entry, and we compare it to binary choices in the results section.

[5]Before the experiment, we conducted a survey with over 1800 responses to ask respondents to select corresponding labels to the primary-earner and the secondary-earner in a household. The details of the procedure are explained in the following section.

between treated female breadwinners and males in the control completely disappears. The increase in women's willingness to compete is mainly contributed by high-ability women. Men are also affected by the exogenous role assignment; male supporters are less likely to enter the competition compared to male breadwinners. By examining the mechanisms behind our results, we demonstrate that the role assignment does not change subjects' confidence or performance. Instead, by eliciting social norms for competitiveness for each role and gender, we argue that the changes in WTC are mainly driven by the social norms for competitiveness associated with the two randomly assigned treatment roles.

To the best of our knowledge, our paper is the first to study the effects of social norms on the gender gap in tournament entry by exogenously and directly inducing distinctive social norms to each gender in the lab, and the first to use WTC scales instead of a binary choice for competition entry. Dargnies (2011) studies the impacts of social identity on WTC, but she focuses on building group identity during the experiment and testing its effects on WTC. The rest of paper is organized as follows. In section II, we explain how the two social norm roles were selected. Section III shows our theoretical hypotheses. In section IV, we present the experimental design and procedures. The analysis and results are presented in section V. Section VI examines the mechanisms behind the results, and section VII concludes.

II Selecting the Social Norm Labels

In order to test for the effects of social norms on WTC, we exogenously change the social norms associated with competitiveness socially prescribed to each gender by randomly assigning subjects two different titles/roles. The two roles have opposite social norms for competitiveness.

In order to elicit social norm labels pertinent to our subject population, we conducted two rounds of surveys. First, a total of 56 undergraduate students were recruited to provide open-ended social norm labels corresponding to primary and secondary earners in a household. Specifically, the scripts asked:

In one word, how would you call someone who is the primary earner in a household:

In one word, how would you call someone who is the secondary earner in a household:

From these 112 responses, the three most frequently used labels for each role were selected for

the second survey. The second survey consisted of the same two questions using a multiple choice format with the top three labels for each role in the first survey. For the primary earner, the top there labels were Breadwinner, Primary, Provider. For the secondary earner, the top three labels were Supporter, Secondary, and Assistant. A total of 3,794 responses were collected in the second survey with 1,897 responses to each question. The results show that 42% of respondents chose "breadwinner" for the primary earner label, and 55% chose "supporter" for the secondary earner label. Thus, the most popular labels from the two surveys "breadwinner" and "supporter" were selected as the labels for our social norms treatment. Under the social identity theory, a man is supposed to be the breadwinner of his family, while a woman is supposed to be the supporter (Eagly and Kite, 1987). We are interested in documenting how women/men change their competitive behavior when they are assigned the role of breadwinner or supporter.

III Theoretical Framework

In this section we discuss our main hypothesis based on the social image model in Bursztyn and Jensen (2017). Bursztyn and Jensen (2017) assume that if an individual i's utility depends on the posterior expectations held by others in a reference group j after they observe her behavior, then she has social image concerns. Let us assume that an individual i can choose one of two types of visible actions $a_i \in \{0, 1\}$, and each action reveals information about i's type $\sigma_i \in \{l, h\}$. Type h ("high" type) is more socially desirable than type l ("low" type) in reference group j which has n individuals in total. The social image term S_{ij} in an individual i's utility function is:

$$S_{ij} = \lambda_{ij} E_i(\omega_j) P_{n-i}(\sigma_i = h|a_i) \tag{1}$$

where $P_{n-i}(\sigma_i = h|a_i)$ represents the probability that other individuals $(n - i)$ in the reference group j treat i as a type h after observing i's choice of action, a_i. ω_j represents the views of social desirability held in group j, and $\omega_j > 0$ when an individual is seen as type h by group j. The term $E_i(\omega_j)$ corresponds to i's expectation about the social desirability of being observed by group j as type h. Finally, λ_{ij} measures how much individual i cares about the group expectation concerning type h on her. If i does not care about her image or what others think about her type, then $\lambda_{ij} = 0$. Thus, we assume for any non-zero λ_{ij}, individual i cares about her social image in group j and is

therefore subject to social pressure. If $\lambda_{ij} > 0$, then individual i is a conformist, and in contrast $\lambda_{ij} < 0$ means i is a contrarian.

To consider the empirical implications of social image, Bursztyn and Jensen (2017) further embed the social image term S_{ij} above within a random utility model. Assume there is a latent variable $\tilde{a_i}$ capturing the social desirability of action a_i. For example, $\tilde{a_i}$ may represent the social desirability of entering the competition, negotiating a promotion or speaking in public. The complete utility function consists of a social image component, as well as some benefits (B_i) and costs (C_i). B_i and C_i are independent with image concerns, and vary by individuals.

$$\tilde{a_i} = B_i - C_i + \lambda_{ij} E_i(\omega_j) P_{n-i}(\sigma_i = h|a_i) + \epsilon_i \tag{2}$$

where ϵ_i is a random error. To sum up, the social image term in an individual's utility is determined by: $P_{n-i}(\sigma_i = h|a_i)$ which represents the likelihood of others in group j to perceive i as type h after observing her action a_i; how socially desirable being of type h is to the reference group j, ω_j; individual i's expectation about the social desirability of her choice action, $E_i(\omega_j)$; and how much the individual cares about her image or type perceived by the inference group, λ_{ij}. We test the social image concerns through this utility function form.

In our experiment, the action, a_i, is "entering the tournament" or not, $a_i \in \{0, 1\}$. Based on the traditional role of each gender, entering the tournament for women is more likely to be perceived as low type l, and it is socially undesirable. This means $P_{n-i}(\sigma_i = h|a_i)$ decreases and $\omega_j < 0$ if a woman chooses to enter the competition. However, men have the opposite gender role which results in high type h, higher $P_{n-i}(\sigma_i = h|a_i)$ and positive ω_j if a man chooses to compete. Therefore, based on equation (2), entering the tournament generates higher social cost for women, thus lowering their utility, but it increases men's utility.

The two exogenously assigned social norm roles of "breadwinner" and "supporter" also have opposite social norms for competitiveness. It is socially desirable for a breadwinner to enter the competition, but it is socially undesirable for a supporter to enter competitions. By assigning the two roles to each gender, we expect the role of "breadwinner" to reduce women's social cost of entering the competition and therefore increase their willingness to compete. In particular, for a female breadwinner, entering a competition should result in $\omega_j > 0$ and a higher $P_{n-i}(\sigma_i = h|a_i)$.

Similarly, the title of supporter is expected to increase the social cost of entering the competition for men and reduce their willingness to compete.

IV Experimental Design

We conducted a between subject design with a baseline control and a treatment group. Our experiment has two stages. The baseline and treatment group only differ in the second stage. The experiment was conducted at Texas A&M University, and subjects were students who were native English speakers recruited by bulk email. A total of 28 sessions were conducted with eight subjects per session. Nine subjects experienced technical issues that affected their performance in the first stage and their subsequent choice in the second stage, and hence they were excluded from the analysis. Ten subjects were not native English speakers and may have experienced difficulties in understanding their assigned roles, so they were also excluded.[6] The final sample consists of 205 subjects.

First Stage: In the first stage, subjects arrived to the experiment and they were seated in rows. Each session had 8 subjects. Following the original design in Niederle and Vesterlund (2007), subjects had five minutes to individually solve a real effort task under a noncompetitive piece-rate payment scheme. The real effort task consists of adding-up five two-digit numbers. Participants were not allowed to use a calculator, but they could use scratch paper. Participants received 10 cents for each correctly solved problem. Contrary to the original design in Niederle and Vesterlund (2007), we do not reveal the subjects' performance in this stage until the end of the experiment. Since our focus is on the effects of social norms on WTC, we try to avoid potential confounding effects of natural ability on competition entry decisions. The purpose of the first stage is to control for ability without any group or competition context.

Second Stage: After the first stage, each participant was randomly paired with another participant in the same room. We use team competition instead of individual competition so that we can assign different group roles which contain different social norms of competitiveness to each group member. In order to identify the effects of social norms, the partner assignment was blind,

[6] Although in the recruitment process, being a native English Speaker was required, there were still some participants who were not native English speakers who registered for the experiment. In order to have exactly 8 participants in each session, we allowed some non-native English speakers to participate.

and no information about partners was disclosed. Four groups were formed in each session.

In the second stage, each subject had 5 minutes to solve a task similar to the first stage. Before they proceed to the task, participants can choose between two possible payment schemes for their group: piece rate and tournament. In the piece rate scheme, each group receives 20 cents for each correctly answered question.[7] In the tournament scheme, each group receives 80 cents for each correct answer only if they win, otherwise they receive nothing.[8] If a group enters the tournament, then they compete with all the other three groups in the session, regardless of the payment scheme chosen by the other groups. This design feature ensures equivalent expected payments for the two payment schemes, and removes potential concerns of participants competing only against those who may have also chosen the tournament scheme.

For the decision to compete, each group member is asked to select the payment scheme for his group using a six point scale: {0, 1, 2, 3, 4, 5}, which is our willingness to compete measure. Table 1 shows the probability of entering the tournament based on the WTC scale, which is revealed to all subjects. Individuals can choose the piece rate payment scheme by selecting 0 or the tournament by selecting 5. One of the two group members' WTC was randomly selected as the binding WTC for the group. This design feature helps us to eliminate strategic considerations. The WTC selection is anonymous between the two group members, and participants do not know the payment scheme for their groups or their performance in this stage until the end of the experiment.

After selecting the payment scheme, each group proceeds to the second round real effort addition task. Subjects worked separately on the task and then the payment of each group is determined by the combination of the number of correct answers of the two group members. The group earnings were equally divided between the two members. This design feature keeps subjects performance consistent and excludes the possibility of individuals competing on the basis of team performance.

Treatment Interventions: The difference between the treatment and the control group is the randomly assigned role to each group member in the second stage. For the treatment group in the second stage, each group member received the following information before answering the

[7]The expected payment under the piece rate payment scheme between the first and second stage is the same.

[8]In related literature, the usual payment for each correct answer is 50 cents. In our case, if we choose 50 cents in the first stage, then the payment for each correct answer would be $4 under the tournament payment scheme in the second stage. It has been shown that scaling up payments results in a significant increase in risk aversion (Holt and Laury, 2002, 2005), and this impact might be heterogeneous by gender. To reduce the effects of the payment itself on competition entry, we lower our payment for the piece rate scheme to 10 cents in the first stage and 20 cents in the second stage.

Table 1: Willingness to Compete Scale

Willingness to compete scale	Probability to enter the competition
0	0%
1	20%
2	40%
3	60%
4	80%
5	100%

WTC scale question:

You were randomly selected to be the Breadwinner (Supporter) of the group, and your partner is the Supporter (Breadwinner) of the group.

In the control group, there is no such role assignment. As previously mentioned, in order to eliminate other factors that may affect competition entry, we inform subjects that the roles are randomly assigned. Recall that the group payment is equally divided, and one of the two members' WTC is randomly selected as the binding WTC for the group. Given all these design features, the only difference between the two group members is the role framing. As such, we can identify the effects of the social norms implied by the role assignment.

Other-Other Allocation: After the second stage, we conducted a hypothetical other-other allocation. Half of the subjects in the treatment were told that they could hypothetically use money from the experimenter ($3) to allocate between the group members who entered the tournament but did not win (i.e., subjects with $0 earnings in the second stage). Subjects can distribute their allocation in the following way (the sum of the two allocations must equal $3):

Group member: breadwinner, freshman year, economics major.

I allocate $_____ to this student;

Group member: supporter, senior year, engineer major.

I allocate $_____ to this student.

The other half of the subjects in the treatment and all the subjects in the control received the same information, except for the descriptions of the group members, in which we focus on gender instead of the assigned role:

Group member: female, freshman year, economics major.

I allocate $_____ to this student;

Group member: male, senior year, engineer major.

I allocate $_____ to this student.

To avoid experimenter demand effects, in addition to gender and assigned roles, we also included irrelevant information, including academic majors and educational levels as shown above. The irrelevant information was randomly assigned and balanced.

Social Norms Belief Elicitation: We follow Krupka and Weber (2013) to elicit social attitudes toward the competitiveness of each role in our experiment. Using a five-point-scale, subjects were asked to indicate how competitive they believe the following individual should be: a female supporter, a female breadwinner, a woman, a breadwinner, a male supporter, a supporter, a man, and a male breadwinner. The social attitudes task was incentivized. If the participant's answer was the same as the most frequent answer in the session, they received a reward of ¢25. The elicited social norms help to explain how they affect individual WTC. This is further discussed in section VI.

Performance Beliefs and Risk Preferences: Finally, we elicit participants' beliefs about their relative performance in the first task. We ask participants to guess whether they performed better than their second-stage partner. We also ask them to state their beliefs about their overall ranking in the session (1-8). Correct answers were incentivized with a ¢25 reward. The performance beliefs can be used to examine whether changes in WTC induced by the role assignment are derived from changes in confidence arising from the roles.

At the end of the experiment, all participants were asked to complete a risk gamble-choice task following Eckel and Grossman (2008) and a demographic survey. The experiment was computerized using Ztree (Fischbacher, 2007), and had a duration of approximate 30 minutes. Participants received payments from the two tasks and the incentivized questions. Table 2 summarizes the experimental procedures and the number of subjects in each experimental condition. The instructions, questionnaires and surveys are available in the Appendix.

Table 2: Summary of Treatments and Number of Subjects.

	First stage (individually)	Second stage (group, WTC)	Female	Male	F-Breadwinner	M-Breadwinner	Total
Baseline Group	Piece rate	No roles	41	35	—	—	76
Treatment Group	Piece rate	Role assignment	73	56	38	26	129

Notes: The required sample size of breadwinner is 21 for female, and 12 for male based on power calculation.

V Results

In this section, we first examine whether conditional on ability, men and women differ in their preferences for performing under a piece-rate *versus* a tournament scheme when competing in teams. We then examine whether different role assignments change their preferences for competition.

V.1 Willingness to Compete in the Baseline

In our experiment, the WTC was elicited in the second stage. Instead of a Yes or No binary choice, participants were asked to indicate their WTC in a 0-5 scale. Figure 1 shows the mean WTC by gender in the baseline. Using a two-tailed Mann-Whitney U-Test, we find that the WTC of males is significantly higher than females ($p = 0.037$). Recall that, in the baseline, there is no role assignment. Participants have no information about their partners, except that the partner is another participant in the room.[9] We always keep the same number of participants in each session (8). Our result suggests that in the absence of specific social norm roles, men are more likely to choose a competitive tournament environment than women when competing in teams. Throughout the paper the reported test statistics refers to a two-tailed Mann-Whitney U-Test, unless otherwise noted.

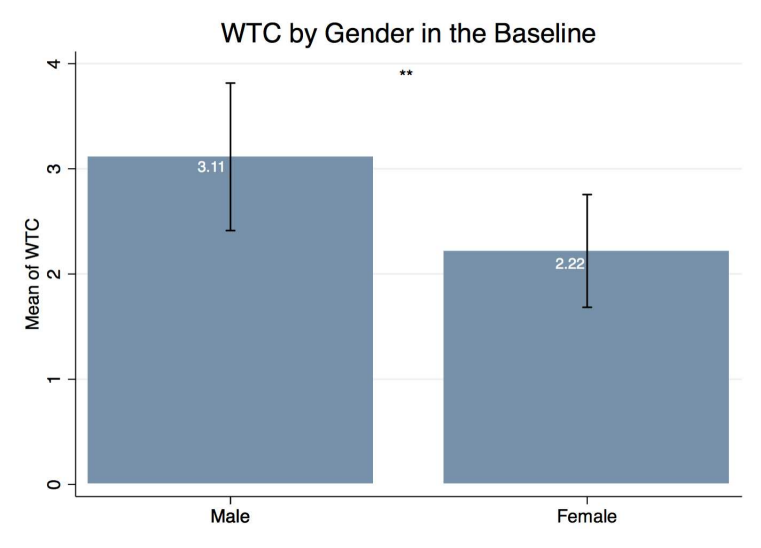


Figure 1: Mean of WTC in the Baseline (with confidence intervals bars)

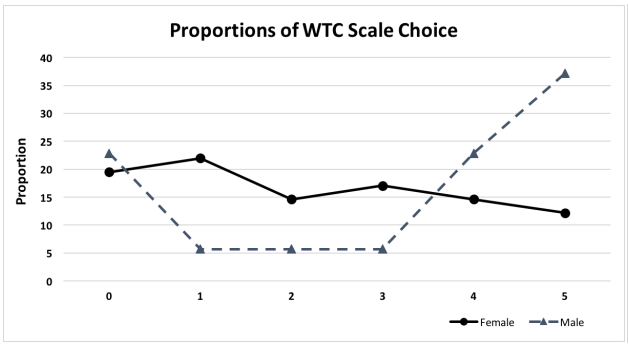
[9] Gender composition of each session is observed by participants, and it is controlled in the following model estimations.

Previous literature using team competitions instead of individual competitions have explored gender differences in tournament entry. Dargnies (2009) finds no gender gap in team competition entry. Healy and Pate (2011) find competing in two-person teams reduces the gender gap by two-thirds. Note that both papers use a binary decision to elicit preferences for competition. In the baseline, if we evaluate $WTC = 0$ and $WTC \neq 0$ instead of our actual WTC scale level $\{0, 1, 2, 3, 4, 5\}$, then there is no gender gap in tournament entry (p = 0.723). This result replicates the findings of Dargnies (2009). According to the corresponding tournament entry probability of each scale number, we further recoded the WTC scales with $0-2$ corresponding to "No" and $3-5$ corresponding to "Yes" in the binary choice. The intuition behind this alternative definition of competition entry is that the probability to enter the competition for subjects who chose 0-2 is 0%-40% (below 50%). Alternatively subjects with a 3-5 scale had a probability to compete of 60%-100% (above 50%). We find that the difference in WTC between man and women is marginally significant (p = 0.059). The results from previous studies and our results recoded to a binary decision suggest that using a binary decision alleviates to a certain degree the gender gap in group competition entry.

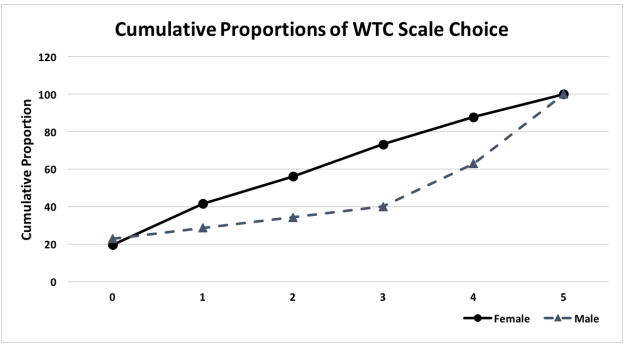
Figure 2 shows that the proportion of men who choose $WTC = 0$ is slightly higher than women (although not significant); however, for subjects with a $WTC \neq 0$, there is a mass concentration for men in the $4-5$ scale. For women, there is a concentration in WTC in the $1-3$ range. We argue that our setting of using a scale to measure the level of WTC reveals people's preferences for competition in more detail. Outside of the laboratory, people can take different degrees of actions between 0 to 1 for competitions. For example, employees can negotiate for a promotion with different degrees of aggressiveness.

V.2 Gender Differences in Performance in the Baseline

In order to rule out individual ability as a driver of WTC, subjects were not informed about their performance in the first stage until the end of the experiment. However, subjects may still form beliefs about their performance before choosing to enter the tournament or not. In the second stage, participants work in teams. Recall that participants do not know the payment scheme applied to their group until the end of the experiment. However, they can speculate about the payment scheme based on their own WTC, especially those with $WTC > 0$. Performance anticipation under



(a) non-cumulative



(b) cumulative

Figure 2: Proportions of WTC Scale Choice by Gender

a group context and a tournament environment may also affect subjects' decisions. Therefore, we examine whether gender differences in performance (ability) exist under the piece rate and tournament payment schemes in the baseline.

Table 3 summarizes the performance of men and women. In the first stage, the average number of correctly solved problems was 9.05 for women and 9.83 for men. The p-values in the last row show that gender differences in performance are not statistically significant. In the second stage, we first show the performance of all participants in order to evaluate whether the group context generates gender differences in performance. The difference in performance under the group context is not significant, which is consistent with Healy and Pate (2011). On average, women correctly solve 9.34 problems, and men 9.40 (p-value = 0.896). Next, we compare the performance of individuals who actually entered the tournament. The difference is still not significant (p-value = 0.823). Previous literature (e.g., Gneezy et al. (2003)) documents that competitions can boost performance; however using Wilcoxon signed-rank tests, we do not find differences in performance between the first and second stage for men or women (see p-values in the last column). This result is reasonable in our case, since performance and payment scheme information was not revealed until the end of the experiment. Given the similarity in performance of men and women, we argue that in the baseline, men have higher WTC than women of similar abilities. Several regression estimations and other potential underlying mechanisms will be explained later.

Table 3: Gender Difference in Performance in the Baseline

	First Stage	Second Stage (all participants)	Second Stage (only those entering comp)	p-value
Men	9.83	9.4	9.08	0.215
Women	9.05	9.34	9.12	0.176
p-value	0.322	0.896	0.832	

V.3 The Impact of Random Role Assignment for the Treatment Group

In this section, we explore how the social norm role assignment impacts participants' WTC. Recall that in the treatment group, before choosing their WTC in the 0-5 scale, subjects were informed that they were randomly selected as the "breadwinner" (or "supporter") of the group. We also informed participants that their partner was assigned the alternative role. Figure 3 shows that the gender gap in WTC decreases by 56% in the treatment group and it is no longer statistically significant (p-value = 0.123). We further analyze whether different roles have asymmetric effects by gender.

First, we compare the WTC of each role to the baseline. Figure 3 shows that the WTC of treated female breadwinners significantly increased compared to women in the baseline (p-value = 0.031). More interestingly, there is no difference in the WTC between treated female breadwinners and untreated men in the baseline (p-value = 0.373). This means that when women are randomly selected as the "breadwinner" of the group and men have no role assignment, the gender gap in WTC disappears. The WTC of treated male "breadwinners" is not statistically different than untreated men in the baseline (p-value = 0.285). We then compare the WTC of treated supporters to the baseline. There are no differences for both genders (p-value = 0.933 for women, p-value = 0.181 for men). Although the difference in WTC between male supporters and men in the baseline is not significant, the decrease in male supporters' WTC closes the gender gap in WTC between male supporters and females in the baseline (p-value = 0.436).

We further compare the WTC within the role assignment (*breadwinner* vs *supporter*). Figure 3 suggests that for both genders, breadwinners' WTC is significantly higher than the WTC of supporters (p-value = 0.049 for women, p-value = 0.016 for men). When we compare the WTC of each role across gender, we find that there is no gender gap in WTC between males and female "supporters" (p-value = 0.513); however male breadwinners are more willing to enter the competition than

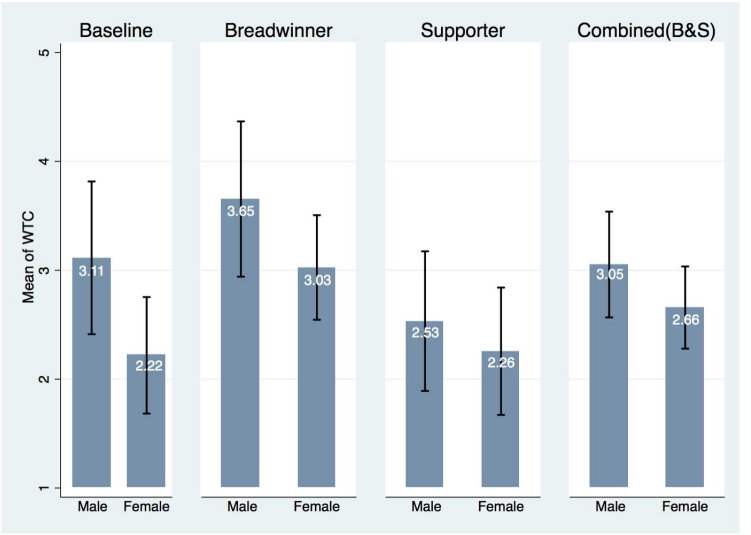


Figure 3: Comparisons of Mean of WTC (with confidence intervals bars)

female breadwinners (p-value = 0.032). There is no difference in WTC between female breadwinners and male supporters (p-value = 0.233). Female breadwinner's WTC is actually higher than male supporter's WTC. Male breadwinner's WTC is statistically higher than female supporter's WTC (p-value = 0.002).

Table 4 summarizes performance in the treatment and it shows that the results are similar to the baseline results presented in section V.2. The p-values in the last row indicate that there are no gender differences in performance in the first and second stage. Using Wilcoxon signed-rank tests, we evaluate the performance between the first and second stage for each role and gender, and the p-values are shown in the last two columns. The results suggest that the role assignment does not change the performance for men or women.[10] This result suggests that the changes in WTC induced by the treatment (role assignment) are not likely due to changes in performance induced by the roles.

Table 4: Gender Difference in Performance in the Treatment Group

	First Stage (individual)	Second Stage (group)	Breadwinner	Supporter	P-value (Breadwinner)	P-value (Supporter)
Men	9.80	9.46	8.5	10.3	0.497	0.811
Women	8.86	8.97	9.32	8.6	0.286	0.441
P-value	0.202	0.163				

[10]Overall, the changes in performance between the first and second stage are not significant (p-value = 0.740 for women, p-value = 0.553 for men).

We estimate ordered probit models for the control and treatment groups. The dependent variable is WTC. The number of correct answers in the first stage is used to control for ability (1st Stage). We also control for confidence, risk preferences and the gender composition of each session. Confidence is indicated by the relative performance beliefs of participants. Risk preferences are indicated by the choice of Eckel and Grossman (2008) incentivized risk gamble task, which has 6 choices (1-6). A higher number means more risk-tolerant behavior (Risk). Gender composition indicates the number of women in each session (No. Female). Table 5 suggests that in the baseline, risk aversion contributes to explain the gender gap in WTC. Risk-tolerant participants are more willing to compete in the baseline. In columns (3) and (4), we pool all data in the treatment, including men and women breadwinners and supporters. First, consistent with figure 3, there are no significant effects for gender. The positive and significant breadwinner coefficient shown in column (4) suggests that the "breadwinner" label makes people more willing to enter the competition compared to the "supporter" label. Columns (5) to (8) display the regression results for women and men in the treatment separately. Compared to "supporters", "breadwinners" have a positive and significant effect on both women's and men's WTC. The male results shown in column (8) suggests that the increase in men's WTC can be explained by their confidence and risk preferences. Furthermore, by focusing only on the comparison between breadwinners and untreated participants in the baseline (columns 9-12), we find that the "breadwinner" label has a larger (and significant) effect on the WTC of females. These results are robust to the inclusion of several control variables. Compared to untreated men in the baseline, there are no significant effects on the WTC of male "breadwinners". These results are all in line with the conclusions derived from Figure 3.

In columns (1) and (2) of Table 6, we estimate an ordered probit model to investigate the gender gap in WTC between female breadwinners and males in the baseline. The dependent variable is WTC. Columns (3) and (4) present the results for male "supporters" and untreated females in the baseline. The results show that there are no gender effects on the WTC across all the estimations. This result suggests that the gender gap in WTC disappears when comparing female breadwinners and males with no role assignment; or when comparing male supporters and females with no role assignment.

We also estimate difference-in-difference models to investigate whether the same role has asymmetrical effects on the WTC of men and women. The results are shown in Table 7. All the

Table 5: Regression Analysis of Ordered Probit Model of WTC.

	Baseline (1)	Baseline (2)	Treatment (3)	Treatment (4)	Treatment (5)Female	Treatment (6)Female	Treatment (7)Male	Treatment (8)Male	Female (9)	Female (10)	Male (11)	Male (12)
Female	-0.498^{*}	-0.390	-0.308	-0.335								
	(0.264)	(0.277)	(0.193)	(0.218)								
Breadwinner				0.401^{**}	0.485^{*}	0.361	0.696^{**}	0.442	0.494^{**}	0.514^{**}	0.336	0.429
				(0.203)	(0.251)	(0.273)	(0.311)	(0.333)	(0.242)	(0.254)	(0.281)	(0.294)
1st Stage		0.014		-0.037		-0.013		-0.061^{*}		-0.002		0.023
		(0.037)		(0.026)		(0.038)		(0.035)		(0.035)		(0.044)
Confidence		0.388		0.335		0.076		0.744^{**}		0.020		0.840^{**}
		(0.300)		(0.241)		(0.333)		(0.376)		(0.340)		(0.334)
Risk		0.142^{*}		0.201^{***}		0.116		0.328^{***}		0.145^{**}		0.076
		(0.079)		(0.069)		(0.091)		(0.104)		(0.073)		(0.089)
No. Female		-0.099		0.132		0.166		0.076		0.012		-0.110
		(0.113)		(0.082)		(0.131)		(0.115)		(0.116)		(0.123)
Observation	76	76	129	129	73	73	56	56	79	79	61	61
Pseudo R^2	0.015	0.043	0.006	0.064	0.015	0.030	0.030	0.122	0.016	0.030	0.008	0.060

Notes: Dependent variable is WTC. Robust standard errors in parentheses. (9)-(12) include only breadwinners in the treatment and all the subjects in the baseline. 1st Stage refers to performance in the 1st stage. Confidence refers to a dummy that takes on the value 1 for subjects who believed that they performed better than their partners. NO. female refers to the number of female in each session. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 6: Ordered Probit Estimations of Gender Gap in WTC

	(1) FBread & MBase	(2) FBread & MBase	(3) MSupporter & FBase	(4) MSupporter & FBase
Female	-0.173	0.083	-0.206	-0.370
	(0.260)	(0.326)	(0.252)	(0.261)
1st Stage		0.046		-0.049
		(0.042)		(0.033)
Confidence		0.476		0.322
		(0.365)		(0.315)
Risk		0.012		0.323***
		(0.078)		(0.086)
No. Female		-0.189		0.050
		(0.147)		(0.101)
Observation	73	73	71	71
R-square	0.002	0.032	0.003	0.074

Notes: Dependent variable is WTC. Robust standard errors are in parentheses. 1st Stage refers to performance in the 1st stage. Confidence refers to a dummy that takes on the value 1 for subjects who believed that they performed better than their partners. NO. female refers to the number of female in each session. $*p < 0.10, **p < 0.05, ***p < 0.01$

regression specifications control for ability, confidence, risk preferences and gender composition of the group. In column (1), we exclude the data of "supporter", and the dummy variable *treatment* takes the value of 1 if the treatment assignment is *breadwinner* and 0 for the *baseline*. In column (2), the "breadwinner" data is excluded, and *treatment* is 1 if the role is *supporter* and 0 for the *baseline*. In column (3), we only use the data of treatment group, and *treatment* is 1 if the role is *breadwinner* and it is 0 if the role is *supporter*. The coefficients for "Treatment" suggest that for males, there is no difference in the WTC between the assigned roles and the baseline, but male breadwinners' WTC is higher than the WTC of male supporters (shown in column 3). The coefficients for "Female" indicate that in the baseline, females' WTC is lower than males' WTC, however there is no difference in WTC between female supporters and male supporters. These results are consistent with the conclusions derived from Figure 3. The interaction terms show that there is no differential effect of assigned roles across gender.

To sum up, we have the following findings:

(1) Compared to the baseline, there is a big increase in women's WTC when they are randomly assigned as breadwinners. This increase is large enough to close the gender gap in WTC between female breadwinners and untreated males in the baseline. There is no difference in the WTC between female supporters and untreated females in the baseline.

(2) Compared to the baseline, there is no change in the WTC of male breadwinners or male

Table 7: OLS Regressions for the Difference-in-Difference Estimates of WTC.

	(1) Breadwinner & Base	(2) Supporter & Base	(3) Breadwinner & Supporter
Treatment	0.527	-0.345	0.775*
	(0.467)	(0.458)	(0.436)
Female	−0.747*	−0.777*	-0.200
	(0.440)	(0.449)	(0.404)
Treated * Female	0.303	0.725	-0.356
	(0.584)	(0.576)	(0.558)
Constant	1.633*	1.412	0.791
	(0.854)	(0.870)	(0.819)
Observation	140	141	129
R-square	0.156	0.153	0.202

Notes: Dependent variable is WTC. Robust standard errors are in parentheses. All regressions control for ability, confidence, risk and gender composition.

supporters.

(3) When comparing within the social norm treatment (i.e., breadwinners and supporters), supporters are less likely to enter the tournament compared to breadwinners for both males and females. The decrease in the WTC of male supporters closes the gender gap between male supporters and untreated females in the baseline.

VI Exploring the Mechanisms Behind the Changes in WTC in the Social Norm Treatment

In this section, we explore the potential mechanisms behind the results presented above. We first ask whether the title of "breadwinner" increases the confidence of women or change their risk aversion. We then test for the effects of social norms. Next, we separate women based on their performance in order to explore the role of performance in explaining the changes in WTC when women are assigned as "breadwinners".

VI.1 Changes in Confidence and Risk Aversion

Recall that in the belief elicitation stage, we asked subjects to guess whether they believe they solved more questions correctly than their partner during the first stage. We also asked subjects to guess their performance rank (1-8) in the first stage. Both questions were incentivized.

Table 8 shows the proportion of subjects who believe they performed better than their partners

Table 8: Proportion of subjects who believe they performed better than their partner (confidence)

	Baseline	Breadwinner	Supporter	p-value Base vs BW	p-value Base vs SP	p-value Base vs SP
Men	80.0%	73.1%	83.3%	0.528	0.732	0.355
Women	78.1%	78.9%	71.4%	0.923	0.509	0.460
P-Value	0.836	0.589	0.260			

in the first stage. Although it has been documented in the literature that women are less confident than men (Niederle and Vesterlund, 2007), the p-values in the last row show there is no gender difference in confidence within the blind group context of our experiment.[11] The p-values in the last three columns suggest no differences in the confidence between the roles by gender. The beliefs for guessing the rank show similar results. We further investigate whether confidence effects on the WTC change by the role assignment using a difference-in-difference approach. We regress WTC on a dummy variable that takes the value of 1 if there is a role assignment and 0 for the baseline. The results are shown in columns (1) and (2) of Table 9. The interaction terms are not significant, which confirms our result that social norm role assignment did not change the effects of confidence on WTC.

Risk preferences were measured using the Eckel and Grossman (2008) gamble choice task. We replicate the robust findings in the literature that women are more risk averse than men ($p - value < 0.01$). We further investigate whether risk changes by the role assignment using a OLS approach, and the results are presented in columns (3) and (4) of Table 9. The results are similar to the confidence results. There are no significant interaction terms, suggesting that risk aversion has no differential impacts on the WTC of different groups. We conclude that the changes in WTC are not likely driven by changes in confidence or risk aversion.

VI.2 The Effects of Social Norms

We follow Krupka and Weber (2013) to elicit social norms for the competitiveness of each role and gender. Participants were asked to:

[11] There are several differences of the belief elicitation question between Niederle and Vesterlund (2007) and our experiment. They asked the subjects to guess their relative performance in the tournament (i.e., to their competitors), but we ask the relative performance to partners; they asked subjects to guess the rank (1-4), but our questions is a "Yes" or "No" question.

Table 9: OSL Estimates of Confidence and Risk on WTC.

	(1) Breadwinner & Base	(2) Supporter & Base	(3) Breadwinner & Base	(4) Supporter & Base
Treatment	0.871	0.481	1.202	0.649
	(0.689)	(0.565)	(0.788)	(0.761)
Confidence	0.745	0.874*		
	(0.523)	(0.515)		
Treat * Confidence	-0.222	-0.577		
	(0.740)	(0.683)		
Risk			0.258**	0.271**
			(0.124)	(0.128)
Treat * Risk			-0.125	-0.202
			(0.168)	(0.194)
Constant	1.420	1.074	1.280	1.610*
	(0.883)	(0.898)	(0.880)	(0.872)
Observation	140	141	140	141
R-square	0.155	0.148	0.158	0.151

Notes: Dependent variable is WTC. Robust standard errors in parentheses. All regressions control for gender, risk/confidence, performance in the first stage and gender composition in each session.

"Please indicate in a 1-5 scale how competitive you think a _____ should be. If your answer is the same as the most frequent answer in the experiment, you will receive a payment of 25 cents."

The social norms were elicited for the following: woman, female supporter, female breadwinner, breadwinner, man, male supporter, supporter, and male breadwinner in a household. Figure 4 shows the mean responses for the social norms of each role and gender in the treatment group.[12] Using Wilcoxon signed-rank tests, we find that most participants believe females should be less competitive than males (p-value= 0.052). The social norms for the competitiveness of "Supporters" are lower than "Breadwinners" ($p - value < 0.01$). Adding the two labels to the gender closes the gender gap of social norms for competitiveness (p-value = 0.470 for Breadwinner, p-value = 0.132 for Supporter). We further find that the social norms for competitiveness of both female and male supporters are lower than females and males without any labels ($p - value < 0.01$ for both genders). In contrast, most subjects believe that the competitiveness of female and male breadwinners should be higher than men and women without labels ($p - value < 0.01$ for both gender). Interestingly, the results are consistent when considering the social norms reported by

[12]In the control group, subjects had not seen the two labels until this section, and some subjects commented that they did not understand the label "Supporter" until they saw "Breadwinner" in the following question. Thus, we focus on the elicitation of social norms in the treatment. Including the social norms elicited in the control does not change the results (see Figure A1).

men and women separately (see Figure A2). From these results, we conclude that the role of "supporter" decreases the social desirability for entering into competition. This means that the social cost of entering the competition is higher for a supporter, regardless of whether the supporter is a man or a woman. However, breadwinners —male or female— have a social norm expectation to be competitive. This question captures $E_i(\omega_j)$ in the theoretical model, which corresponds to i's expectation about the social desirability of being observed by group j as a high type h. These results are aligned with our main hypothesis described in the theoretical framework. The social cost of entering the competition for a female breadwinner is lower than that of a woman.

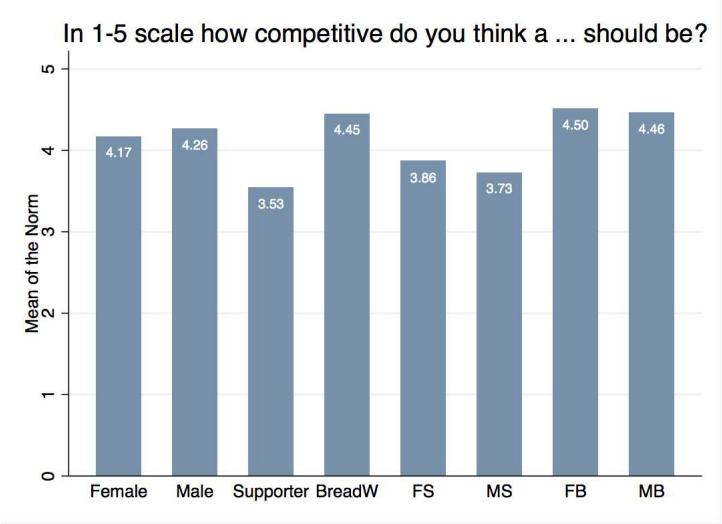


Figure 4: Elicited Social Norms for Competitiveness in Treatment

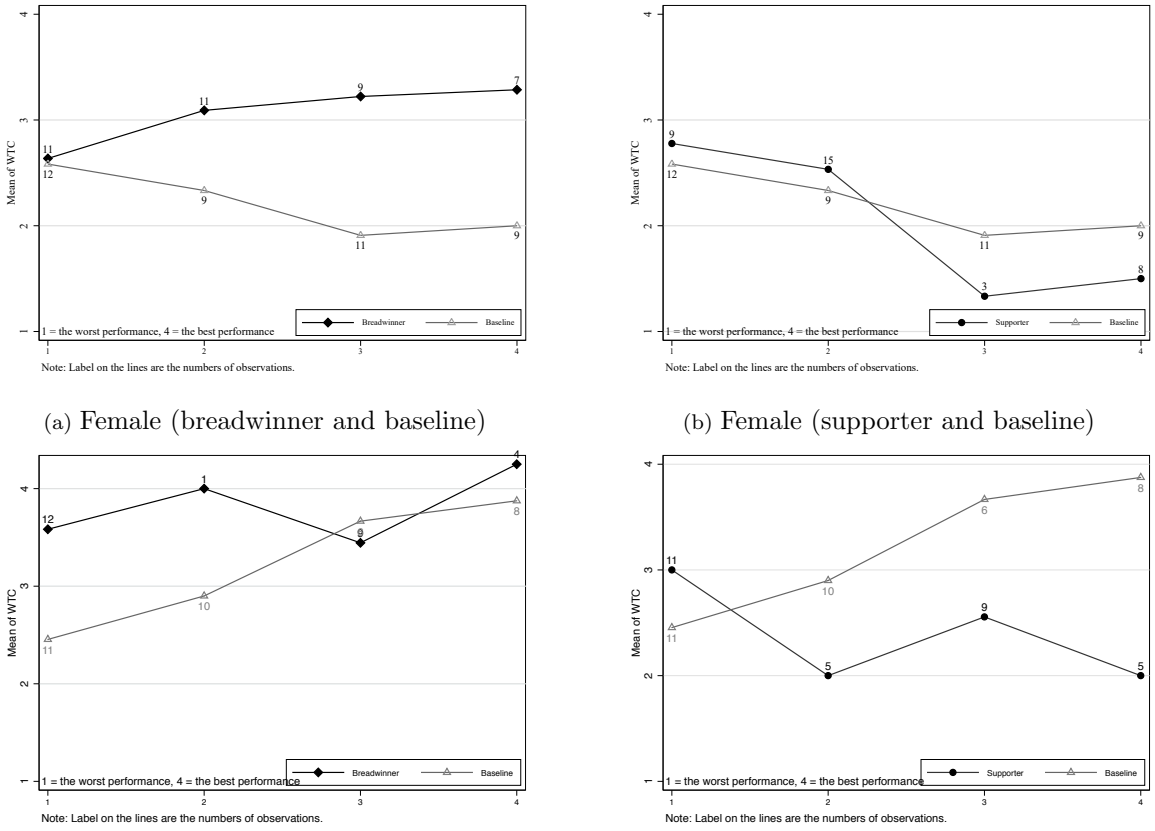
Recall that after the second stage, subjects were asked to do other-other allocations (using the experimenter's money) to participants who entered the competition, but lost the competition and received \$0 payments. Both females and males allocate more to females (Wilcoxon signed-rank tests, p-value $= 0.033$ for women, p-value $= 0.051$ for men).[13] Participants allocate more money to supporters compared to breadwinners, but the difference is not significant (see Figure A3, A4).[14] To some extent, the allocation results in our experiment suggest that women and supporters have a similar status in competition as perceived by all the participants.

[13] Our results are in line with the dictator games literature which find women tend to receive more money as recipients (Engel, 2011).

[14] Previous literature shows a tendency to distribute more to poor recipients in dictator games (e.g., Eckel et al. (2005); Brañas-Garza (2006)).

VI.3 WTC by Performance

The purpose of our research is to evaluate whether women, particularly women with high ability, can be encouraged to become more competitive. In this section, we explore the WTC of women by performance. In particular, we are interested in the linkage between the increase in WTC and performance. If the increase in WTC is driven by high-ability women, then the results of our experiment have important policy implications for social efficiency.



(a) Female (breadwinner and baseline)

(b) Female (supporter and baseline)

(c) Male (breadwinner and baseline)

(d) Male (supporter and baseline)

Figure 5: WTC by Performance Quantiles

Figure 5 shows the mean WTC for women (panels a and b) and man (panels c and d) conditional on their performance quartile during the first stage.[15] Panel a of Figure 5 shows the results for the baseline and female breadwinners. The graph shows that the increase in WTC of female breadwinners is mainly contributed by high-performing women. The difference in WTC between the baseline and female breadwinners in the first two best quantiles (3&4) is significant (p-value =

[15] We use performance in the first stage since there is no group pairing effect in the first stage, hence, the performance in the first stage is the performance before participants make the tournament entry decision.

0.040). In panel b of Figure 5, we compare the WTC by performance of women in the baseline and women supporters. As expected we do not find similar patterns of changes in WTC by performance quantiles in this case. We also show the WTC by performance quantile for males (see Figure 5-c & 5-d). We do not find similar results for male breadwinners, but we find that the WTC of male supporters in the top two performance-quantile is lower than males in the baseline within the same quantiles (p-value = 0.048). This means the decrease in WTC of male supporters is also mainly contributed by high-performing men.

VII Concluding Remarks

PricewaterhouseCoopers (PwC) recently surveyed over 3,600 professional women (aged 28-40) to find out about their career development experiences and aspirations.[16] The results of the survey show that 82% of women are confident in their ability to fulfill their career aspirations and 73% are actively seeking career advancement opportunities. However, 42% feel nervous about how starting a family might impact their careers. Affected by traditional social norms, women are expected to spend more time in child care and household activities, and hence 38% of surveyed women saw motherhood and flexibility as a penalty on their careers. As a result, reducing the limits of social norms for competitiveness on women becomes important. Especially for high-ability women, since it is costly for themselves, their families and society if they don't pursue professional positions for which they are the best suited candidate. PwC's global chairman Bob Moritz states "It must go hand in hand with efforts to mitigate any unconscious biases and gender stereotypes that have traditionally impacted career success and progression in workplaces around the world".

However, gender stereotypes remain extremely powerful because they are evoked by highly visible, biological characteristics and they are deeply rooted in culture. Changing stereotypes is almost impossible during a short period of time, but other social norms for females can be promoted to mitigate the effects of traditional gender stereotypes and change their behavior. We use a simple experiment and provide evidence that females' willingness to compete is limited by social norms. We show females' WTC significantly increases when they are randomly assigned the role of "breadwinners". Although the role is randomly assigned, and there is no real responsibility or power

[16] PricewaterhouseCoopers (doing business as PwC) is one of the Big Four auditors in the world. The report of the survey can be found at www.pwc.com/timetotalk.

bestowed on this role, it still gives women a "nudge" to surpass the limits of traditional gender roles in competition entry. This result is further reinforced by examinations of the underlying mechanism behind the results: we do not find changes in subjects' confidence or performance. Becoming the "breadwinner" reduces the social cost or backlash effects of entering the competition for women. According to our social norm elicitation, under the title of "breadwinner", it is socially desirable for women to enter the competition. The role assignment also affects males. When men are randomly assigned as the "supporter", their WTC decreases.

Importantly, the increase in female breadwinners' WTC is mainly contributed by high-ability women. This means that relaxing social norm constraints encourage women with high ability to enter the competition. Our random role assignment improves efficiency in terms of improving the match of ability and competition entry. Interestingly, at the end of the experiment, we ask subjects which role they prefer, breadwinner or supporter, and we found men and women both prefer to be breadwinners. The effects of the role nudge in our experiment imply that restructuring institutions, such that assigning leadership encouraging titles to women, have the potential to reduce gender disparities in economic and labor market outcomes.

In our experiment, for simplicity and clean identification, we do not disclose any information about partners. Future work may investigate how subjects change their WTC when the gender of the other group member (role) is disclosed to them, or when the role is assigned based on performance. It is also possible that social norms may have heterogeneous effects on women of different ages and women in different cultures, which can be another venue for future work.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Apicella, C. L., E. E. Demiral, and J. Mollerstrom (2017). No gender difference in willingness to compete when competing against self. *American Economic Review 107*(5), 136–40.

Balafoutas, L. and M. Sutter (2010). Gender, competition and the efficiency of policy interventions.

Baldiga, N. R. and K. B. Coffman (2016). Laboratory evidence on the effects of sponsorship on the competitive preferences of men and women. *Management Science*.

Bartling, B., E. Fehr, M. A. Maréchal, and D. Schunk (2009). Egalitarianism and competitiveness. *American Economic Review 99*(2), 93–98.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics 130*(2), 571–614.

Blackstone, A. M. (2003). Gender roles and society.

Booth, A. L. and P. Nolen (2009). Gender differences in competition: the role of single-sex education. Technical report, CEPR Discussion Paper 7214.

Brañas-Garza, P. (2006). Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior & Organization 60*(3), 306–320.

Brown, S. L. and B. P. Lewis (2004). Relational dominance and mate-selection criteria: Evidence that males attend to female dominance. *Evolution and Human Behavior 25*(6), 406–415.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'acting wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Cochard, F., H. Couprie, and A. Hopfensitz (2018). What if women earned more than their spouses? an experimental investigation of work-division in couples. *Experimental Economics 21*(1), 50–71.

Dargnies, M.-P. (2009). Does team competition eliminate the gender gap in entry in competitive environments?

Dargnies, M.-P. (2011). Social identity and competitiveness.

Datta Gupta, N., A. Poulsen, and M. C. Villeval (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry 51*(1), 816–835.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–90.

Dreber, A. and M. Johannesson (2008). Gender differences in deception. *Economics Letters 99*(1), 197–199.

Eagly, A. H. and S. J. Karau (2002). Role congruity theory of prejudice toward female leaders. *Psychological review 109*(3), 573.

Eagly, A. H. and M. E. Kite (1987). Are stereotypes of nationalities applied to both women and men? *Journal of Personality and Social Psychology 53*(3), 451.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1–17.

Eckel, C. C., P. J. Grossman, and R. M. Johnston (2005). An experimental test of the crowding out hypothesis. *Journal of Public Economics 89*(8), 1543–1560.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics 14*(4), 583–610.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics 121*(2), 673–697.

Folke, O., J. Rickne, et al. (2016). All the single ladies: Job promotions and the durability of marriage. Technical report.

Fryer Jr, R. G. and S. D. Levitt (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics 2*(2), 210–40.

Gneezy, U., K. L. Leonard, and J. A. List (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica 77*(5), 1637–1664.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Greitemeyer, T. (2007). What do men and women want in a partner? are educated partners always more desirable? *Journal of Experimental Social Psychology 43*(2), 180–194.

Große, N. D. and G. Riener (2010). Explaining gender differences in competitiveness: Gender-task stereotypes. Technical report, Jena economic research papers.

Healy, A. and J. Pate (2011). Can teams help to close the gender competition gap? *The Economic Journal 121*(555), 1192–1204.

Hitsch, G. J., A. Hortaçsu, and D. Ariely (2010). Matching and sorting in online dating. *American Economic Review 100*(1), 130–63.

Hochschild, A. and A. Machung (2012). *The second shift: Working families and the revolution at home*. Penguin.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Holt, C. A. and S. K. Laury (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review 95*(3), 902–912.

Ibarra, H., N. M. Carter, C. Silva, et al. (2010). Why men still get more promotions than women. *Harvard Business Review 88*(9), 80–85.

Kamas, L. and A. Preston (2010). What can social preferences tell us about charitable giving? evidence on responses to price of giving, matching, and rebates. In *Charity with Choice*, pp. 165–199. Emerald Group Publishing Limited.

Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association 11*(3), 495–524.

Niederle, M., C. Segal, and L. Vesterlund (2013). How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science 59*(1), 1–16.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives 24*(2), 129–44.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annu. Rev. Econ. 3*(1), 601–630.

Oakley, A. (2016). *Sex, gender and society*. Routledge.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology 74*(3), 629.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association 10*(5), 1189–1213.

Sutter, M. and D. Rützler (2010). Gender differences in competition emerge early in life.

Appendix.

A1.Instructions

WELCOME

In the experiment today, you will be asked to complete two different tasks. None of these will take more than 5 minutes. At the end of the experiment we will pay you based on your performance in these two tasks. The method we use to determine your earnings varies across tasks. Before each task we will describe in detail how your payment is determined.

Your total earnings from the experiment are the sum of your payment from the two tasks.

New page:

Task 1

For Task 1 you will be asked to solve a set of problems independently. You will be given 5 minutes to answer these problems. For calculation questions, you cannot use a calculator, however you are welcome to write the numbers down and make use of the provided scratch paper. Your answers to the problems are anonymous.

You get 10 cents per problem you solve correctly. Your payment does not decrease if you provide an incorrect answer to a problem. We refer to this payment as the piece rate payment. You will not be informed of how many questions you correctly solve and your corresponding payment in this task until the end of the experiment.

Please do not talk with one another for the duration of the experiment. If you have any questions, please raise your hand.

New page:

Task 2

For task 2 you will be randomly paired with a partner (one of the participants in this room, but you will not be told who she/he is), and you two will work as a group anonymously. Four groups in total will be formed in this section. As in Task 1 each group will be given 5 minutes to solve a similar set of problems. The sum of answers that the two members correctly solve will be the final number of correct answers of the group. The final payment of each group in this task will be equally split between the two group members.

New page:

Your payment in this task will also depend on your choice of payment scheme: piece rate or tournament. If you choose *piece rate*, your group will be paid in the same way as Task 1, but with 20 cents per problem you solve correctly.

If you choose *tournament*, your group will be paid based on the performance of your group relative to that of the other three groups, regardless the payment scheme chosen by the other groups. For instance, if the number of problems solved correctly in your group is the largest among all the four groups, your group will receive **80** cents per correct answer, four times the payment from the piece rate; otherwise your group will receive no payment for this task. If there are ties, all these winner groups will receive 80 cents per correct answer.

The choice and answers that you and your partner made will keep anonymous. You will not be informed of the corresponding payment of your group in this task until the end of the experiment.

New page:

Now, you were randomly selected to be the *BreadWinner* of your group, and your partner is the *Supporter*. (only for treatment)

Please indicate in 5-scale how much you want to choose the tournament payment scheme for your group. Your partner will also be asked this same question.

Your and your partner's choices will be randomly selected to be the final choice of your group.

You (your partner) will not know the choice made by your partner (you) in this task. You will be informed of the final payment scheme of your group and your corresponding payment in this task at the end of the experiment.

New page:

Please indicate in 5-scale how much you want to choose the tournament payment scheme for your group. If you choose scale 1, and your choice is selected as your group's choice, then the probability of entering the tournament will be 20% (i.e., 80% of probability to use piece rate); if you choose scale 2, the probability of entering the tournament will be 40%. Scale 5 means entering tournament for sure.

Scale	Description
0	Piece rate
1	Entering the Tournament with probability 20%.
2	Entering the Tournament with probability 40%.
3	Entering the Tournament with probability 60%.
4	Entering the Tournament with probability 80%.
5	Entering the Tournament for sure.

Your choice is:

New page:

In this stage, if the experimenter has 3 dollars, and now you can use this $3 (not your money) to allocate money to the group members who enter the tournament but did not win the competition (i.e., receive nothing in the above stage).

The sum of your allocations to the two group members must be equal to $3.

New page:

Please indicate in the five-scale how competitive do you think a woman should be if she is the supporter in her household? If your answer is the same as the most frequent answer for this question in today's experiment, then you will get extra 25 cents.

New page:

Do you think that your performance in Task 1 is better than your partner? If your guess is correct, then you will get extra 25 cents.

Yes

No

New page:

Please estimate the rank of your performance relative to the rest of the participants in Task 1? For example, if you check 1, you consider yourself as better than all other seven participants. If

your guess is the same as your rank, then you will get extra 25 cents.

1, 2, 3, 4, 5, 6, 7, 8

New page:

If you were given $100 and had an opportunity to invest in a risky asset which has a 50% chance of earning 2.5 times of your investment and 50% of earning zero, how much of the $100 will you invest?

New page:

If you can choose your role in a household, what is your option?

Breadwinner

Supporter

New page:

Please make a choice among the following six gamble choices. Each choice has two events A and B, and each event's chance of occurring is 50%. After you make your choice, the system will randomly choose an event for that choice, you will get the corresponding payoff for that event. Please select the gamble that you choose.

New page:

Do you think men or women generally do better in the "math problem" task that you just did?

Demographic survey

Please answer the following survey questions.

(1) Please enter your age in years.

(2) Please indicate the HIGHEST level of education you have completed.

Some High School or less

High School Diploma

Some College

2 year/Associates Degree

4 year/Bachelor's Degree

Some Graduate School

Graduate Degree

(3) Including yourself, how many people live in your household?

(4) Please indicate your gender.

Male

Female

(5) Please indicate your race.

Asian/ Pacific Islander

African American

Caucasian/ White

Native American/ Indigenous

Hispanic

Other (Please list below)

(6) Please indicate your household yearly income for 2016. (Include all forms of income, including salary, interest and dividend payments, tips, scholarship support, student loans, parental support, and allowance)

Less than $30,000

$30,000 - $39,999

$40,000 - $49,999

$50,000 - $59,999

$60,000 - $69,999

$70,000 - $79,999

$80,000 - $89,999

$90,000 - $99,999

$100,000 - $149,999

$150,000 or more

(7) Your major?

Word Selection Survey

Survey1

In one word, how would you call someone who is the primary earner in a household: _____

In one word, how would you call someone who is the secondary earner in a household: _____

Survey 2

1. In one word, how would you call someone who is the primary earner in a household:

Breadwinner

Primary

Provider

Other (please list) _____

2. In one word, how would you call someone who is the secondary earner in a household:

Secondary

Supporter

Assistant

Other (please list) _____

Elicited Social Norms

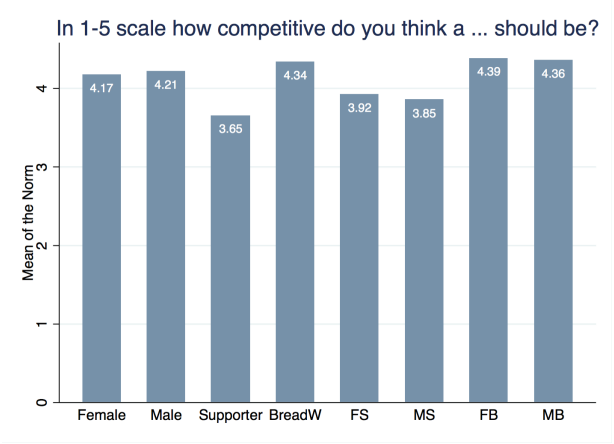
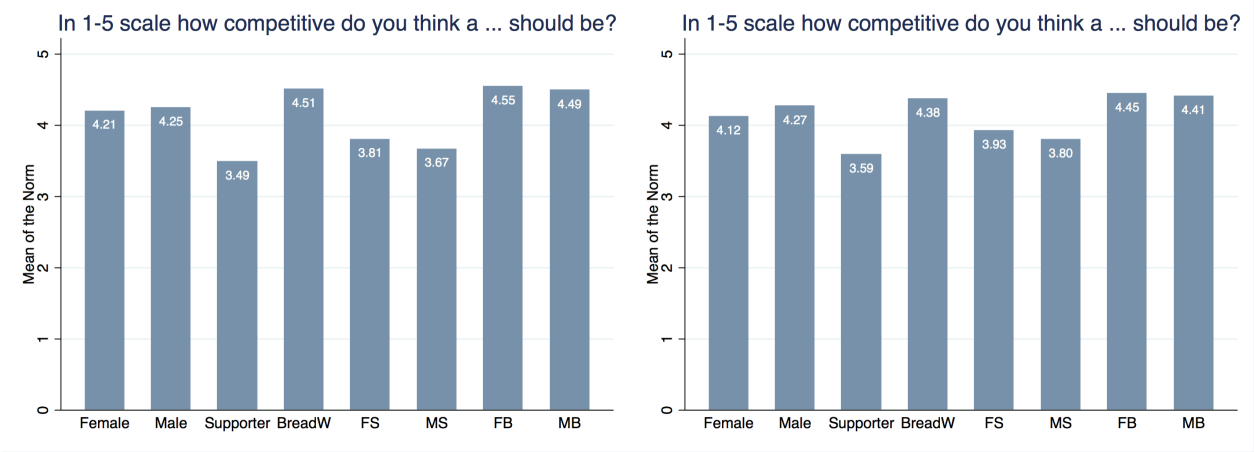


Figure A1: Elicited Social Norms for Competitiveness (including baseline & treatment group)



(a) Female (b) Male

Figure A2: Elicited Social Norms for Competitiveness by Gender

Average Allocations

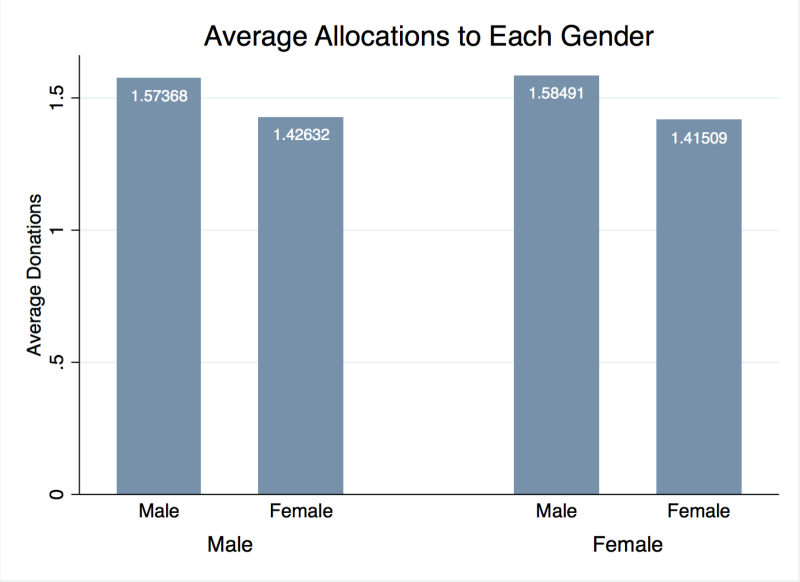


Figure A3: Average Allocations to Each Gender from Both Men and Women

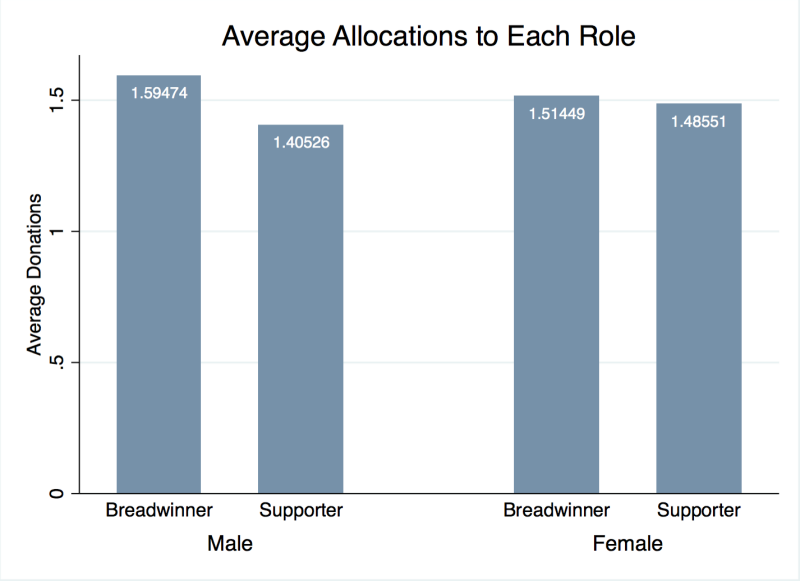


Figure A4: Average Allocations to Each Role from Both Men and Women