GENDER GAP IN LEADERSHIP SELF-SELECTION AND RETENTION: AN EXPERIMENTAL ANALYSIS

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfilment of the Requirements for the Degree

Bachelor of Arts

By:

Kexin (Willow) Ma

May 2024

GENDER GAP IN LEADERSHIP SELF-SELECTION AND RETENTION: AN EXPERIMENTAL ANALYSIS

Kexin (Willow) Ma

May 2024

Mathematical Economics

Abstract

This paper investigates the gender gap in leadership participation and the mechanism behind this observation, by examining the role of backlash and its impact on self-selection. Through an online experiment simulating the real-world work environment, I find evidence suggesting that women are significantly less likely than men to self-select into leadership roles initially. Once in a leadership role, a substantive proportion of both men and women stepped out in the following periods, with no significant gender difference. Although the results do not perfectly align with the existing literature, they provide the grounds for future studies to continue.

KEYWORDS: Gender differences, leadership, experiment

JEL CODES: D91, J16, C92

ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED

UNAUTHORIZED AID ON THIS THESIS

Signature

Table of Contents

Abstract II

List of Tables V

List of Figures VI

Acknowledgements VII

1 Introduction 1

2 Literature Review 3

3 The Experiment 6

3.1 Design and Implementation 6

3.2 Estimation Strategy 9

4 Results 11

4.1 Self-selection into manager roles 11

4.2 Retention rate of the manager role 13

5 Conclusion 18

References 21

Appendix 24

List of Tables

1 Summary statistics of participants 10

2 Percentages of male and female participants who chose to be the Manager 12

3 Self-selection into manager roles in round 1 14

4 Percentages of male and female participants who remain as managers . . 16

5 Participants who remain in manager roles in rounds 2 and 3 17

List of Figures

1 Percentages of participants who chose to be the Manager 13

2 Participants who remained in manager roles across three rounds 16

Acknowledgements

I would like to thank my thesis advisor, Professor Esther Redmount, for supporting the work of this thesis. Thank you to Dr. Jessica Hoel for providing feedback during the initial design stage of this experiment. Special thanks to Dr. Guanyi Yang for continued support in my academic endeavors.

Thanks to the Van Skilling Award from the Economics and Business Department and the Academic Opportunities Grant from the Colorado College for funding this study.

1 Introduction

The gender gap in leadership participation has long been observed and documented in existing literature, however, the study of mechanisms leading to this observed gap is still in a nascent stage. Studies have concluded that this difference can be attributed to women being less willing to lead and make decisions on behalf of others (Alan et al., 2019; Johnson & Helgeson, 2002; Jones & Linardi, 2014; Mayo et al., 2012). Chakraborty and Serra (2023) suggested that a significant factor leading to this observed gap is in one of the nature of leadership roles, that leaders have to make decisions that bring negative effects to certain groups of people. The possibility of receiving backlash resulting from these decisions prevents women in particular from self-selecting leadership roles.

This paper aims to study the effect of backlash on people's behavior in leadership roles, particularly the gender difference. I start by examining the gender difference in initial self-selection into leadership roles when knowing the possibility of receiving backlash. I also examine how men and women behave in response to receiving backlash, by analyzing their likelihood to leave the leadership roles. I attempt to explain the extent to which people step out of the leadership role in relation to receiving backlash.

I employ an online experiment to simulate a collaborative working environment while incorporating different roles at work, taken from Chakraborty and Serra (2023) with a twist. In the experiment, subjects work in groups of three, where one of them assumes the role of manager, and others assume the role of worker. They work in three rounds to complete a threshold public goods game where they have to decide how much to contribute to earn the return. The level of

return is different for each role of the group. The manager has to make decisions on worker's rankings, which will affect their level of payouts. The two workers in each group must be assigned to either Rank A or Rank B, where there can only be one worker in each rank. In this way, the experiment elicits backlash from the lower-ranked worker due to the unfairness of their payouts. At the beginning of each round, participants have to make choices on whether or not to be the manager or remain as the manager. The managers also have to make decisions every round on workers' rankings. These choices provide information on the proportion of participants who self-selected for the manager role initially, as well as who opt out of the manager role in the following rounds.

I find evidence of a gender gap in the initial self-selection for being the manager. A significantly larger percentage of males who were provided with the choice chose to be the manager than females. On the other hand, a substantive proportion of managers chose to step out of the role in the following rounds. However, there are no significant gender differences in the likelihood of stepping out. There is only a marginal effect of backlash observed on the likelihood of stepping out, which suggests that other social preferences were in play, such as a preference for fairness.

The results do not perfectly align with the existing literature, and the experimental design can be improved in various aspects. However, these findings are still suggestive and sets the grounds for the directions of future studies.

2 Literature Review

It is well documented that there exists a gender gap in leadership participation in the United States. Besides the exogenous organizational barriers integrated with the management hierarchy (Lyness & Grotto, 2018), a key factor hindering women's participation in leadership roles is self-selection into these positions. In particular, women have less willingness to lead because they seek to avoid public visibility, and make decisions on behalf of others.

Jones and Linardi (2014) conducted an experiment on altruism and visibility, showing that "wallflower behavior" is particularly strong in women, where they avoid standing out and tend to respond to visibility by seeking to minimize potential scrutiny. Similarly, Alan et al. (2019) found that women are less willing to perform under public scrutiny. In their longitudinal study, lacking social confidence is the most important obstacle to women's willingness to make decisions on behalf of others, which is a major component of a leader's job.

In addition to sensitivity to public scrutiny, women are also more sensitive and likely to internalize negative feedback from others. Mayo et al. (2012) performed a longitudinal study among MBA students, and showed that women are more likely to align their self-ratings with the peer reviews they receive. Johnson and Helgeson (2002) found that when receiving external evaluations, women's self-esteem was much more affected by negative feedback than positive feedback, while men's self-esteem stayed relatively unaffected by either type of feedback. Therefore, under the threat of receiving negative feedback as a leader, women are less likely to self-select into these roles.

Chakraborty and Serra (2023) performed a laboratory experiment to identify the gender difference in self-selection into leadership roles under the threat of backlash. They found that when solely given the choice to determine whether to be the leader, women are as likely as men to self-select into leadership roles. However, when given the possibility of receiving backlash from workers, women significantly refrain from self-selecting into leadership positions, compared to their male counterparts. In other words, women are more sensitive to the possibility of receiving backlash. And in response to this possibility, they are more likely to select out of leadership roles than men do.

Moreover, when considering the gender composition of the groups, Grossman et al. (2015) found that women show more hesitation for being the leader, and being in the vulnerable position to achieve a beneficial outcome in a mixed-gender environment. This observation can be partially attributed to the observed increasing willingness to cooperate with female leaders and followers than males. Born et al. (2022) also found that women are less willing to lead male-majority than female-majority teams. They revealed the mechanisms in effect are women's confidence, influence, and expected support from team members. In short, one of the main barriers keeping females from leadership roles is the fear of backlash and disagreement from their group members, as well as unwillingness to be in a visible and vulnerable position.

Departing from the existing literature, this paper aims to identify the lack of willingness to lead in women, and how it varies over time. Specifically, I investigate the likelihood of staying in or leaving leadership roles, in response to receiving backlash. Employing a similar experimental setup as Chakraborty and

Serra (2023) with a twist, I record the intertemporal choices of subjects and analyze the effect of various factors on these choices.

3 The Experiment

3.1 Design and Implementation

The experimental design follows the setup of Chakraborty and Serra (2023), with a slight modification. The experiment consists of three main stages. In the first stage, subjects were presented with the instructions and asked to fill out a short survey about their age and gender. Depending on the subject's gender, they were led to choose a fictitious name that would be used to identify themselves throughout the experiment from a pre-determined list of either male or female names. This design follows the result from Bertrand and Mullainathan (2004), as well as the design from Chakraborty and Serra (2023), where names were used to convey implicit information about gender. In order to rule out the possible race connotation, all names on the list were picked from the most popular names used by white Americans. Gender-neutral options were also included for non-binary identifying subjects, however, due to the small sample size, these subjects were not considered in further analysis.

In the second stage, subjects were grouped into three to complete three rounds of group projects. Subjects remain in the same group for all rounds. The design of different roles in each group was following Chakraborty and Serra (2023). Within each group, one person will assume the role of Manager, and the others will assume the role of Worker. The Manager has to decide on the ranking of two workers and place them into either Rank A or Rank B. At the beginning of the first round, one person in the group was randomly selected to choose whether they wanted to be the Manager of the group. If the person did not opt-in to be the Manager, the next randomly selected person in the group would be offered the same choice. If

none of the three members of the group opted in to be the Manager, the roles for each person will be randomly assigned by the system.

In each of the three rounds of group projects, subjects were asked to complete a threshold public goods game. Public goods games and their variants have been widely used in previous studies on leadership (Alan et al., 2019; Ertac & Gurdal, 2012; Moxnes & van der Heijden, 2003). As a variant of the public goods game, the threshold public goods game is used in this context to better simulate cooperative decision-making. This mechanism allows for the possibility of free-riding and coordination between group members (Cadsby & Maynes, 1998). When under proper coordination, the Rank B worker can have the possibility to free-ride or reach the threshold upon minimal contribution. This provides the chance to reduce the earning gap between different roles and reduces the probability of the Manager receiving severe backlash from the Rank B Worker. Moreover, the threshold public goods game involves public pressure and scrutiny (Iris et al., 2019), which puts more responsibility on the Manager where their decisions can be directly related to the interest of the workers.

Subjects were endowed with 100 Experimental Currency Units (ECU) every round regardless of their roles. Each group member will make a decision every round to contribute some amount of their endowment to the group project. The threshold for group projects to be completed is 150 ECU. If the total contribution of the group exceeds the threshold, the group project is completed and everyone in the group earns a return, where the amount of the return differs by one's role. The return is 400 ECU for the Manager, 250 ECU for the Rank A Worker, and 100 ECU for the Rank B Worker. If the group project is not completed, all

contributions will be returned to the contributor, and every member will keep the initial 100 ECU as their earnings for this round. Subjects were able to see the result of their own total earnings after each round of the group project, as well as the amount of contribution made by each group member in this round.

To further simulate the real-world work environment, following the design from Chakraborty and Serra (2023), subjects in the same group were allowed to communicate, and Managers had to communicate their decisions knowing that they would receive the reactions from their workers. Before the individual decisions on contributions were made for each round, the Manager would have to send an individual free-form message to the two workers. After each group project was completed, the workers had to send a free-form message to the Manager. Particularly, the Rank B Worker was allowed to send up to five angry emojis in addition to their message to express their discontent about the Manager's decision (Chakraborty & Serra, 2023).

In rounds two and three of the group project, the person assuming the role of Manager will be given the choice again of whether they want to remain as the Manager. If they opt out, the same process of role selection as the first round will be repeated until another person assuming the role of Manager is determined. The Manager, regardless of being the same or different person from the previous rounds, would have to decide on workers' ranking allocation again for the current round.

In the final stage of the experiment, subjects were asked to fill out a post-study survey, which collects information on demographics, previous leadership

experience, and personality measures based on the Extra-Short Form of the Big Five Model (C. Soto & John, 2017; C. J. Soto & John, 2017).

The experiment was programmed and run through the online behavioral research platform oTree (Chen et al., 2016). The participants of this experiment were recruited through Prolific, sampling the general US working-age population, with an age range of 22-55. The study involved a total of 114 participants, of which 57 percent were women, and 70 percent were white or Caucasian. The median age of the participants is 28.5. 70 percent of the participants own at least a Bachelor's degree. Detailed summary statistics are displayed in Table 1.

Subjects were paid a $3 show-up fee once they completed the study. Their bonus payment was accumulated through three rounds of games and was converted from ECU to dollars at the exchange rate of 1000 ECU per dollar. The average session length is 18 minutes, and the average total payment per participant is $4.

3.2 Estimation Strategy

The decisions of participants to be the Manager in round 1 are estimated through a linear probability model by the equation below:

$$Y_i = \beta_0 + \beta_1 F_i + \beta_2 G_i + \beta_3 D_i + \beta_4 R_i + \beta_5 L_i + \beta_6 A_i + \epsilon_i \tag{1}$$

Y_i is the probability that participant i selects to be the Manager. In this specification, the effect of gender on participants' probability of being the Manager is indicated by β_1. The control variables include participants' age G_i, level of education D_i, race R_i, previous leadership engagement L_i, and the Big Five agreeableness measure A_i (Chakraborty & Serra, 2023).

Table 1: Summary statistics of participants

	Male	Female	M=F (p-value) (Total)
Age	34.00	29.94	0.012
	(8.79)	(7.81)	
Race	0.78	0.65	0.141
	(0.42)	(0.48)	
Level of Education	0.64	0.75	0.218
	(0.48)	(0.43)	
Past Leadership	0.58	0.52	0.575
	(0.49)	(0.50)	
Big Five Agreeableness Index	11.17	11.84	0.162
	(2.52)	(2.40)	
Observations	45	65	110

Note: The average age, race, level of education, past experience with leadership, and Big Five agreeableness index and standard deviations are reported. Race is reported by 1 = White or Caucasian, 0 = Others. Education is reported by 1 = Bachelor's degree or higher, 0 = Less than Bachelor's degree. Big Five agreeableness index ranges from 1 to 15.

In rounds 2 and 3, a similar estimation strategy is applied for the retention of the Manager, but also accounting for the backlash received from the previous round. The estimation follows the equation:

$$Y_{it} = \beta_0 + \beta_1 F_i + \beta_2 G_i + \beta_3 D_i + \beta_4 R_i + \beta_5 L_i + \beta_6 A_i + \beta_7 E_{it-1} + \beta_8 E_{it-1} * F_i + \epsilon_i \quad (2)$$

where Y_{it} is the probability that participant i selects to be the Manager in round t, and E_{it-1} is the number of angry emojis received by the Manager in the previous round. The interaction term between gender and the number of angry emojis received captures the difference between genders in reaction to the backlash received.

4 Results

Table 1 reports the p-values for the balance tests in terms of demographic variables between male and female subject pools. The average age of male participants (34) is significantly higher than that of female participants (29.94, $p = 0.012$). The percentage of participants who are white or Caucasian in males (78 percent) is higher than the percentage of those in females (65 percent), but not significant ($p = 0.141$). There are also no significant differences in the percentage who own a Bachelor's degree or higher across genders. There is also no significant difference in the percentage who have past leadership experience across genders. Contradicting the existing literature, there is no significant difference in the average score of agreeableness index between male and female subjects ($p = 0.162$).

In what follows, I will present and discuss the main findings of this paper, in terms of the effect of gender on self-selection into manager roles, and the retention rate in manager roles across gender.

4.1 Self-selection into manager roles

The pattern of self-selection into manager roles across all rounds is displayed in Figure 1. Overall, although there is a difference in the percentages of male and female participants who selected to be the Manager in all three rounds, the differences are insignificant due to the high variance except for the first round. In this case, since the choices in rounds two and three are a combination of participants remaining as managers and participants replacing those who opted out, we only consider the choices in round one when exploring the gender difference in self-selection into manager roles.

Table 2 shows the percentages of male and female participants who chose to be the Manager in round one, among those who were given the choice, as randomly selected by the system. The p-values obtained from the T-tests are also reported. 83.33 percent of male participants who were given the choice decided to be the Manager, while only 51.51 percent of female participants who were given the choice decided to do so. The difference between male and female participants is significant at the 95% level ($p = 0.012$).

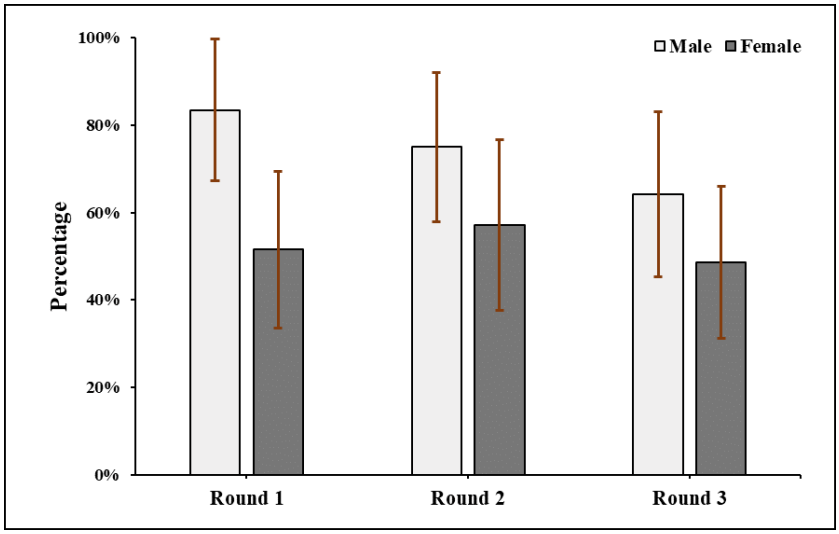
Table 3 displays the estimation result of participants' choice to be the manager using a linear probability model. The model without control variables suggests a strong negative relationship between gender being female and the probability of selecting to be the Manager. Accounting for the control variables, i.e. age, level of education, race, past leadership experience, and Big Five agreeableness index, the impact of gender on the probability of one self-selecting to be the manager is negative and significant at the 90% level ($p = 0.07$). The estimate suggests that female participants are on average 27 percent less likely to select to be the manager in round one than their male counterparts.

Table 2: Percentages of male and female participants who chose to be the Manager

	Male	Female	M=F (p-value)
Round 1	83.33%	51.51%	0.012
	(0.08)	(0.09)	
Round 2	75.00%	57.14%	0.164
	(0.08)	(0.10)	
Round 3	64.29%	48.57%	0.219
	(0.09)	(0.09)	

Percentages (and standard errors) are reported as the proportion of participants who were given the choice and chose to be the manager.

Figure 1: Percentages of participants who chose to be the Manager



Note: The figure shows the percentages (and 95% confidence intervals) of men and women who chose to be the Manager among those who were given the choice.

4.2 Retention rate of the manager role

Figure 2 demonstrates the percentage of male and female participants who chose to remain as the Manager among those who assumed the role of manager in the previous round. As shown in the figure, both men and women exhibited noticeable proportions who opted out of the manager role. Only 50 percent of male participants and 35.3 percent of female participants who assumed the role of manager in round one stayed in the manager position for all three rounds.

The percentages of men and women who remained in the manager role are reported separately for rounds two and three in Table 4. In round two, 65 percent of males and 64.71 percent of females who were the Manager in round one chose to remain as the Manager. In round three, 76.92 percent of males and 54.54 percent of females who were the Manager in round 2 chose to remain as the Manager. The differences between male and female participants for both rounds two and three are not significant ($p = 0.986$ and 0.266). This suggests that there is no observed

Table 3: Self-selection into manager roles in round 1

	(1) Without Controls	(2) With Controls
Gender	-0.318**	-0.273*
	(0.118)	(0.148)
Age		0.006
		(0.009)
Level of Education		-0.003
		(0.135)
Race		0.045
		(0.165)
Past Leadership		0.047
		(0.166)
Agreeableness		0.011
		(0.028)
Constant	0.833	0.510
	(0.077)	(0.346)
Observations	57	57
R-squared	0.108	0.127

Note: Gender: 1=Female, 0=Male
Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

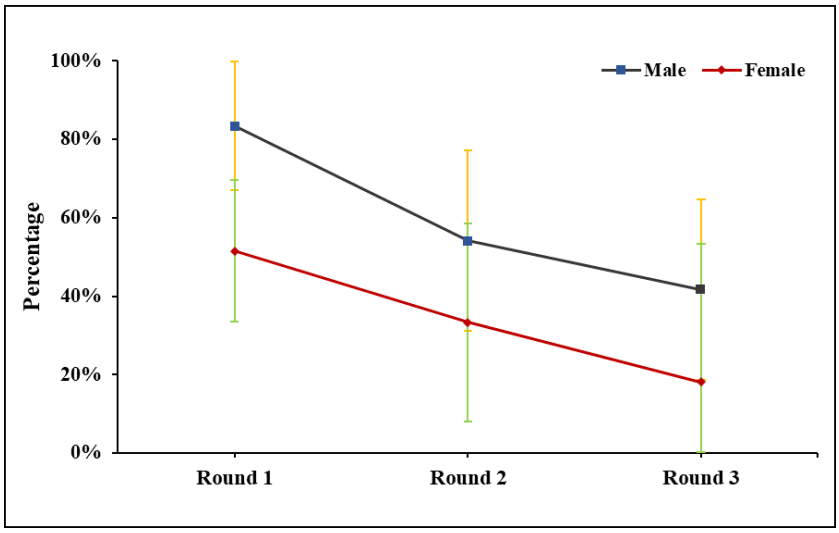
gender gap in the tendency to opt out of the manager role after receiving backlash. The gender gap observed in Figure 2 can be mainly attributed to the initial gap in self-selection into manager roles in the first round.

Table 5 displays the estimation results from the linear probability model which predicts the probability of one remaining in the manager role in round t, in relation to the backlash received (represented by the number of angry emojis sent by the Rank B worker) in round $t-1$. Both the tendency of retention in rounds two and three exhibits a positive relationship with the number of angry emojis received,

which contradicts the hypothesis and the existing literature. The coefficient for the effect of emojis received in round 1 on the retention rate in round 2 is significant at the 90% significance level ($p = 0.072$), and the magnitude of the coefficient is substantive. However, the coefficient for round three becomes significantly smaller in magnitude, and statistically insignificant ($p = 0.378$). Moreover, the coefficients for both the gender term and the interaction terms between gender and the number of emojis received have opposite signs for rounds two and three. Given their coefficients are all insignificant and small in magnitude, one cannot conclude a clear pattern that informs the difference in participants' response to backlash between genders.

The failure to conclude any patterns can be attributed to two main reasons. First, the number of observations is too small to generate significant estimations. Second, due to the nature of the threshold public goods game as well as the relatively small stakes in the experimental design, not enough backlash was elicited, hence insufficient to generate informative estimates. In this case, there are no clearly observed effects of backlash received on the likelihood of remaining as the Manager. However, we do observe a significant proportion of participants who chose to opt out of the manager roles in rounds two and three. It is reasonable to infer that the tendency to opt out of manager roles can be attributed to social preferences other than fear of backlash, such as a preference for fairness, which is not captured in this particular experimental design.

Figure 2: Participants who remained in manager roles across three rounds



Note: This figure shows the percentage (and 95% confidence intervals) of men and women who were the manager in the previous round and chose to remain as the manager in the next round.

Table 4: Percentages of male and female participants who remain as managers

	Male	Female	M=F (p-value)
Round 2	65.00%	64.71%	0.986
	(0.11)	(0.12)	
Round 3	76.92%	54.54%	0.266
	(0.12)	(0.16)	

Note: Percentages (and standard errors) are reported as the proportion of participants who chose to remain as managers among those who were managers in the previous round.

Table 5: Participants who remain in manager roles in rounds 2 and 3

	(1)	(2)
	Round 2	Round 3
Number of Angry Emojis Round 1	0.308	
	(0.165)	
Number of Angry Emojis Round 2		0.066
		(0.073)
Gender	0.117	-0.369
	(0.207)	(0.368)
Gender x Round 1 Emojis	-0.294	
	(0.229)	
Gender x Round 2 Emojis		0.028
		(0.141)
Age	0.010	0.028**
	(0.010)	(0.007)
Race	0.094	-0.411
	(0.251)	(0.284)
Level of Education	0.068	-0.237
	(0.200)	(0.218)
Past Leadership	-0.161	-0.620
	(0.234)	(0.300)
Big Five Agreeableness Index	-0.00832	0.010
	(0.038)	(0.055)
Constant	0.283	0.513
	(0.542)	(0.687)
Observations	37	24
R-squared	-0.130	-0.006

Note: Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

5 Conclusion

In this paper, I explored whether men and women behave differently in the tendency to self-select into leadership roles, as well as their rate of retention in leadership roles in response to receiving backlash. By employing an experiment, I was able to assess the effect of gender on the likelihood of self-selecting into leadership roles, as well as the effect of backlash on participants' tendency to opt out of manager roles.

The results show significant evidence of gender difference in the likelihood of initial self-selection into manager roles. Women are significantly less likely to opt in as the manager than men initially. The results also suggest that a significant proportion of participants opted out of the manager roles in the following rounds, while there are no significant gender differences in these proportions. In the estimates of the impact of backlash on the likelihood of remaining as managers, the effect of backlash received is only marginally significant. Meanwhile, there are no observed gender differences in response to backlash. While these findings do not align with the existing literature, they can be attributed to the following reasons.

First, the experimental design, although similar to that of Chakraborty and Serra (2023), has been modified in several mechanisms. The choices of whether to be the manager were provided to random participants selected by the system, and no bases or justifications for this leadership were provided to the participants. In this case, the managers are more likely to opt out of the position, because they did not "earn" the position themselves. This amplifies the social preference for fairness. Meanwhile, the threshold public goods game, although incorporating a large difference in returns for different roles, still preserves the room for group

members to free-ride or invest the minimum. The Manager was also able to express a sense of fairness and compensation by investing more in the group project. Either way, backlash from the lower-ranked worker will be significantly reduced. The relatively low stakes for the threshold public goods games were also a factor of not enough backlash being elicited.

Another important factor that leads to insignificant results is the insufficient sample size. Although there were more than 110 participants being surveyed in total, the actual number of participants who were making management decisions was much smaller. This insufficiency was particularly profound when analyzing the retention rates in rounds 2 and 3. A revised experimental design may improve this insufficiency, where all participants can be asked to make the selection first, and then decide on the manager based on certain criteria, such as a trivia quiz, or some kind of real-effort task. When contextualizing the results from lab or online experiments, one should be aware that the stakes are relatively low compared to what is in the real world. Therefore, it is important to consider the possible mitigation of behaviors or reactions performed by subjects.

The future directions of this study include implementing a control group since this is supposed to be a randomized controlled experiment. A control group not incorporating the possibility of backlash will allow us to separate and make causal estimations on the effect of backlash on participants' decisions. It also allows us to see whether men and women behave differently in the initial self-selections in the absence of backlash. It is also worth revising the experimental design to provide enough bases and justifications for both the manager role and managers' decisions on workers' rankings, as well as the nature of the group project mechanism, to

make sure enough backlash will be elicited. Despite the imperfections, the current study still provides suggestive results that set the grounds for future research to continue.

References

Alan, S., Ertac, S., Kubilay, E., & Loranth, G. (2019). Understanding gender differences in leadership. *The Economic Journal*, *130*(626), 263–289. https://doi.org/10.1093/ej/uez050

Bertrand, M., & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, *94*(4), 991–1013. Retrieved February 20, 2024, from http://www.jstor.org/stable/3592802

Born, A., Ranehill, E., & Sandberg, A. (2022). Gender and Willingness to Lead: Does the Gender Composition of Teams Matter? *The Review of Economics and Statistics*, *104*(2), 259–275. https://doi.org/10.1162/rest_a_00955

Cadsby, C. B., & Maynes, E. (1998). Gender and free riding in a threshold public goods game: Experimental evidence. *Journal of Economic Behavior & Organization*, *34*(4), 603–620. https://doi.org/https://doi.org/10.1016/S0167-2681(97)00010-3

Chakraborty, P., & Serra, D. (2023). *Gender and leadership in organizations: The threat of backlash* (tech. rep.). Technical report, Working Paper.

Chen, D. L., Schonger, M., & Wickens, C. (2016). Otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/https://doi.org/10.1016/j.jbef.2015.12.001

Ertac, S., & Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups [Gender Differences in Risk Aversion and Competition].

Journal of Economic Behavior & Organization, *83*(1), 24–30. https://doi.org/https://doi.org/10.1016/j.jebo.2011.06.009

Grossman, P. J., Komai, M., & Jensen, J. E. (2015). Leadership and gender in groups: An experiment. *The Canadian Journal of Economics / Revue canadienne d'Economique*, *48*(1), 368–388. Retrieved February 23, 2024, from http://www.jstor.org/stable/43818279

Iris, D., Lee, J., & Tavoni, A. (2019). Delegation and public pressure in a threshold public goods game: Theory and experimental evidence. *Environmental and Resource Economics*, *74*, 1331–1353. https://doi.org/https://doi.org/10.1007/s10640-019-00371-6

Johnson, M., & Helgeson, V. S. (2002). Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly*, *26*(3), 242–251. https://doi.org/10.1111/1471-6402.00063

Jones, D., & Linardi, S. (2014). Wallflowers: Experimental evidence of an aversion to standing out. *Management Science*, *60*(7), 1757–1771. Retrieved December 1, 2023, from http://www.jstor.org/stable/42919632

Lyness, K. S., & Grotto, A. R. (2018). Women and leadership in the united states: Are we closing the gender gap? *Annual Review of Organizational Psychology and Organizational Behavior*, *5*(1), 227–265. https://doi.org/10.1146/annurev-orgpsych-032117-104739

Mayo, M., Kakarika, M., Pastor, J. C., & Brutus, S. (2012). Aligning or inflating your leadership self-image? a longitudinal study of responses to peer feedback in mba teams. *Academy of Management Learning & Education*,

11(4), 631–652. Retrieved December 1, 2023, from http://www.jstor.org/stable/23412349

Moxnes, E., & van der Heijden, E. (2003). The effect of leadership in a public bad experiment. *Journal of Conflict Resolution*, *47*(6), 773–795. https://doi.org/10.1177/0022002703258962

Soto, C., & John, O. (2017). Short and extra-short forms of the big five inventory–2: The bfi-2-s and bfi-2-xs. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Soto, C. J., & John, O. P. (2017). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Appendix

Experimental Protocol

Instructions

Thank you for agreeing to participate in this project! To thank you for your time, you will be paid a $3 show-up fee for completing the experiment.

In addition, you will have the chance to earn additional money in this experiment.

Earnings during the experiment will be denominated in Experimental Currency Units, or ECU. At the end of the session, your earnings throughout the experiment will be converted to dollars at the exchange rate of $1 for 1000 ECU.

In this experiment, you will be working with two other participants as a group. One group member will assume the role of Manager and the other two group members will assume the role of Worker.

The main role of the manager is to decide which worker will be rank A and which worker will be rank B in the group.

Each group will be collectively completing 3 rounds of group projects. In each round, you start with an endowment of 100 ECU.

You can choose to contribute some amount of your endowment to the group project. The group project takes **150** ECU to be completed. Once it is completed, each person in the group gets an additional payout, with a different rate for the Manager, rank A Worker, and rank B Worker.

If the group project is completed:

- The Manager gets a 400 ECU return.

- The Rank A Worker gets a 250 ECU return.
- The Rank B Worker gets a 100 ECU return.

If the group project is not completed, all the endowments you contributed will be returned.

The total earnings for each round of the group project are determined as follows:

- Manager gets 400 ECU + leftover endowment if the project is completed.
- Rank A Worker gets 250 ECU + leftover endowment if the project is completed.
- Rank B Worker gets 100 ECU + leftover endowment if the project is completed.

If the project is not completed, every member in the group will keep a 100 ECU earning, the same as your initial endowment for this round.

Let's take a moment to make sure you understand the instructions so far.

1. Suppose you are the Manager. You contributed 50 ECU to the project, and the project is completed, what will be your total earnings?

   (a) 500

   (b) 450

   (c) 400

2. Suppose you are the Rank B Worker. You contributed 200 ECU to the project, but the project is not completed. What will be your total earnings?

(a) 100

(b) 150

(c) 200

3. How much does each person have to contribute ON AVERAGE to complete the group project?

(a) 50

(b) 100

(c) 150

At the beginning of round 1, one person in the group will be randomly selected to choose whether they want to be the Manager. If the person refuses the role, another randomly selected person will be given the choice. If all three members of the group refuse to be the Manager, the roles will be randomly assigned by the system.

At the beginning of rounds 2 and 3, the Manager in the previous round will be asked again whether they want to continue being the Manager. If they opt out, the same process will be repeated as in round 1 where other members of the group will have the chance to be the Manager.

You and your group members will make decisions 3 times across 3 rounds. In addition to the $3 base payment, you get the additional payment accumulated during the experiment converted to US dollars.

At the beginning of each round, the Manager will have to decide which worker will be Rank A and which worker will be Rank B. After the Manager makes the

allocation decision, each worker will be informed about the Rank they have been assigned.

After assigning ranks A and B to the workers, the Manager will have to send a message to the Rank A worker and a message to the Rank B worker. The message chat box cannot be left blank. In the chat box, the Manager can write anything they wish to communicate to each worker. After the Manager submits the individual messages, each worker will see the message sent to them.

After the group project is completed, the Workers have to send a message to the Manager. They can write anything they wish to communicate to the Manager. In addition, the Rank B worker can send one or more angry faces to the Manager to express their disapproval of the ranking decision. In particular, the Rank B worker can send up to 5 angry faces to the Manager, as shown below. The messages and angry faces sent will not affect your earnings or others' earnings. ¡br¿

Please remember to be courteous to your co-workers. If you use any profanity or bullying language, we will not pay you and we will report you to Prolific.

1. What is your gender identity? [male, female, others]

2. What is your age in years?

We would like to use a name for you in this project, but to protect your privacy, we will not ask for your real name.

[If male] [Jacob, Michael, Joshua, Jeremy]

[If female] [Heather, Jessica, Hannah, Emily]

[If other] [Kai, Ridley, Lowen, Ellis]

Here are your two group members:

XX whose age is XX.

XX whose age is XX

Threshold public goods games

This is round X.

Do you wish to be the manager? [yes, no]

Please select the corresponding name of the person who you want to be the Rank A Worker. [p1name, p2name]

You are the [Manager/Rank A worker/Rank B Worker].

XX is the Rank A worker.

XX is the Rank B worker.

[For Managers]

As the Manager of your group, you would have to send a message to each of your workers.

What would you like to say to the Rank A worker?

What would you like to say to the Rank B worker?

[For Workers]

Here is a message from your Manager:

The group project X starts!

You are endowed with 100 ECU.

How much do you want to contribute?

You started with an endowment of 100 ECU, of which you contributed X.

XX contributed X [For group member 1].

XX contributed X [For group member 2].

Your group contributed X in total, your group did/did not complete the project.

Your earning for this round is therefore X.

[For Workers]

What message would you like to send to your Manager?

[For Rank B Workers]

How many angry faces would you like to send to your Manager? [0-5]

[For Managers]

Here is a message from your Rank X worker: [with angry emojis if applicable]

Post Survey

Understanding

1. What do you think of the study and these instructions so far? Was it fun? Was it boring?

2. How difficult was it for you to understand the tasks today?

3. How difficult was it for you to make choices in the study today?

Demographics

1. What is your marital status? [single, partnered, married, divorced, widowed]

2. What best describes your sexual orientation? [homosexual, heterosexual, bisexual, others]

3. Do you have children? [yes, no]

4. What is the primary race/ethnicity with which you identify? [White or Caucasian, Black or African American, Asian or Asian American, Native American, Native Hawaiian or other Pacific Islander, Hispanic/Latinx]

5. What is your highest level of education? [less than high school graduate, high school graduate, some college/vocational training, associate's degree, bachelor's degree, master's degree, Ph.D.]

6. Have you ever been in a leadership position at work? [yes, no]

Personality

Survey taken from The Big Five Inventory–2 Extra-Short Form (BFI-2-XS) (C. J. Soto & John, 2017).