Gender and Leadership: An Experimental Analysis

Yaxiong Li

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Economics, the University of Auckland, 2016.

Abstract

We conduct three different decision-making experiments to explore gender differences in labour market interactions. In our first experiment, participants are grouped in five, and are asked to play a corporate turnaround game. We rely on this game to explore: (1) if there are systematic gender differences in the willingness to lead and (2) whether followers are equally likely to follow messages from male and female leaders. In our second experiment, we use a principal-agent game to study gender differences in employment contract interactions between employers and workers. Employers are given the option to choose between a relational contract and a transactional contract. In our third and final experiment, we further explore gender differences in trust and reciprocity using a trust game. We find that, compared to men, fewer women volunteer to lead, particularly when the leader's gender is revealed to the followers. However, we show that when leader messages and action choices are similar, even if not identical, groups achieve similar levels of coordination success regardless of the leader's gender. We do not find evidence of systematic resistance to female leadership, even though such an anticipated backlash may be causing the female reluctance to lead. This suggests that the leadership gap in the workplace may be partially due to supply-side factors with fewer women volunteering to lead.

JEL Classification Codes: C91, D91, J16

Acknowledgements

I want to express my deep gratitude to my supervisor Professor Ananish Chaudhuri, who not only guided me through my studies but also provided extensive feedback on my dissertation chapters and cared for me like a family member. I could not have imagined having a better supervisor and mentor for my Ph.D. study.

I would like to express my sincere gratitude to my co-supervisor Dr Erwann Sbai, who provided invaluable advice on the econometric analysis in my dissertation. I am grateful to Associate Professor John Hillas, Professor John Panzar, Dr Alan Rogers, and Dr Ryan Greenaway-McGrevy for their feedback and support.

My sincere thanks to MSA Charitable Trust and Kelliher Charitable Trust for supporting my studies via scholarships.

This thesis has improved vastly from insightful comments provided by Professor Pushkar Maitra, Professor Lata Gangaharan, Professor Lionel Page, Assistant Professor Maria Recalde, and Dr Ananta Neelim. I also thank participants at various conferences for their feedback. This includes the Economic Science Association North-American Conference in Los Angeles (2019), the Australia New Zealand Workshops on Experimental Economics in Auckland (2018) and Melbourne (2019), the annual conferences of the New Zealand Association of Economists in Auckland (2018) and Wellington (2019), and the Australia Gender Economic Workshops in Melbourne (2019) and in Brisbane (2020) for valuable feedback.

I am also grateful to my family and friends. I owe a special debt of gratitude to Chenchen (Isabella) Li, Dr. Tony So, Mrs Annette Lazonby, and members of RWC. I am responsible for any and all errors in this thesis.

Contents

1. Introduction

The “gender gap” in the workplace is a well-documented phenomenon. This gender gap takes two forms. The first is the gender gap in wages: that mean and/or median earnings is higher for men than for women. The second refers to the gender gap in leadership roles: there are fewer women as we move up organizational hierarchies. This thesis focus on the second phenomenon. The aggregate statistics have shown that there are far fewer women in elite leadership roles than men, despite the improvement of women in labour force participation and in educational attainment. This thesis utilises economic decision-making experiments to study the reason of the gender gap in leadership roles.

The reasons for a gender gap in leadership roles comes from both the demand-side and the supply-side of the labour market. On the supply-side, explanations look at gender differences in psychological attributes and preferences between men and women. Women may be reluctant to take on leadership roles due to stereotype attributes and preferences. On the demand side, explanations include discrimination or biased perceptions towards female leaders. Biased perceptions can lead to undervaluation of a woman’s effectiveness as a leader.

In this thesis, we focus on four primary research questions by utilizing three laboratory experiments. The first and second research question ask whether there are systematic gender differences in the willingness to lead, and whether there are differences in the perception of female leaders compared to male leaders, in the sense of followers’ willingness to follow male leaders more than female leaders. Our third research question asks whether there are systematic gender differences in leadership styles. In particular, we examine whether male leaders are more likely to rely on rewards and penalties and follow a transactional leadership, and if female leaders are more likely to rely on mutual trust and reciprocity and follow a transformational leadership. Our final research question relates to gender differences in trust and reciprocity.

To study our first two research questions, we examine a minimum effort coordination game, which serves as a good vehicle for simulating intra-organizational coordination problems. We use a modified version of the Brandts and Cooper (2006, 2007) corporate turnaround game. We include two sets of experiments. In our first experiment, both leaders and followers face the same payoffs. In our second experiment, leaders face a different payoff matrix compared to followers. Within these two experiments we also change two additional factors. Firstly, we change whether the leader's gender is revealed to followers or not. Secondly, we change the nature of messages that leaders can send to their followers. In one treatment, the message sent is pre-determined and written by the experimenter. The leader only gets to choose how frequently to send this message. In a second treatment, leaders are permitted to write free-form messages. Overall, we do not find significant differences in followers' perception towards male and female leaders, but we do find that women are less likely volunteer for leadership positions.

To study our third research question, we use a principal-agent game (Fehr, Gächter, and Kirchsteiger, 1997) to test the proposition that in terms of the work-place, female leaders tend to be more transformational in the sense of relying more on mutual trust and reciprocity based relationships, compared to male leaders, who rely more on explicit "carrots" and "sticks". We implement two treatments. In the fixed matching treatment, participants are matched with the same person for the entire session of the experiment, whereas in the random re-matching treatment, participants are randomly re-matched with a different person in each round. We do not find dramatic differences in behaviour. Women showed a slight preference for transactional contracts and making more generous offers. When it comes to worker behaviour, women tend to shirk a little more. However, these differences are neither major nor permanent.

We study a repeated version of the Berg, Dickhaut, and McCabe (1995) trust game with random re-matching to investigate whether there are systematic differences in the cognitive demand effect and examine whether gender interacts with such an effect. Previous studies from the trust game generally report more trust from men than women or no difference between the genders, while women reciprocate more compared to men or no difference between the genders. Given that most of the studies focused on one-shot game, one possible explanation is the misunderstanding or misinterpretation between participants and experimenters due to there being little scope to learn the game. Given that most studies use neutral and context-free language, there may well be cognitive demand effects. The interpretation of the task by participants may be different from that of the experimenter.

We designed five treatments. Our first treatment is a Private knowledge treatment. Participants in this treatment are only provided with context-neutral written instructions. In our second treatment, which we refer to as a Common knowledge treatment, participants are provided with the same context-neutral written instruction as in the private knowledge treatment, except that the instructions are read aloud to the participants by the experimenter prior to the start of the session. In our other three context treatments, participants receive the same written instructions as in the private knowledge treatment and common knowledge treatment, which are read out loud. However, two additional paragraphs are added to the instruction. We use context neutral languages in our Context-neutral treatment. In Context-loaded A treatment, here we explicitly use the words “trust” and “trustworthiness” in one of the additional paragraphs. The idea is to make the motivations explicit to the participants. In the Context-loaded B treatment, the additional two explanation paragraphs are identical to Context-Loaded A treatment except here we interchange the order of the two paragraphs.

We find that both genders show a higher level of trust when the game is explained to them. We find that men exhibit more trust and reciprocate more compared to women when we remove strategic uncertainty. Contrary to with many previous studies, we do find men are significantly more reciprocal than women when strategic uncertainty is eliminated. We do not find any evidence that cognitive demand effects differ by gender.

The thesis proceeds as follows. Chapter 2 provides a literature review of relevant studies on the gender gap in leadership roles. Chapter 3 provides the study of the gender gap in the willingness to be a leader and perception towards female leaders. In chapter 4, we report on a study of gender differences in leadership styles from experiments simulating an employer-employee relationship. Chapter 5 provides results for gender differences trust and reciprocity and the relations to cognitive demand effect. Chapter 6 concludes.

2. Literature Review

In this chapter, we survey the relevant literature on the gender gap in leadership roles. We first examine existing explanations for the differences between males and females from the supply-side in the labour market. Next, we summarize the literature relating to the demand-side, including discrimination both from the upper level and the lower levels in the organizational hierarchies of the employing organisations.

We first review one of the most recent explanations for gender differences in wages and leadership roles, namely gender differences in psychological attributes and noncognitive skills. The various examinations of this explanation in the literature includes both lab and field experimental evidence and survey evidence. We divide our examination of the literature on discrimination regarding women's representation in senior roles through the perspectives of upper level and lower levels of organizational hierarchies, by first outlining literature regarding employer discrimination in entry or promotions, and then summarizing studies concerning employees' perceptions towards female leaders.

Our research question examines gender differences in the willingness to lead. We review the literature on gender differences in psychological attributes. We also look at a few papers regarding personal traits that are relevant to the gender gap in leadership roles. Next, we look at perceptions toward female leaders. We review perceptions both from the employer, i.e. upper level of the organizational hierarchies, and from the employee i.e. lower level of organizational hierarchies.

We then focus on gender differences in leadership styles and gender differences in the workers' behaviour. We summarize relevant papers regarding different leadership styles with gender identity norms. Our final research focus is on gender differences in trust and reciprocal behaviour.

2.1. Gender differences in psychological attributes

2.1.1. Gender differences in competition

Senior positions in organizations and the government typically take place in highly competitive environments. Several papers show that men exhibit more competitiveness than women. This relatively recent research suggests that women may systematically underperform compared to their male counterparts in competitive settings, and some of them may choose to avoid such environments. These findings provide one explanation for the lower chances of success for women when competing for promotions or in entering male-dominated and competitive fields.

Gneezy, Niederle, and Rustichini (2003) conduct controlled lab experiments to test gender differences in performance in competitive environments. They have groups of three male and three female students perform a task of maze solving. Students are given 15 minutes to solve as many mazes as they can under two different compensation schemes. Under a piece-rate scheme, students are paid a fixed rate for each maze solved. Therefore, their payment depends on their own performance. Under a tournament scheme only the student who solves the largest number of mazes is paid proportionally to their output. Therefore, their performance depends on their performance relative to that of the other students and is more uncertain. Gneezy et al. (2003) find no significant gender difference under the piece-rate scheme. However, under the tournament scheme the performance of the male students increases significantly more than that of the female students. Given that the nature of the payment is uncertain in the tournament treatment, one explanation of the gender difference is that the female students are more risk-averse. Gneezy et al. (2003) conduct a third treatment to test this possibility. Under this treatment, only one participant is paid, and this "winner" is chosen at random. They find no significant gender differences in performance under this treatment.

Dutta-Gupta, Poulsen, and Villeval (2013) extend the work by Gneezy et al. (2003), looking at whether an individual's competitiveness is influenced by the gender of their matched person. In their study, each participant is informed about their co-participants' gender, and the participants then choose between a piece-rate payment scheme and a tournament payment scheme. The participant who chooses piece-rate payment scheme is paid at 4 points for every maze he or she solves, regardless of choice made by their co-participant. In the case where both individuals choose the tournament payment scheme, the one with most solved mazes gets 6 points per maze while the other gets 1 point. If one participant chooses the tournament scheme, and the other participant chooses a piece-rate, the participant who selected the tournament scheme gets 6 points per maze. Dutta-Gupta et al. (2013) conduct two treatments. In the "weak gender information procedure," participants are informed about their co-participant's pseudonym without explicitly revealing that the pseudonym reflects gender. In the "strong gender information" treatment, gender information is common knowledge. They find that the percentage of women who choose tournaments under both treatments is significantly lower as compared to men. Men compete less against other men than against women in the "strong gender information procedure", but not in the "weak gender information procedure". The gender of the co-participant does not have a significant impact on the women's choice under both treatments.

In the two papers we have just described participants are exogenously assigned to payment schemes. Niederle and Vesterlund (2007) report on experiments in which the participants choose the payment scheme and study whether there is any gender difference in their choices. Participants in their experiment all perform a number adding task under a piece-rate pay scheme, and then under a tournament scheme. After experiencing both compensation schemes, participants are asked which scheme they would like to apply to their next performance. Under a piece-rate payment scheme, participants are rewarded based on

absolute performance, whereas under a tournament scheme, participants are rewarded based on relative performance. Niederle and Vesterlund (2007) conclude that women shy away from competition, as only one-third of women selected the tournament treatment compared to almost three-quarters of men. This is consistent with the findings by Vandegrift and Yavas (2009) where they find that men enter tournament treatments more frequently than women.

The gender differences in the willingness to compete may be linked to a cultural difference. Gneezy, Leonard, and List (2009) study two distinct societies, the patriarchal Maasai society and the matrilineal Khasi society. These two societies represent very different cultures in terms of gender. The experimental task carried out by Gneezy et al. (2009) is to toss a tennis ball into a bucket, where the shot is considered successful if the tennis ball stays inside the bucket. Participants from two societies are informed that they have 10 chances to toss the tennis ball and they are matched in pairs with someone else from another group. Participants are asked to choose between a piece-rate payment scheme and a tournament payment scheme. In the piece-rate payment scheme, participants are paid X for each successful shot regardless of the performance of the matched person from another group. In the tournament payment scheme, participants are paid 3X for each successful shot if they outperform the person with whom they are matched. Gneezy et al. (2009) find that men are more competitive among the Maasai while women are more competitive among the Khasi, suggesting that socialization is an essential determinant of gender preferences for competitiveness.

Andersen, Ertac, Gneezy, List, and Maximiano (2013) carry out similar studies by comparing the competitiveness of children in matrilineal and patriarchal societies. They assign the same tennis ball task, but in their design, the children only have 5 chances instead of 10. In the patriarchal society, they find no significant gender differences at the young age

but find that girls become less competitive while boys become more competitive around puberty. In the matrilinear society, there is no significant difference at any age.

2.1.2. Gender differences in negotiation

Negotiation is one of the key characteristics that is associated with leadership. In order to access resources and opportunities, one should seek out, instead of shy away, from chances to negotiate. Gender differences in negotiation behaviour have been both in the laboratory and in the field.

Babcock and Laschever (2003) claim that women are less likely to engage in competitive negotiations than men. In this study, they find that women with MBAs who had recently graduated from Carnegie Mellon were less likely to initiate negotiations over salary but simply accepted their employer's initial salary offer. Only 7 percent of these women have attempted to negotiate their salary compared to 57 percent of the men. This resulted in a 7.6% lower starting salary of female MBA graduates compared to male MBAs from the same program. This small gap in starting salaries can translate to substantial differences over time; therefore, it is a significant contributor to earnings differentials in the long term.

Exley, Niederle, and Vesterlund (2020) also examine gender differences in negotiation in lab experiments. Participants are randomly matched into worker-firm pairs, and then re-matched randomly each round. Both parties faced 10 rounds of negotiation opportunities that determine how to split a joint revenue. In the "choice" treatment, workers are given negotiation opportunities with the firm. They can choose to avoid negotiation by accepting the computer-generated suggested wage or can otherwise reject the suggested wage and enter into a negotiation. In the "always" treatment, workers observe the suggested wage and enter negotiations. For negotiation that reaches agreement, the worker receives the agreed-upon wage while the firm receives the joint revenue minus the wage. For negotiation that fails to

reach an agreement, both parties receive a payoff based on the suggested wage, subject to a penalty of $5. They conclude that women know when to ask, in the sense that women do enter negotiations when the outcome benefits them, while they avoid negotiations that are financially harmful for them.

Chaudhuri, Cruickshank and Sbai (2013) conduct a laboratory experiment using a principal-agent game adapted from Fehr et al. (1997, 2007) to examine gender differences in employer and worker behaviour. Participants are randomly assigned into either an employer role or an employee role, and they are randomly re-matched with a different person in each round. Chaudhuri et al use two experiments, each with two treatments. In experiment 1, under the trust treatment, employers offer a wage and suggest an effort level. If the worker accepts the offer, the worker will choose an actual effort level, which can be different from the suggested effort level. Under the penalty treatment, employers specify a required effort level and have the chance to specify a fine, which is imposed if the worker is detected to be providing less than the required effort. The monitoring technology is imperfect and detects shirking with only a probability of one third. Employers receive the fine only if the worker is shirking, they have chosen to use the monitoring technology, and the shirking is detected. Shirking is defined as the worker's effort level being less than the suggested effort level.

Experiment 2 has the same two treatments except here employers are endowed with an additional 25 dollars at the beginning of the game. Chaudhuri, Cruickshank and Sbai (2013) find that female employers offer significantly higher rent than male employers in both experiment under the penalty treatment, but this behaviour dissipates over time. They do not find any significant gender differences in rent offer under trust treatment. They find no significant gender differences in the likelihood of shirking from workers.

Leibbrandt and List (2015) analyse a natural field experiment to examine gender differences in choosing workplaces where salary is negotiable and in the initiation of salary negotiations. They place advertisements for two different job tasks, each with two different advertisements. One job task involves sports and is considered to be more "masculine" while the other job task is considered to be "gender neutral". For each task one job advertisement explicitly mentions that wages are negotiable while the other advertisement only mentions the level of salary. Leibbrandt and List (2015) find that men are more willing to negotiate for a higher wage than women when there is no explicit information regarding whether wages are negotiable, this gender difference vanishes when the job advertisement explicitly mentions that the wages are negotiable. Women are also more likely to accept a lower wage when wages are explicitly advertised as negotiable.

Andersen, Ertac, Gneezy, List, and Maximiano (2018) conduct a laboratory experiment and a field experiment to study negotiation culture in matrilineal and patriarchal societies. The field experiment looks at bargaining in a real marketplace that involves participant selection into a buyer role or a seller role. They also examine an alternating-offer bargaining game in a lab experiment where roles are randomly assigned. In the lab experiment, it is common knowledge that the price at which the seller is willing to supply is equal to zero while the maximum price buyer is willing to pay is 150. Sellers and buyers simultaneously choose their initial offer, it is then determined by a coin toss which offer is implemented. The game then proceeds with alternating offers. The seller and buyer then either accept or reject the offer after the first offer is made. The game ends if an offer is accepted. In case of a rejection, nature determines whether the game continues or ends. The seller and buyer can make a counteroffer if the game continues. The probability of game ends equals to $r/10$, where r is the number of rejections, indicating that the likelihood of the game ending increases with each rejection, and the game can last for a maximum of 10 rounds. The field

experiment was carried out in the Burra bazaar in Shillong city and involved hiring locals as buyer subjects from Khasi (matrilineal tribal society) and Hindu (patriarchal society) to bargain towards the acquisition of a particular good. Buyers are given 30 Rupees for each purchase, and they are incentivized to negotiate as they can retain the remaining money after each transaction. Andersen et al. find no significant difference in earnings between males and females in the patriarchal society. However, females earn significantly more on average than males in the matrilineal society. This result is consistent with women not being naturally disadvantaged in bargaining compared to men.

There are studies that show that women face a higher social cost than men when negotiating. Riley-Bowles, Babcock, and Lai (2007) conducted four experiments to examine whether women encountered more social costs than men when negotiating for higher compensation. In experiment 1, participants respond on how likely they are to hire someone based on resume and interview notes under two conditions. The resume was identical across conditions with a gender-neutral first name. The gender of the candidate is then revealed in the interview notes. In the No Ask condition, participants are only provided with interview notes. Participants in Ask condition are provided with an additional note indicating that the candidate had asked for other benefits, including a higher compensation. Riley-Bowles et al. (2007) find that evaluators penalize a female candidate more for initiating negotiations than a male candidate.

In experiment 2, Riley-Bowles et al. (2007) adjust the scenario to make the job applicant an internal job candidate who had received an offer for a new position. Participants access one of six interview transcripts. There are three conditions in this experiment: No Ask, Moderate Ask and Strong Ask. They find that women face a higher social cost from negotiating for higher compensation than men. The evaluator's willingness to work with a male candidate remains unchanged when the male candidate attempted to negotiate.

Experiment 3 is similar to experiment 2, except participants now use a videotaped interview rather than a transcript to undertake the evaluation. With male evaluators, women incur a higher social cost for negotiating a higher compensation than men, but this does not happen when the evaluator is female. In experiment 4, participants are given two potential strategies regarding their salary and benefits offer, which are identical to the No Ask and Ask scripts in experiment 3. Women are found to be less likely to negotiate with a male evaluator.

2.1.3. Gender differences in risk preferences

Attitudes towards risk may be associated with the gender gap in leadership roles. The literature has suggested that there is a positive correlation between market wages and occupational choice. The standard explanation for this is that it is due to compensating wage differentials with agents who are risk-averse. Individuals with a lower willingness to take risks tend to self-select into occupations with lower earning variance. If women are more risk-averse than men, then they may not choose high-profile, high wage occupations.

In a lab experiment, Holt and Laury (2002) present participants with 10 different binary lottery choices, where participants must choose between two options for each paired lottery choice. Option A is a low-risk lottery where the potential payoffs differ slightly, whereas option B is a high-risk lottery where the potential payoffs differ widely. The payoff values for both outcomes remain unchanged, but the probabilities associated with the outcomes change. Each individual's level of risk aversion can be measured by the switching point from option A to option B. For example, a risk-neutral person would choose option A for the first four rows and then switch to option B in the fifth row. A risk-loving individual would switch to option B before the fifth row, while a risk-averse individual would switch after the fifth row. Women are found to be slightly, but statistically significantly, more risk-averse than men in the low-payoff treatment. This gender gap disappears in the high-payoff treatment.

Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2011) apply both survey data and field experiments to investigate an individual's willingness to take a risk. The survey data of more than 22000 individuals is sourced from the German Socio-Economic Panel, while their field experiment of 450 participants is drawn from the adult population. Participants in survey and field experiment answer the same general risk question. Participants in field experiment also make decisions in a lottery experiment with real monetary payoffs. They find that women are less willing than men to take risks in a number of different areas of life.

Croson and Gneezy (2009) review sizeable experimental literature on gender differences in risk preferences. They conclude that those experimental findings are broadly consistent with women being more risk-averse than men, both in lab settings and in investment decisions in the field. They investigate three explanations for the gender difference in risk preference. First, women are more emotional than men in risky situations, making them more risk-averse when facing a risky environment. Second, gender differences in overconfidence is another explanation for gender differences in risk preference and vice versa. Third, a difference in the interpretation of risky situations could make women more risk-averse.

Supporting the first explanation, Eriksson and Simpson (2010) conduct an international survey from India and the U.S. to test this hypothesis and find a consistent result that women are more emotional about outcomes than men. Supporting the second explanation, Niederle and Vesterlund (2007) claim that men are more overconfident about their relative task performance as compared to women. And, supporting the third explanation, Arch (1993) finds that women are more likely to see a risky situation as a threat, whereas men are more likely to view it as a challenge. Therefore, women tend to be more risk-averse and are therefore more likely to avoid risky situations than men. Notably, the gender difference in

risk preferences does not extend to managerial positions. Women in managerial positions have similar risk preferences as men.

Eckel and Grossman (2008b) review both lab and field experimental studies on risk preference. The findings from field experiments show that women are more risk-averse than men, and most studies in the lab experiments come to the same conclusion. However, Eckel and Grossman (2008b) argue that both lab and field experiments fail to take into account many factors such as knowledge, wealth, marital status and other demographic factors that may influence the results. Most studies summarized in Eckel and Grossman (2008b) and Croson and Gneezy (2009) compared gender differences in risky gambles or choice between gambles. One issue is the lack of comparability across studies as they differ in many factors, such as the structure of the gamble, the potential monetary payoff, and the degree of risk variance. Another issue is the consistency of measures of risk aversion across different treatments. Therefore, the findings do raise concerns regarding the measurability of risk attitude.

Charness and Gneezy (2012) provide an overview from 15 sets of experiments with one underlying investment game to examine whether the willingness to take on financial risk is higher for men as compared to women. Given that the relevant data is based on the same investment game and most of those experiments are designed to investigate gender differences, they conclude that the finding that women tend to be more financially risk-averse than men is robust.

2.1.4. Other personality traits

The willingness to guess is an individual trait that has the potential to influence performance in many situations. Individuals who are less likely to answer under uncertainty may also be less likely to propose opinions and offer advice. If women are less likely to volunteer their

ideas, it would be more difficult to recognize their talent which may hinder their career advancement.

Coffman (2014a) studies gender differences in the willingness to guess using lab experiments. Participants are provided with the U.S. and World History SAT II subject tests. In the "no penalty" treatment, participants get one point for every correct answer. In the "low penalty" treatment, participants get one point for every correct answer and are penalized one-quarter of a point for the wrong answer. In the "high penalty" treatment, participants are penalized one point for every incorrect answer. Baldiga (2014a) finds that women skipped significantly more questions as compared to men in the case where penalty is imposed. This gender difference remains after controlling for knowledge of the material, confidence level, and risk preferences.

Coffman (2014b) conducts lab experiments to examine gender differences in the willingness to contribute ideas to a group. Participants are randomly assigned into groups of two. Each participant faces five questions from six different categories: arts and literature, entertainment and pop culture, environmental science, history, geography, and sports and games. For each question, the participant decides how willingly to put his or her answer forward as the group answer. Coffman (2014b) finds that women are significantly less likely to contribute their ideas when the question is perceived as a "male-typed" question. Similarly, men are significantly less likely to contribute their idea when the question comes from a "female-typed" category. Coffman (2014b) concludes that women in male-typed domains are less likely to contribute ideas, which leads to a lower recognition of their expertise that is not driven by discrimination.

Babcock, Recalde, Weingart and Vesterlund (2017) argue that women's slow progress in organizational hierarchies may be due to their willingness to undertake "low-promotability"

tasks. Field evidence suggests that women are more likely to volunteer themselves into low-promotability tasks, e.g., serve on a faculty senate committee. To understand the gender differences in response to requests, they conduct lab experiments to examine the allocation of a task that everyone prefers to be completed by someone else.

In their first experiment, participants are anonymously paired in groups of three and can volunteer to invest. If one of the group volunteers within two minutes, then the volunteer receives $1.25 while the other two members of the group receive $2.00. If no member volunteers, then all members receive $1.00. The second experiment is the same as the first treatment, but all members are of the same sex and are aware of this. The third experiment is again similar to the first experiment, but an additional person is added. The additional person is called the requestor and is shown the photos of the other three members and chooses one who is then asked to invest. The requestor receives $1 if no one invests and receives $2 if an investment decision is made by any of the group members. The payoff for the group members are the same as the first two experiments. The fourth experiment is designed to explore the role of beliefs. In this experiment, groups play the same game as in the first experiment, but each member is then asked to predict the outcome of a randomly chosen session from the first experiment. The fifth and final experiment is designed to examine gender differences in altruism. In this experiment, participants are provided demographic information of the other participants in the three-person group. Each participant faces a sequence of 6 decisions, each time choosing between Option A and Option B. Option A stays the same through the sequence of decisions with payoffs the same as they would have been in the first experiment with the participant volunteering, that is, with payoffs of $1.25 to the participant and $2.00 to each of the other two group members. Option B varies over the decisions with payoffs being as they would have been in experiment 1 with a fixed one of the other group members, called Group member 1, volunteering with a probability that increases from 0 in the first decision

by increments of 20% to probability 1 in the final decision. Thus, the payoffs for this option increase from $1.00 for all group members in the first decision to $2.00 to the participant and Group member 2 and $1.25 for group member 1 in the final decision. The decision at which the participant switches from choosing Option A to choosing Option B is taken to be a measure of their altruism.

Overall, Babcock et al. (2017) find that women are more willing than men to volunteer, more likely to be asked to volunteer, more likely to accept requests to volunteer to make the low return investments and are expected by other participants to volunteer more than men. They interpret this as evidence that women are more willing to volunteer for low-promotability tasks. Examining the switching point from option A to option B in experiment 5, they find conclude that the gender difference in investing is not driven by altruism.

Exley and Kessler (2019) conduct lab experiments to examine gender differences in self-promotion. Participants are provided with 20 questions from the Armed Services Vocational Aptitude Battery (ASVAB). They find that women rate their past performance less favourably in than men, even in situations where they performed better than men. This gender gap in self-promotion notably persists after controlling for confidence and after participants are informed about their absolute and relative performance. This gap in self-promotion may contribute to the gender gap in leadership roles.

2.2. Perceptions towards the female leader

The aggregate statistics on labour force participation and educational attainment for women have improved notably. However, the gap in elite leadership roles between men and women remains. One explanation for the gap in leadership roles is discrimination or biased perceptions towards female leaders. Biased perceptions can lead to undervaluation of a

woman's effectiveness as a leader. Theoretical perspectives such as "expectation theory" (Ridgeway 2001), "role congruity theory" (Eagly & Karau, 2002), and "lack of fit theory" (Heilman, 2012) all support this explanation. These theoretical explanations posit the existence of a mismatch between gender stereotypes and job stereotypes.

The outcome of this mismatch is a negative expectation of job performance for both genders in gender-incongruent domains. The biased expectation towards women's performance leads to a biased process of information and employment decision-making. These negative expectations lead to the presumption that women do not fit the perceived requirements for success and lack the competence necessary for success in male-typed positions (Heilman, 2012). Consequently, this facilitates discrimination at various stages of women's careers.

2.2.1. Discrimination from below

Female leaders are evaluated more harshly compared to their male counterparts when they attempt to establish their authority in an authoritative manner, which is considered traditionally masculine (Eagly, Makhijani, & Klonsky, 1992). Women are traditionally assigned a lower status and are likely to be regarded as inappropriate if they attempt to demand or negotiate for higher levels of compensation.

Gangadharan, Jain, Maitra, and Vecci (2016) combined survey data with an artefactual field experiment in villages in the Indian state of Bihar regarding leadership in public good provision. They find that when women are group leaders, men contribute less compared to the case where men are group leaders. Gangadharan et al. (2016) name this behaviour as male backlash and suggest that this is driven by rooted social norms linked with male identity instead of real or perceived ineffectiveness of the women leader, or the women being thought of as tokens for powerful elites.

According to expectation theory, followers are more likely to ignore and dismiss actions by female leaders because they do not view them as legitimate leaders. (Ridgeway, 2001). Grossman, Komai, and Jensen (2015) conduct lab experiments to examine gender differences in leadership using a leader-follower collective action game. Participants are randomly assigned into a three-person group in each round, each with an endowment of $10, and must decide whether to invest in a joint investment project. The leader is randomly selected and needs to make an investment decision on the joint project after being informed of the group's payoff scenarios from the joint project. Grossman et al. (2015) use four treatments. Treatment with only male participants is referred to as "All Male treatment", and treatment with only female participants is referred to as "All Female treatment". "Gender Signalling treatment" and "No Gender-Signalling treatment" are conducted with mixed gender groups. In the "Gender Signalling treatment", followers are informed about the leader's gender in addition to the leader's investment decisions. Grossman et al. (2015) find those female leaders are more likely to invest in the single-gender treatment and the "No Gender-Signalling treatment" compared to male leaders. However, female leaders are less likely to invest in a mixed-gender environment where followers' refusal to follow imposes a negative impact on them. They do not find a difference in follower's perceptions towards female leaders.

Brands, Menges, and Kilduff (2015) claim that the follower's perception towards different leadership styles depends not only on the gender of the leader but also on the social network context of the work. They conduct three studies encompassing both experimental and survey data to examine gender differences in charismatic attributions. They find that female leaders are viewed as less charismatic compared to male leaders when the network structure is perceived to be centralized around one or a few individuals. On the other hand,

female leaders are viewed as more charismatic than male leaders when networks are seen as cohesive.

2.2.2. Discrimination from above

Female gender roles are incongruent with cultural images of leadership and are therefore perceived as systematically mismatched with a leadership role (Eagly and Karau, 2002). Riley-Bowles et al. (2007) show that evaluators view women as more demanding and less nice when they choose to negotiate compared to women who choose not to negotiate. They find that the social cost is significantly higher for women who negotiate for higher pay. On the other hand, this disinclination to work with employees who negotiate for compensation is generally negligible for men.

Bohnet, van Geen and Bazerman (2016) conduct lab experiments to test for gender bias in the evaluation of job candidates. Participants are assigned to either a "candidate" role or an "evaluator" role. Candidates participate in a verbal or a math task in stage 1, and they are paid according to their performance. In stage 2, evaluators are informed about the candidates' gender and past performance and are asked to select one of the candidates. In a separate evaluation treatment, the evaluator is offered a choice of a candidate who performed at the average or slightly sub-average level in stage 1 and offered the choice of that candidate or a random draw from the remaining candidates. In the joint-evaluation treatment the evaluator is offered a choice of two candidates a male and a female with one having average performance in stage 1 and the other slightly sub-average performance. The evaluator's payment is based on the chosen candidate's future performance. Bohnet et al. (2016) employ a 2 x 2 x 2 x 2 design. First, their design differs in evaluation mode, one is joint evaluation while the other one is a separate evaluation. Second, the level they report of the candidates' past performance. Third, they vary the gender of the candidate. Last, candidates participate in

either a mask task or a verbal task. The experiment setting involves a conflict between the individual performance information favouring one of the candidates and the group stereotype favouring the other candidate, assuming that women are stereotypically advantaged in the verbal task while men are stereotypically advantaged in the maths task. They find that evaluations are significantly influenced by the gender of the candidate consistent with the gender role thought to be congruent with the task. However, the joint-evaluation mode reduces this evaluator bias as evaluators in this mode focused more on individual performance rather than the gender of candidates.

Fernandez-Mateo and Fernandez (2016) study the sources of underrepresentation of women in top management roles using data from an executive search from a pool of 10,970 individuals by a search firm. They find that fewer women apply for top management roles than men and that they are slightly less likely to be interviewed, both when they are competing directly with a male candidate for a job and when they are not. However, in either case there is no significant difference in the likelihood of women being hired compared to men. They argue that the gender gap in top management roles are driven both by demand-side and supply-side actors at the initial stages of the recruitment process but that once women are considered for a position, the likelihood of them being hired is no less than it is for men.

2.3. Gender differences in Leadership Styles

Sandberg (2013) suggests that women should "lean in", implying that women should use a more assertive negotiating tactic, a trait usually associated with men. One approach to understanding the gender gap is by examining the differences between women and men in attributes relevant to leadership. A potential explanation for the gender leadership gap, especially in male-dominated areas of organizations, is that male and female leaders adopt different leadership styles.

Eagly and Johannesen-Schmidt (2001, p. 781) define leadership style as "relatively stable patterns of behaviour that are manifested by leaders". Leadership is generally regarded as a traditionally male role (Koenig, Eagly, Mitchell, and Ristikari, 2011), but research has separated the leadership styles into different categories. Rosener (1990) adopts the concepts first used by Burns (1978) to categorize different leadership styles. Men are described as "transactional" leaders who view job performance as a series of transactions with subordinates that involve rewards and punishments. Rosener (1990) finds that men are more likely to use the power that comes from their organizational roles. On the other hand, women are more "transformational" leaders who rely more on a democratic style and less on explicit rewards and punishments.

Eagly and Johnson (1990) undertake a meta-analysis of 162 studies on leadership and find little difference between male and female leadership styles. However, they find some support for the view that women adopt a more democratic style, while men tend to take a more autocratic style. Women are considered to be democratic because they are regarded as generous, understanding, compassionate, and concerned about others. In contrast, men are considered to be autocratic because they are regarded as aggressive, self-sufficient, rational, dominant, and competent.

Eagly, Karau, and Makhijani (1995) carry out a meta-analysis of studies looking at the effectiveness of male and female leaders. Their analysis reveals that women are more effective leaders in female-oriented organizational settings, while men are more effective leaders in male-oriented settings. Eagly, Johannesen-Schmidt and van Engen (2003) extended those findings and show that women are more likely than men to have a transformational leadership style, which is consistent with prescriptive gender roles, and is considered to be a more effective leadership style for them than the traditional male leadership styles. Engen and Willemsen (2004) carry out another meta-analysis and find that women adopt

transformational leadership styles more often than men. They also show that the type of organization and work setting is a significant influence on the differences in leadership styles.

Eagly and Carli (2003, p. 813) point out that "job candidates' leadership styles are surely among the attributes given special scrutiny in interviews of candidates for managerial positions, and managers fired from their positions are often critiqued for their leadership styles". Cuadrado, Navas, Molero, Ferrer, and Morales (2012) study gender differences in leadership styles and the impact of organizational outcome variables and find that male subordinates evaluate female leaders as more autocratic and aggressive than male leaders. This result could be due to the need of power to achieve success in their organization. Based on role congruity theory (Eagly & Karau, 2002), Cuadrado et al.'s study indicates that women who occupy leadership positions in "traditional"[1] organizations may conform to male norms in order to be viewed as legitimate in the leadership role. They conclude that "when female leaders work in contexts in which women predominate in managerial positions (that is, in decision making), and whose main activity is considered female sex-typed, they show a more congruent style with the feminine stereotype" (p. 3103). The outcomes of evaluations for female and male leaders depend on whether the organizational context is considered to be feminine or masculine.

Koch, D'Mello, and Sackett (2015) undertake a meta-analysis from experimental studies and examine gender stereotypes and bias in organizational decisions. Their analysis supports the role congruity theory (Eagly & Karau, 2002), which suggests that gender bias increases with the incongruence between stereotypical gender traits and the gender stereotype associated with the job. Koch et al. (2015) find that women are more likely to face

[1] In traditional organizations organization activities are divided into 1) male-sex-typed organization activities traditionally carried out by men, 2) female sex-typed organization activities traditionally carried out by women, and 3) gender-balanced organization activities that are not performed mainly by either sex.

discrimination in male-dominated organizational structures. However, there is no significant gender stereotype bias in female-dominated or integrated organizational settings for either gender.

2.4. Gender differences in Trust and Reciprocity

Buchan, Croson, and Solnick (2008) use the trust game to examine gender differences, both behaviourally and attitudinally. In their study, participants are divided into four conditions. In the "number identification" condition, senders and receivers are identified to each other by a number. In the other two "mutual name identification" conditions, either the sender's or the receiver's gender-identifying first name was revealed. In the last condition, the gender-identifying first name for both the sender and the receiver are revealed. They find men show a higher level of trust compared to women, and women are more trustworthy compared to men.

Chaudhuri and Gangadharan (2007) examine gender differences by using a within-subjects design for both a trust game and a dictator game. Participants play both roles in the trust game and the dictator game. They find that men are more trusting compared to women, but there are no significant gender differences in the level of reciprocity.

Haselhuhn, Kennedy, Kray, Van Zant, and Schweitzer (2015) examine how men and women respond to trust violations. They conduct three studies. Participants play seven rounds and are told they will be matched with a randomly selected person which in fact is a common, computer-simulated counterpart. In rounds 1-4, participants are exposed to trustworthy behaviour from the receivers (the computer-simulated counterpart), where the receiver returns at least half of the amount tripled. In rounds 5 to 6, participants are exposed to untrustworthy behaviour where receivers keep the entire amount passed to them. In round 7, participants are

told that this will be the last round. In study 2, participants are exposed to untrustworthy behaviour then followed by trustworthy behaviour. The result from study 1 and 2 show that women are more likely than men to regain trust and exhibit a higher level of trust after a violation. In their last study, participants are either assigned into a control treatment which involves no trust violation, or they are assigned into a trust violation condition. In this study, participants are provided with a typical business transaction context rather than the normal abstract setting of the trust game. They do not find any gender differences in the initial level of trust; however, women show a higher level of trust after a trust violation.

Dittrich (2015) uses heterogeneous participants from the German population instead of typical university students to play an anonymous online experimental trust game. Dittrich (2015) finds men exhibit a higher level of trust as well as a higher level of reciprocity compared to women. On average, men sent 45.11€-cent while women sent 38.56€-cent; women returned 38.7%, whereas men returned 41.6%. In addition, they find an inverse U-shaped relationship between age and trust (and reciprocity) for men, but no significant age effects for women.

Shahriar, Unda, and Alam (2020) examine gender differences in the repayment of microcredit. They recruited participants from rural Bangladesh to play a trust game and a microloan repayment game. Participants are randomly paired and assigned the role of the sender and recipient. Participants who played as a recipient in the trust game continue to play the microloan repayment game, which involves two treatments of individual loans or joint liability-based loans. They find that women are more reciprocal compared to men, and they are more likely to repay loans.

Kvaløy and Luzuriaga (2014) use the trust game to examine gender differences in trust and reciprocity when participants are playing with other people's money. They find no

significant differences between senders in typical trust game with senders who make decisions on behalf of others. However, women return less compared to men in both conditions. Findings from Kvaløy and Luzuriaga (2014) are similar to the result of Chaudhuri and Sbai (2011). Utilizing a trust game and a gift exchange game, they find no significant gender differences in trust under the gift exchange game, but women show a lower level of reciprocity compared to men. In addition, they find that any early gender differences dissipate over time; there are dynamic effects that are not captured in one-shot interactions. They conclude that gender differences reported in earlier research using one-shot plays might overestimate such differences, and real-life situations are better modelled by repeated interactions rather than one-shot games. Therefore, the dynamic aspects of behaviour are essential, and it appears that there are no strong gender differences in the level of trust or reciprocity in such repeated interactions.

3. Gender and Leadership

3.1. Introduction

The "gender gap" in the workplace is a well-documented phenomenon. However, it is important to recognize that this gap, in reality, refers to two phenomena, which are certainly related but still, not exactly the same. The first is the gender gap in wages: that mean/median earnings for men is higher than that for women. The second refers to the gender gap in leadership roles: There are fewer women as we move up organizational hierarchies.

In this chapter, we focus on the second phenomenon; the gender gap in leadership and explore two issues. First, we look at whether there are systematic gender differences in the willingness to lead. Second, we explore whether there are differences in the perception of female leaders compared to male leaders, in the sense of followers' willingness to follow male leaders more than female leaders. In particular, we intend to see whether messages from male and female leaders – even when the content of the message is identical – have a different impact on the actions of the worker. In order to study this issue, we rely on the minimum effort coordination game paradigm, which serves as good vehicle for simulating intra-organizational coordination problems. We use a modified version of the Brandts and Cooper (2006, 2007) corporate turnaround game. We explain the details of the game below in the section on experimental design.

We present results from two sets of studies. In one, both leaders and followers face the same payoffs while in another the leaders experience a payoff matrix that is different from that for the followers. Within these studies we manipulate two factors: in one treatment, the followers get to learn the leaders' gender while in another they do not. We also manipulate the nature of messages that leaders can send to their followers. In one treatment, the message sent is pre-determined and written by the experimenter. The leader only gets to choose how frequently to send this message. This ensures that, if and when sent, the message coming from

the leader is identical across genders. In a second treatment, leaders are allowed to write free-form messages. This generates a 2x2 protocol consisting of four treatments: (1) Leader's gender not revealed; pre-set message; (2) Leader's gender revealed; pre-set message; (3) Leader's gender not revealed; free-form message and (4) Leader's gender revealed; free-form message.

Overall, we find that fewer women volunteer to the leadership role compared to men. This is true whether the payoff matrix for the leaders is the same as that of the followers or not. But the decision to lead or not does depend crucially on whether the leader's gender is revealed to group members or not. When gender information is revealed, there is a significant gap in the rate at which men and women volunteer to lead with men being much more willing; when gender is not revealed, this gap shrinks considerably. However, when it comes to leader effectiveness, we do not find evidence of significant differences. By and large, messages and actions taken by leaders are similar, even if not identical, and this, in turn, implies that groups led by women enjoy similar levels of coordination success as those led by men.

Our results provide evidence that the gap in leadership roles may be due, at least partly, to supply side rather than demand side factors. If women are more reluctant to avail of leadership roles, then this may partially explain the existing gender gap in leadership. To an extent, this reluctance to volunteer for leadership may be predicated on the assumption that followers are less likely to follow female leaders. We do not find evidence of any systematic resistance to female leadership. This seems to offer a learning opportunity that oft-held presumptions may not be necessarily correct and provides support for more pro-active equity-based practices in the workplace. We proceed as follows. In Section 2, we provide a brief review of the relevant background literature. In Section 3 we outline our experimental design and procedures. We present our results in section 4 and make concluding remarks in Section 5.

3.2. The gender gap in wages and leadership roles

Despite the progress in female educational attainment and increasing parity in paid hours of work, and occupational choice, women are still earning less than men under same or similar jobs (Goldin, Katz and Kuziemko, 2006; Bertrand, Goldin and Katz, 2010; Blau and Kahn, 2000; 2006, Drolet, 2001, Goldin, 2014, Weichselbaumer and Winter-Ebmer, 2005). This is true not only for women in the general labour force, but also for those who have graduated from MBA and other professional programs, presumably with the aim of pursuing ambitious managerial/professional careers. (Babcock and Laschever, 2003; Bertrand, Goldin and Katz, 2010; Carter and Silva, 2010). Wood, Corcoran and Courant (1993), in a study of law school graduates find that even after controlling for the "motherhood penalty" and other factors such as school performance and work-history, one-third to one-quarter of the wage gap is left unexplained. Jena, Olenski and Blumenthal (2016) confirms a significant gender gap in wages among more than 10,000 physicians at 24 public medical schools in United States after controlling for factors such as age, experience, specialty, faculty rank, measures of research productivity and clinical revenue. Moreover, the gender gap in wage at the top of the wage distribution has declined much more slowly than at the middle and bottom (Blau and Kahn, 2017).[2]

We refrain from elaborating any further on the gender gap in wages since our focus in this paper is on the gender gap in leadership. Women currently hold 5% of CEO positions in S&P 500 companies. Among the CEOs of the largest publicly listed companies in the European Union, only 5.5% are women. Women constitute only 9.4% of the 540 C-level

[2] The OECD defines the gender wage gap as the difference between median earnings for males and females relative to the median earnings for males. It should be noted that there are significant cross-country differences. Looking at data for 2015-2018, we find that the average gender wage gap for all OECD countries is 13.5%. At the most unequal end we have Korea (with a difference of 35%), followed by Japan (25%), Israel (22%) and then Canada, USA and Finland (around 18%). At the other extreme, we have Belgium, Greece, Costa Rica, Denmark and Italy, all hovering around 5% with Ireland, Norway and Sweden at about 6% and New Zealand at a little less than 8%.

executives among Canada's 100 largest publicly traded companies. (Catalyst, 2018). Sandberg (2013) comments that at the time of writing: Of the 195 independent countries in the world, only 17 were led by women. Women held just about 20% percent of seats in parliaments globally and about 14% of executive officer positions, 17% of board seats, and constituted 18% of elected congressional officials in the US.

Of course, if more men than women are occupying higher paid jobs, this will also show up as a higher average wage for men. Some of the factors that contribute to the gender gap in wages almost certainly also create the gender gap in leadership roles. Early research in the gender wage gap attributed this phenomenon mainly to differences in human capital and either taste-based or statistical discrimination. (E.g., Altonji and Blank, 1999). But, as Bertrand (2011) points out, in recent years a large body of research – to a large extent experimental – suggests the possibility that observed gender gaps may arise due to important differences in psychological attributes and preferences between men and women. Such differences may include gender differences in risk preferences, in attitudes towards competition and negotiation and in other-regarding preferences. This, in turn, may also have implications for gender differences in occupational choice, work-place strategies and consequently wages and career advancement.

Experimental evidence suggests that women tend to shy away from participating in highly competitive environments (Gneezy, Niederle and Rustichini, 2003; Niederle and Vesterlund, 2007); are less likely to engage in negotiations for pay and promotions (Babcock and Laschever 2003; Exley, Niederle and Vesterlund 2016) and are more risk-averse compared to their male counterparts (Croson and Gneezy, 2009; Charness and Gneezy, 2012, Coffman, 2014). Evidence suggests that as the environment becomes more competitive, the performance and participation of men increase relative to that of women. (Andersen, Ertac, Gneezy, List and Maximiano, 2013; Vandegrift & Yavas, 2009, Gneezy, Leonard and List, 2009). In addition, recent studies have found that women are more willing to take on tasks

that are less likely to benefit their performance evaluations or advancement prospects whereas men tend to focus more on high-promotability tasks. (Babcock et al., 2017).

Babcock and Laschever (2003) and Niederle and Vesterlund (2007) argue that greater female aversion towards competition may explain why one finds fewer women occupying positions of power. Sandberg (2013) suggests that in order to be successful in the work-place women need to adopt more assertive negotiating tactics, i.e., they need to "*lean in*" more, a trait usually associated with males. Bohnet (2016) not only provides a comprehensive overview of what we currently know about the gender gap, but also makes numerous policy suggestions as to how we can go about creating a more equal workplace.[3]

The experimental literature looking at gender differences is large. We refer the interested reader to Eckel and Grossman (2008a) and Croson and Gneezy (2009) for reviews primarily with regards to differences in risk attitudes and other-regarding preferences. Babcock and Laschever (2003) and Bohnet (2016) provide comprehensive discussions of the topic from a labour economics perspective.[4]

This line of inquiry into the gender gap in leadership overlaps with research in leadership; specifically whether men and women tend to adopt different styles when it comes to dealing with employees. Rosener (1990), building on concepts introduced by Burns (1978), argues that men typically tend to be "*transactional*" leaders and see job performance as a series of transactions with subordinates involving rewards for services rendered and punishments for inadequate performance. Women on the other hand are seen as being more

[3] Such include, for instance, the introduction of "blind" auditions for orchestras (Goldin and Rouse, 2000) or the use of structured interview questions with the same questions being asked in the same sequence of all job candidates as well as evaluating job candidates contemporaneously rather than sequentially. (Bohnet, van Geen and Bazerman, 2015).

[4] For a selection of findings from the psychological point of view, see Walters, Stuhlmacher, and Meyer (1998) or Sax (2005).

"*transformational*", relying less on explicit rewards and punishments and more on a democratic and participative style.

Eagley and Johnson (1990) undertake a meta-analysis of 162 studies on leadership and find little difference between male and female leadership styles. They find some support for the view that women adopt a more democratic style while men tend to adopt a more authoritative style.[5] Eagley, Karau, and Makhijani (1995) undertake a further meta-analysis which extends the analysis of leadership styles to the issue of leadership effectiveness. They report that men and women are equally effective as leaders except that men tend to be more effective in occupations that are typically defined in more masculine terms such as the military while women are more effective in occupations defined in primarily feminine terms such as nursing.[6]

However, some other studies find that men are perceived as more effective leaders because gender role expectations spill over into leadership roles. Indeed, men are found to exert more influence than women in mixed groups and are more resistant to female leadership. (Gangadharan, Jain, Maitra and Vecci, 2016). Followers are likely to ignore and dismiss actions by female leaders because they do not view women as legitimate leaders. (Ridgeway, 2001). Grossman, Eckel, Komai and Zhan (2019) find that when leaders are appointed exogenously, women are assessed less positively and rewarded less generously than equally effective men in resolving intra-organizational coordination problems and Reuben and Timko (2018) find in the case of elected leaders, male leaders enjoy greater benefits, at least in the initial interactions.[7]

---

[5] The authors suggest that these differences may arise in part from the fact that women, being outnumbered by men as leaders, face greater resistance from employees and feel the need to seek greater employee input.
[6] See Moran (1992) for a succinct overview of much of this work.
[7] Gallup has been tracking workers' preferences for their superior's gender since 1953 when 66 percent of American adults said they would prefer a male boss and only 5 percent said they would prefer a female boss. Although this gap has narrowed as of 2013, it still remains. According to 2013 Pew Research Centre survey, around three-quarters of men and women who are currently working or ever worked have no preferences regarding their boss's gender, but among those who do have a preference, both genders are more likely to prefer a male boss.

3.3. Experimental Design and Procedure

3.3.1. Experimental Design

We rely on a modified version of the Brandts and Cooper (2006, 2007) corporate turnaround game which, in turn, is derived from the minimum effort (weak-link) game of Van Huyck, Battalio and Beil (1990). We provide the instructions in the Appendix 1. Each session consists of 20 rounds with the first ten rounds being identical. Participants are randomly assigned to groups of five at the start of the session, and group composition remains unchanged for the entire session. In each of the first ten rounds, each participant simultaneously chooses an effort level {0, 10, 20, 30, or 40}, where earnings depend on the participant's effort choice and the minimum effort level chosen in the group in that round. Earning for each player is determined by the underlying equation:

$$\pi_i = 400 - 5(E_i) + B\left(\underset{i\in\{1,2,3,4,5\}}{\text{Min}}(E_i)\right) \quad (1)$$

where π_i indicates earning for player i, E_i is the effort level player i chooses and B is constant set equal to 10 and *Min* (E_i) is the minimum effort level chosen in the group for that round. The earnings are depicted in Table 3.1. The values are denoted in experimental dollars with 400 experimental dollars equal to 1 New Zealand dollar.

Table 3.1: Payoff matrix for (i) first 10 rounds of Experiment 1 and (ii) all rounds of Experiment 2

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Choice of hours by particular a player	0	400	400	400	400	400
	10	350	450	450	450	450
	20	300	400	500	500	500
	30	250	350	450	550	550
	40	200	300	400	500	600

Note: This payoff matrix shows the payoff to all players for the first 10 rounds of Experiment 1 (before leaders are selected) and for all 20 rounds of Experiment 2 (both before and after leader selection)

In this payoff matrix, any common effort level (hours) chosen by the five players constitutes a Nash equilibrium with everyone choosing 40 hours being the payoff dominant outcome while everyone choosing zero hours is the secure outcome. Each player faces the dilemma that higher effort level (and therefore, higher disutility of effort) is required in order to achieve higher earnings, but a higher effort level also entails more risk in the form of reduced payoffs, if even one member of the group deviates and chooses a lower effort level. Earlier studies suggest that over time play typically approaches the secure Nash equilibrium where all players choose the lowest effort level and that players find it difficult to coordinate to the payoff dominant equilibrium for any length of time. (Van Huyck et al., 1990; Knez and Camerer, 1994, 2000, Chaudhuri, et al. 2009).

Prior to the start of round 11, participants are given an opportunity to volunteer to be the group leader. If more than one person wishes to be a leader, then the leader is chosen randomly from the group of contenders.[8] We conduct **two separate** experiments. In **Experiment 1**, following the selection of the leader, the payoff matrix from the leader changes. The leader's payoff is now determined by the following equation:

$$\pi_L = 100 + [(60 - 4B) \times \min_{i \in \{1,2,3,4,5\}} (E_i)] - \mathrm{c} \qquad (2)$$

If the leader's choice of hours exceeds that of the workers, then there is an additional cost to the leader of the form:

$$c = 2\left(E_L - \mathrm{Min}(E_{(-i)})\right) \qquad (3)$$

where E_L is the leader's effort level and Min $(E_{(-i)})$ refers to the effort level chosen by the all group members excluding the leader. We think of this as a psychological cost of feeling "ripped-off" when the leader chooses an effort level that is higher than that chosen by one or more followers. Table 3.2 shows the payoff for leaders.

[8] A leader is chosen randomly among group members if no one volunteers. But it never happened that there were no volunteers.

Table 3.2: Payoff matrix for leaders for rounds 11 through 20 in Experiment 1

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Leader's choice of hours	0	100	100	100	100	100
	10	80	300	300	300	300
	20	60	280	500	500	500
	30	40	260	480	700	700
	40	20	240	460	680	900

Note: In Experiment 1, all players face the payoff matrix shown in Table 1 for rounds 1 through 10 before selection of leaders. For rounds 11 – 20, the followers continue to play with that same payoff matrix, while the payoff matrix for the leaders change to the one shown in Table 2.

In **Experiment 2**, the payoff matrices for both leaders and followers are the same for all 20 rounds and shown by the payoff matrix in Table 3.1. Other than the change in the leader's payoff matrix in Experiment 1, the two studies are conducted in the exact same way. Clearly, the payoff matrix in Table 3.2 (for Experiment 1) presents both gains and losses for the leader. We note a few points. From Table 1, it is clear that any team member can guarantee him or herself a payoff of 400 by choosing zero hours in the secure equilibrium. Now, looking at the leader's payoff matrix in Table 3.2, we find that if the minimum effort in the group ends up being zero or 10 hours, then the maximum the leader can earn is 300. It is only when the team manages to coordinate to a minimum effort level of 20 hours or more that the leader makes more than 400, which any team member can make in the secure equilibrium. If the team minimum is 20 hours, then the leader makes 500. The leader makes more if the team manages to coordinate to a higher minimum.

If we find greater female reluctance to lead in Experiment 1 where the leader's payoff changes, then this may partially be attributable to differences in risk attitudes. To control for this, we run Experiment 2[9], where the payoff matrix remains unchanged for both leaders and followers. We also control for risk preferences using the Holt-Laury lottery choice mechanism

[9] Experiment 2 only partially controls for risks, some risks remain. The fact that we do not get different results for experiment 2, may due to the fact that the gender difference is not due to differences in risk aversion and/or because we have not sufficiently eliminated the risk.

(Holt and Laury, 2002). In what follows, we will periodically remind readers of the difference between Experiments 1 and 2; an easy mnemonic is "*1 for change, 2 for none*"; meaning that in Experiment 1, the leader's payoff changes following leader selection and prior to Round 11 while in Experiment 2, both leaders and followers continue to face the same payoff matrix following leader selection.

3.3.2. Treatments

We vary the experimental design along two dimensions: whether the gender of the leader is revealed or not; and whether the leader sends a fixed message, that is provided by the experimenter or the leader is allowed to write free-form messages. This generates a 2X2 protocol consisting of four treatments: (1) Leader's gender not revealed; pre-set message; (2) Leader's gender revealed; pre-set message; (3) Leader's gender not revealed; free-form message and (4) Leader's gender revealed; free-form message.

In our pre-set message treatments, participants are informed that if they choose to be a leader, then they will be provided with a message that they can send to the other members of their group. In each round, the leader moves first by choosing the number of hours. The leader can also choose to re-send the fixed message or not. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours and the content of the message will be revealed to the group members. The fixed message that leaders could disseminate is:

> *You should choose to work 40 hours in each round. NOTICE, from the payoff matrix, that if every participant in a group follows the message then every participant will earn 600 experimental dollars. However, if even one of the participants does not*

follow the message and chooses a number different from 40, then each participant will make less money than if everyone chose 40.[10]

In free-form message treatments, leaders are asked to type a message which they can send to their group members. Figure 3.1 shows a screenshot for a leader at the beginning of round 11 under Free-form message treatments. They can write a different message each round. In each round, the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or write a different message. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours and the content of the leader's message is revealed to the group members. The employees then choose their hours simultaneously. It is important to note that the person selected to be the leader must send a message prior to Round 11. From that point, the leader gets to choose whether to send any further messages or not. Not sending any more messages is an option as is the option to send a message prior to every round.

[10] An astute reader may note that in Experiment 1, this is true of all the followers but not of the leader since the leader's payoff matrix changes and the leader gets to make more than 600 if everyone coordinates to 40 hours. This instruction is accurate for Experiment 2, where there is no change in payoffs for either followers or leaders. We decided to use this language in the interests of consistency. Otherwise, we would have had to use the word "followers" for Experiment 1 and "participants" for Experiment 2. This would have introduced a potential confound. In Experiment 1, we felt that given that the followers are getting this message *after* they have seen the leader's effort choice and the leader's message, it would be obvious in the context that here participant means the followers, who are receiving this message. In Experiment 2, this is correct that choosing 40 hours means all participants earn 600 units in each round.

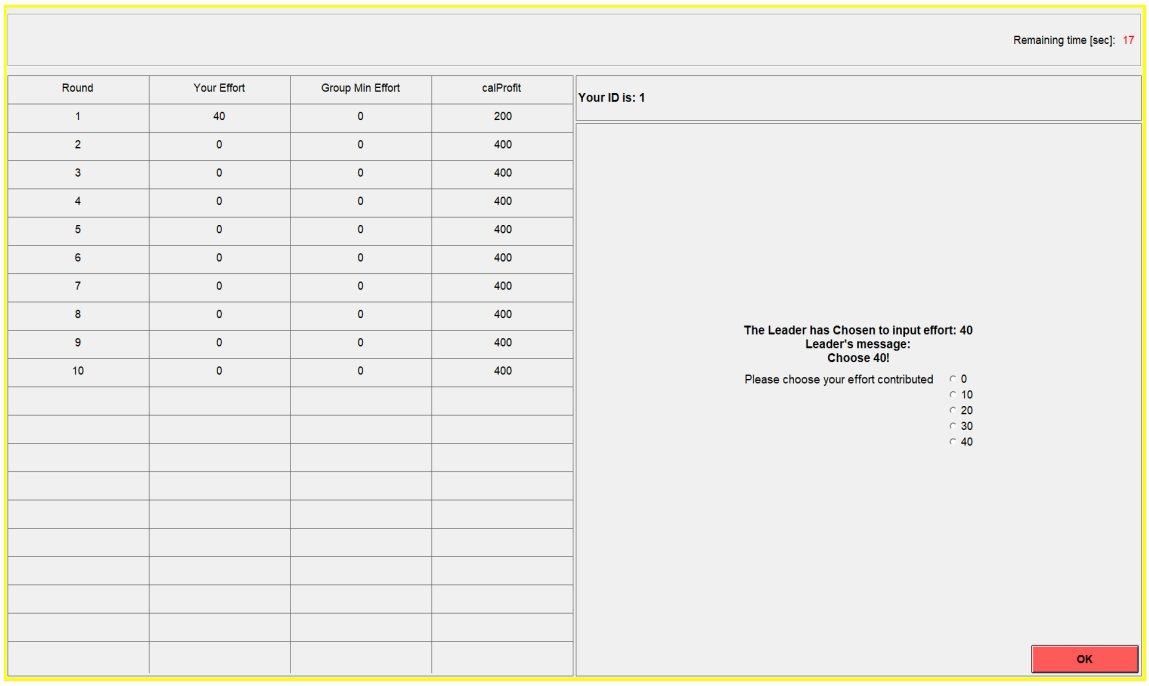
Figure 3.1: Screenshot for Leader under Free-form message treatment

In gender revealed treatments, the leader's gender is revealed to the group members in addition to the choice of hours and content of the message. Figures 3.2A and 3.2B are screenshots for team members in Free-form message treatment with gender not revealed (Figure 3.2A) and revealed (Figure 3.2B) at the beginning of round 11. The main difference is that in the former the followers do not know the leader's gender while in the latter this information is available on the screen. Each participant sees a screen that indicates if he or she has been selected as a leader. If not, the screen shows the ID number and gender of who has been selected as the leader. Each participant sees a screen that indicates if he or she has been selected as a leader. If not, the screen shows the ID number and gender of who has been selected as the leader. Leaders are identified by ID number only. So, in the gender revealed treatments, the gender information is made salient over multiple screen messages. [11]

[11] In this experiment, we chose to have the leader's gender being revealed as being male or female; in the sense that the followers get a message saying: "Your leader is female/male." One drawback to doing things this way is that this may create experimenter demand effects. There are different ways one can go here and all of them pose difficulties. One option used often is to ask participants choose from a list of proto-typical male or female names, typically European ones such as Adam or Lisa. But the University of Auckland is one of the most international universities in the world. We were uncertain whether asking a large number of non-European

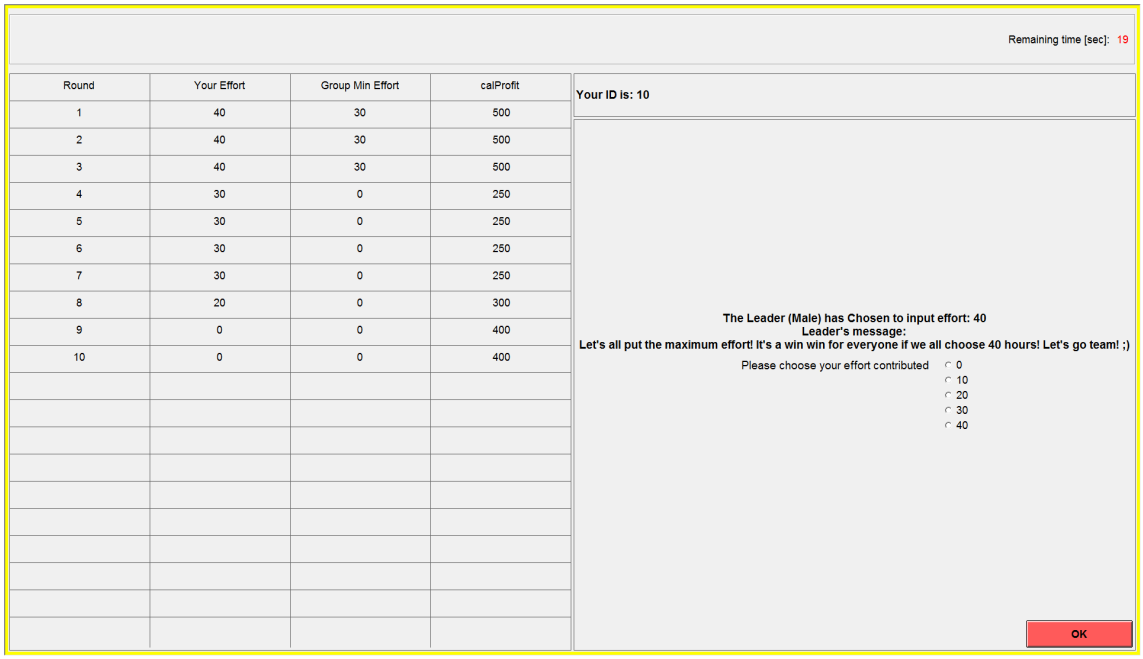
Experiment 1 results clearly demonstrate that any differences, if they arise at all, are more prominent in the gender revealed treatment. Therefore, in experiment 2, we only conduct the gender revealed treatments.

Figure 3.2A: Screenshot for followers in Free-form message <u>gender not revealed</u> treatment



students to explicitly choose an European pseudonym would be well received. On the other hand, if we used non-European names, it was not clear if the gender would automatically be salient to everyone. Other studies have used virtual avatars. It is our feeling that asking participants to adopt another name or a virtual avatar etc., at times, start to approach the line separating truth and deception. There is increasing scrutiny and concern about deception in experiments even if benign. We decided to avoid this and to go with the label male/female. This is partly because as Zizzo (2010) points out: experimenter demand effects become less of a concern where the demand effect explicitly relates to the variable, which is being manipulated and is of primary interest. Here, as in many other cases, one faces design trade-offs. We chose to go with this particular approach.

Figure 3.2B: Screenshot for team members in Free-form message gender revealed treatment



Prior to choosing whether to volunteer as a leader, each group member knows whether the leader's payoff will change and how (Experiment 1) or not (Experiment 2). Each person also knows whether they are in the pre-set message treatment (i.e., the message is provided to them by the experiments and all they need to do is to decide how often to send this message out to group members) or the free-form one. They also know that the leader has to send a message prior to Round 11 but can choose whether or not to do so for the subsequent rounds.

3.3.3. Post-experiment tasks

Following the conclusion of the 20 rounds of the coordination game, participants are asked to take part in the Holt and Laury (2002) lottery choice game with choices being incentivized. Participants are presented with 10 different binary lottery choices and must choose between either option A or option B for each paired gamble. (See Appendix 4.)

An individual's degree of risk aversion is determined by the point at which that person switches from option A to option B. A risk-neutral person would switch from option A to B

at the fifth choice. An individual who switches to Option B earlier than the fifth choice (i.e. choices 1 through 4) is risk-loving, while an individual who switches to option B at the sixth choice or later is risk-averse. Participants are informed that this is a separate task which will be paid according to their choices. At the end of the session the computer randomly chooses one of the gambles (rows) and each participant is paid based on whether (s)he chose Option A or Option B for that gamble. In our overall sample, we have 27% of all participants with inconsistent risk preferences, 42% of all participants are risk-averse, 25% of all participants are risk-neutral and 6% of participants are risk-loving. In our regression analyses below, when we control for risk preferences, we will ignore the participants with inconsistent preferences, leading to a loss in the number of observations in some cases.[12]

After completing the lottery task, participants completed a demographic survey (see Appendix 8) collecting information regarding participant's gender, field of study, year in the undergraduate program, age, income, whether they were born in New Zealand, and their ethnicity. In some sessions, we also asked participants to fill out a questionnaire regarding their attitudes toward female leadership.[13]

[12] The existence of subjects who demonstrate inconsistent preferences in the Holt-Laury task is a well-documented drawback of this particular way of measuring risk preferences.

[13] Participants are asked to answer the following five questions. (1) Females do not possess good leadership quality. (2) Because leadership is viewed as a masculine trait, females will not be viewed as strong leaders. (3) Females are too emotional to lead effectively. (4) Females are capable of performing effectively in any leadership position. (5) Male leaders connect with the public better than female leaders. All questions are answered on a 1-5 scale from "Strongly Disagree" to "Strongly Agree" with appropriate scores reverse scored as relevant. Using various non-parametric tests, we do not find any significant differences between male and female responses regarding their view of female leaders. Therefore, in what follows, we do not elaborate on these survey responses.

3.3.4. Experimental Procedure

A total 330 students took part in Experiment 1, and 85 students took part in Experiment 2. Tables 3.3 and 3.4 provide details for experiment 1 and 2, respectively. All sessions in this experiment were conducted in the DECIDE laboratory at The University of Auckland using Z-tree (Fischbacher, 2007). Participants were recruited via an email announcement and they were students from undergraduate courses without any prior experience with the turnaround game.

Table 3.3: Number of participants in different treatments in Experiment 1

	Pre-set message; Gender not revealed	Pre-set message; Gender revealed	Free-form message; Gender not revealed	Free-form message; Gender revealed	Total
Male	21	40	18	86	165
Female	24	45	17	79	165
Total	45	85	35	165	330
Total 5-person Groups	9	17	7	33	66

Table 3.4: Number of participants in different treatments in Experiment 2

	Pre-set message; Gender revealed	Free-form message; Gender revealed	Total
Male	24	17	41
Female	21	23	44
Total	45	40	85
Total Groups	9	8	17

Participants are directed to computer cubicles once they enter the lab. There are dividers between each cubicle so that participants are separated from one another are unable to see any other participant's screen. This prevents any opportunities to observe the decisions made by any other participants during the experiment. They are also cautioned against communicating with others. Participants know that they are randomly assigned to a group of five and that the group composition will remain unchanged for the entire time. Each

participant is assigned a subject ID number and never learn the actual identity of any of the others in his or her group. Participants know that the experiment consists of two parts of 10 rounds each. We read them the instructions for the first 10 rounds at the beginning. Participants then play the first 10 rounds of the game. After the conclusion of the first ten rounds, participants receive the instructions for the second part of the experiment. The instructions are read out loud to them. Once they are finished reading the instructions and ready to proceed, they first see a screen which asks them whether they wish to volunteer to be a leader or not. Once all participants have entered their decisions, one of them is randomly selected as the leader, other participants get to know who was selected as leader (via subject ID only), and in the gender revealed treatments, participants also learn whether this leader is male or female. Then, the leader enters his/her effort choice for Round 11 along with the message. In the pre-set message treatments, the leader chooses effort as usual and also chooses whether to send this pre-set message to the followers. The followers then get to see the leader's message and effort choice and enter their own effort choices. Everyone then gets to see the outcomes (the effort choices of the leader and of the followers) and earnings from that round. In the free-form message treatment, the procedures are similar. Except, in Round 11, the leader has to write a free-form message besides making an effort choice. From Round 12 onwards, the leader has the choice of whether to write the message or not.[14]

Each session lasts approximately 60 minutes. At the end of the session, participants are paid their earnings in cash plus a $5 show-up fee. Participants are told that their earnings are private information and they should keep this private. Once paid, they are free to leave. Average payoff for the leadership game is NZ $29 including the show-up fee. In addition, participants would have earned another additional amount of approx. NZ $5, on average, depending on the outcome of the Holt-Laury lottery-choice game.

[14] We provide two diagrams in the Appendix 5 to show the experiment procedure.

3.3.5. Hypotheses

First, the literature regarding gender and competition suggests that volunteering for leadership may well be perceived as entering into a competitive situation and if so, then we might expect women to exhibit a lower willingness to volunteer. Second, the result of Gangadharan et al. (2016) suggest that women may anticipate a backlash from followers and therefore, may be less willing to volunteer in those treatments where the leader's gender is revealed. Last, based on the results of Grossman et al. (2016), we might not anticipate differences in followers' response to male and female leaders. Our three hypotheses are list as follows:

Hypothesis 1: Women will be less likely to volunteer themselves for leadership positions than men.

Hypothesis 2: The tendency of women to be less likely to volunteer themselves for leadership positions than men will be stronger in gender revealed treatments.

Hypothesis 3; There will be no significant gender difference in followers' response to male and female leaders.

3.4. Results

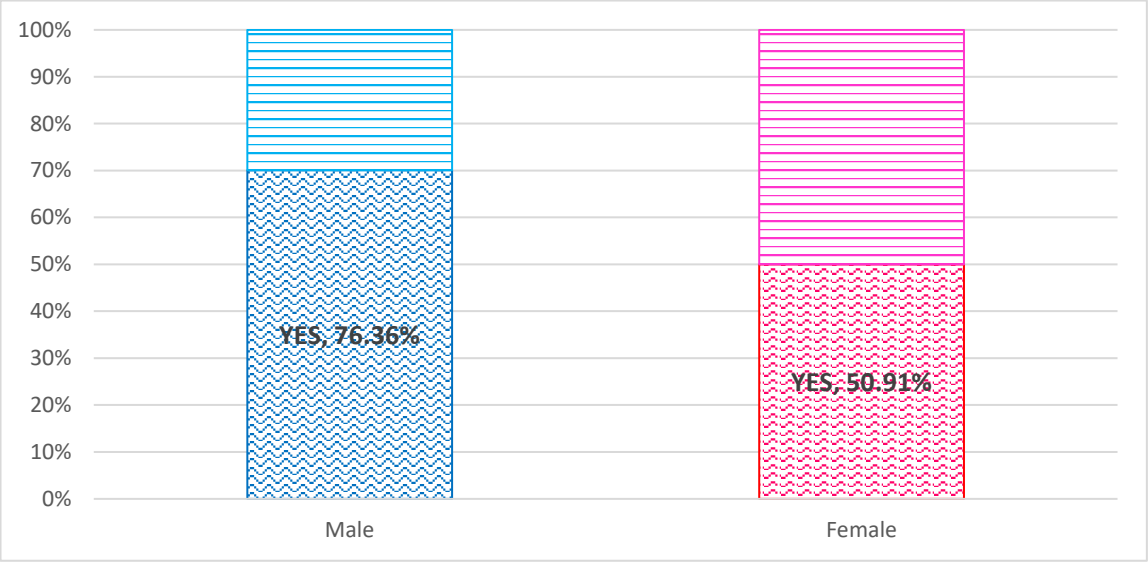
We report our results in two sections. First, we examine whether men and women differ systematically in their willingness to lead. Second, we look at the impact of leaders' gender on the coordination level of group members in both pre-set message treatment and free-form message treatment.

3.4.1. Willingness to lead

As noted, upon the conclusion of the first 10 rounds and prior to beginning Round 11, participants are asked whether they wish to be a leader or not. We find that in both Experiment 1 (where the leader faces a different payoff matrix) and Experiment 2 (where the payoff matrix remains unchanged) a much larger proportion of men volunteer to lead. Figure 3.3, Panels A

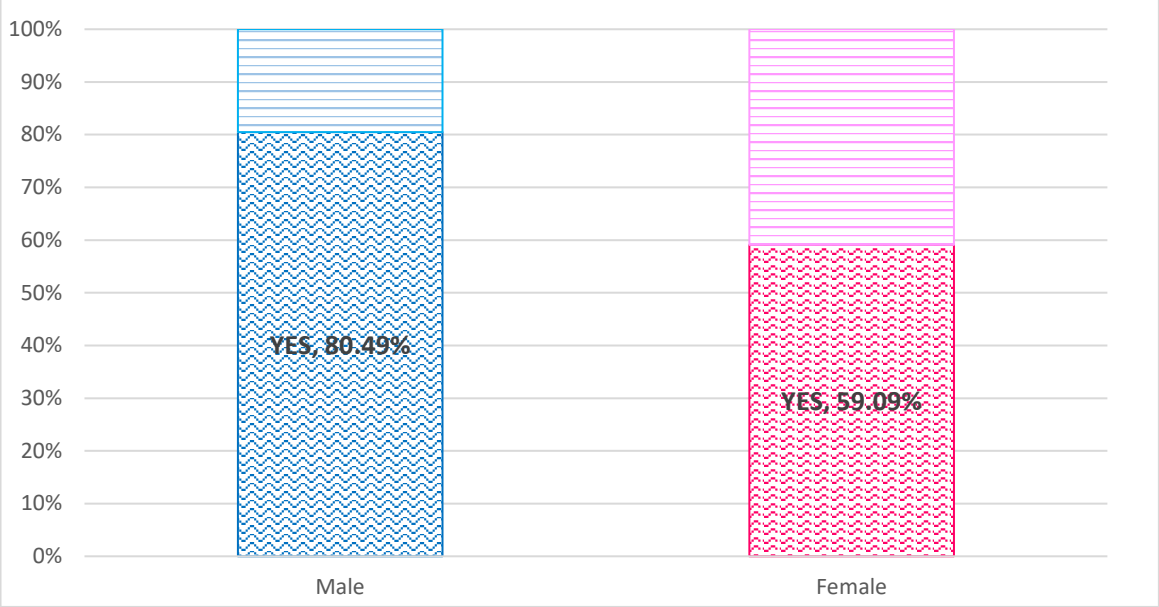
and B show the breakdown. In Experiment 1, 76% men volunteer to lead as opposed to only 51% of women. There is a similar discrepancy in Experiment 2, with nearly 80% of men expressing a willingness to lead while only about 60% of women do so. The fact that there is not much change between experiment 1 and experiment 2, suggests that it is not the change in the payoff matrix that is driving the results.

Figure 3.3A: Difference in the willingness to lead in Experiment 1



Note: Data for 165 males and 165 females from Experiment 1; simple count of how many men and women volunteered to lead prior to Round 11

Figure 3.3B: Difference in the willingness to lead in Experiment 2



Note: Data for 41 males and 44 females from Experiment 2; simple count of how many men and women volunteered to lead prior to Round 11

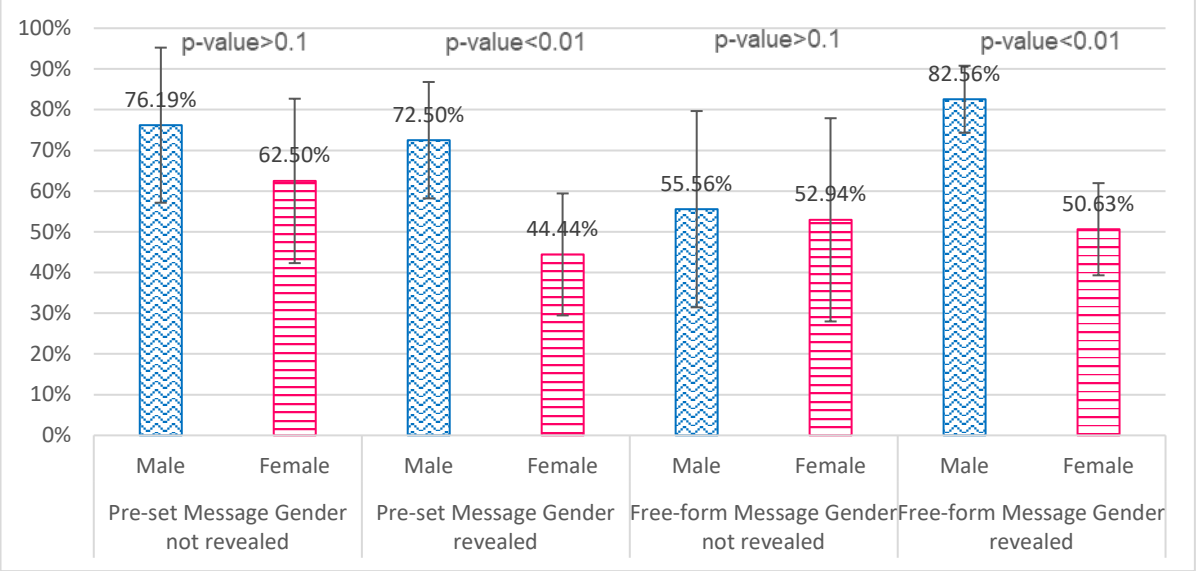
In Figure 3.4, Panels A and B, we provide a more detailed break-down of what happens when the leader's gender is revealed to followers as opposed to when it is not. We start by looking at Panel A of Figure 3.4, which shows the gender breakdown for the pre-set and free-form message treatments separately. This figure reveals some striking differences. It is clear that women are much more reluctant to lead, when the leader's gender is made known to the followers. If we look at the first and third pair of bars, which represent the situation when gender is not revealed, we find that any gender differences in the willingness to lead are not large. 76% of men volunteer as opposed to 63% women in the pre-set message gender not revealed treatment. The corresponding figures are 56% for men and 53% for women in the free-form message gender not revealed treatments. Using a sample proportions test, these differences are not statistically significant.

But if we now compare the second and fourth pairs of bars, for the gender revealed treatments, then the differences are dramatic. In the pre-set message gender revealed treatment, 73% of men express a willingness to lead, while only 44% of women do so. In the free-form message, gender revealed treatment, the differences are even more pronounced; 83% for men and 51% for women. Both of these differences are highly significant using a sample proportions test. ($|z|=2.84$, $p<0.01$, m=40, f=45 for the pre-set message gender revealed treatment and $|z|=5.33$, $p<0.01$, m=86, f=79 in the free-form message gender revealed treatments.)[15] The results of Experiment 2, where we only look at the gender revealed condition, are similar except that the proportions of men and women volunteering to lead in the pre-set message treatment are not significantly different ($|z|=1.05$, $p>0.1$, m=24, f=21)

[15] In order to increase the statistical power of our tests, we also carry out similar tests after pooling the data across the two message treatments in Experiment 1. Given that the decision regarding leadership is taken before sending any messages, we look at the gender-revealed and gender not revealed treatments in aggregate. This allows us to compare the decisions made by 39 males and 41 females in the two message treatments where gender is not revealed. There are no significant differences here. ($|z|=0.76$, $p>0.1$). But if we combine over the two message treatments, where gender is revealed, then we have 126 decisions by males and 124 by females. In this latter case, there is a significant gender difference in willingness to lead ($|z|=5.37$, $p<0.01$).

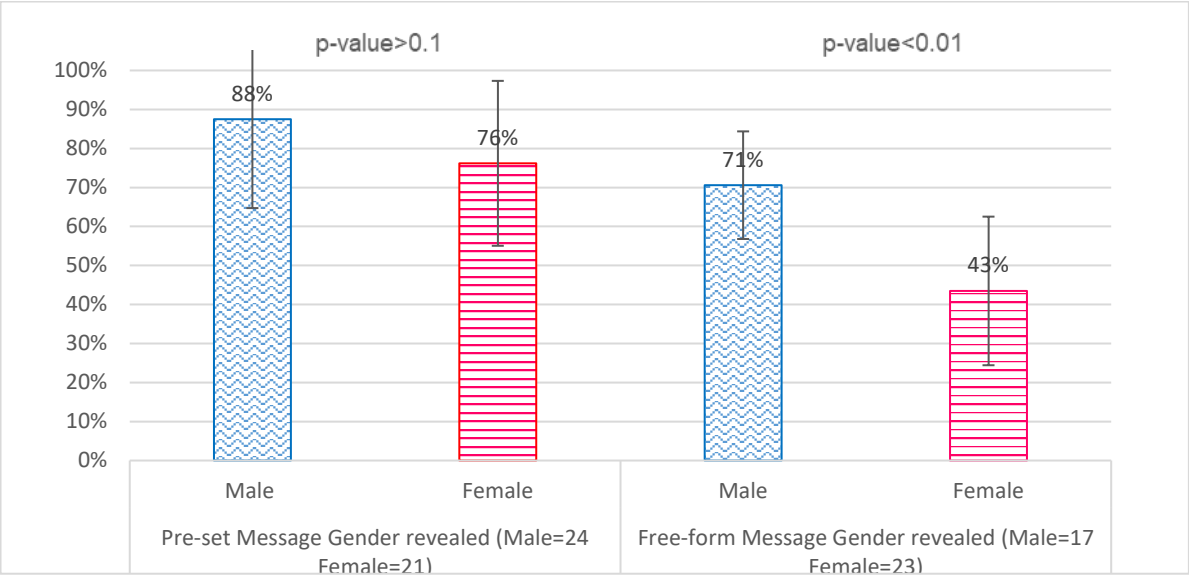
while the difference in the free-form message treatment is significant (|z|=1.86, p<0.1, m=17, f=23).

Figure 3.4A: Difference in willingness to lead broken up by treatments (gender revealed vs. not revealed; pre-set message vs. free-form) in Experiment 1



Note: Pre-set message gender not revealed: 21 males and 24 females; pre-set message gender revealed 40 males and 45 females; free-form message gender not revealed: 18 males and 17 females; pre-set message gender revealed: 87 males and 79 females.

Figure 3.4B: Difference in willingness to lead broken up by treatments (pre-set message vs. free-form) in Experiment 2



Notes: In Experiment 2, there is no gender not revealed treatment since we undertake only gender revealed treatments.

In Table 3.5, we present the results of probit regresions for differences in the willingness to lead. The dependent variable is an individual's response before round 11 (=1 if willing to be leader, =0 otherwise). The first two columns (Models 1A and 1B) present results for Experiment 1, while the third and fourth columns (Models 2A and 2B) do so for Experiment 2. Given that the leadership choice occurs prior to sending messages, in this table we have combined the pre-set and free-form messages for both Experiments 1 and 2. In each case, the first specification (Models 1A and 2A) includes a female dummy (Female= 1 for female, = 0 for male), average effort in the first 10 rounds, average earnings in the first 10 rounds, and demographic controls. The second specification (Models 1B and 2B) control for risk preferences using the decisions made in the Holt-Laury (2002) lotter-choice task.[16] The female dummy is negative and significant at 1% level in Experiment 1, clearly suggesting that fewer women choose to be a leader across both message treatments. For Experiment 2, the female dummy is negative and significant at 5% when we do not control for risk preferences but loses significance once we do so.

[16] As noted previously, we lose oobservations when we control for risk preferences due to excluding subjects who make inconsistent choices in the Holt-Laury task.

Table 3.5: Probit Model for leadership choice prior to round 11

	Experiment 1		**Experiment 2**	
Choice	**Model 1A**	**Model 1B**	**Model 2A**	**Model 2B**
Female	**-0.261*****	**-0.231*****	**-0.223****	-0.120
	(0.046)	**(0.055)**	**(0.089)**	(0.103)
Average effort round 1-10	**0.007*****	**0.011*****	0.009	0.010
	(0.003)	**(0.003)**	(0.006)	(0.006)
Average earnings round 1-10	-0.001	**-0.001****	**-0.002*****	**-0.002****
	(0.001)	**(0.001)**	**(0.001)**	**(0.001)**
Risk Averse	-	0.018	-	0.032
	-	(0.059)	-	(0.105)
Constant	**0.640****	**0.657****	**0.694*****	**0.764*****
	(0.024)	**(0.028)**	**(1.1045)**	**(0.049)**
Demographic Control	Yes	Yes	Yes	Yes
pseudo R^2	0.1127	0.1073	0.157	0.178
Wald χ^2	48.76	33.29	16.35	11.54
Prob>χ^2	0.000	0.000	0.038	0.2407
Number of Observations	330	241	85	59

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

We conclude this section by noting the following results. First, regardless of whether leaders face a different payoff matrix or not, across both our experiments, a much larger proportion of men volunteer to lead. This is borne out by sample proportions tests and parametric probit regressions. This difference in the willingness to lead persists even after we control for risk preferences using the Holt-Laury lottery-choice task. The female reluctance to lead is much more pronounced when participants know that the leaders' gender will be made known to the followers. When the leaders' gender is not revealed, the differences between the proportion of men and women volunteering to lead are usually not significant. But when this information is made public, women exhibit greater reluctance to lead. A plausible conjecture regarding the greater female reluctance to lead is that female leaders anticipate greater resistance and backlash from followers as documented by Gangadharan,

Jain, Maitra, and Vecci (2016). We now turn to exploring to what extent this presumption of greater resistance to female leadership is borne out in the latter half of the game.

3.4.2. Perception and efficacy of leaders

In this part of our chapter, we show that having a leader help with reducing coordination failure. Our aim is to compare the performance of the groups led by male and female leaders respectively. Are there systematic differences in the level of effort exerted, the degree of coordination success or earnings in male and female led groups? Therefore, here, we focus on the gender revealed treatments only. In Experiment 1 (where leader payoff changes), we have 17 groups with 11 male leaders and 6 female leaders in the pre-set message treatment, while we have 33 groups with 19 male and 14 female leaders in the free-form message treatment. In Experiment 2 (where leader payoff remains unchanged), we have 9 groups with 4 male leaders and 5 female leaders in the pre-set message treatment, and 8 groups with 2 male leaders and 6 female leaders in the free-form message treatment. This information is provided in Tables 3.3 and 3.4 above. Given that here the unit of observation is a group of five people, we will be dealing with a relatively small number of observations.

The dynamic nature of the process following the selection of a leader creates potential confounds since differences will arise endogenously. In an ideal setting, we want to hold everything other than leader's gender constant. One way of doing this is to have an exogenously appointed leader and then look at follower responses. This is the approach adopted by Grossman et al. (2019). However, introducing an exogenous leader has an element of artificiality and may well raise questions about the authenticity of such a leader, who has no prior experience with the group or the task at hand. We felt that allowing endogenous emergence of leaders, who already have experience with the coordination problem, makes the setting more realistic. In any event, such endogenous appointment is the logical extension of

the first part of our study, which allows participants to express a willingness to lead. This does create a trade-off between a degree of realism in the design and the need for experimental control. As we explain below, we control for potential confounds as far as practicable and note relevant caveats along the way.

In the pre-set message treatment, we make sure that the message sent by the leaders is exactly the same. Given that this message is exhorting followers to choose the maximum possible effort level (forty hours), it stands to reason that leaders, when sending this message prior to Round 11, will also end up choosing the maximum possible effort level. Therefore, any differences that arise in the pre-set message treatment should be driven primarily by the leader's gender, since we expect no differences in the message sent by the leader and, few, if any, differences in the leader's effort choice in Round 11. Of course, over time, things may start to diverge; male and female leaders may choose different effort levels and also differ in the number of messages they send over the course of the second half of the game. However, we do know that behaviour in these weak-link games is extremely path dependent and so controlling for variations early on guarantees that the initial conditions are similar for the male and female led groups. If the leader's gender matters, then we would expect to see differences emerging in group performance even in this somewhat sanitized environment with tight control over the leader's actions and messages.[17]

On the other hand, in the free-form message treatment, we are allowing for much greater variation. Leaders here can differ along a number of dimensions: the content of their messages; their effort level, especially if the message chosen asks followers to choose an

[17] In the pre-set message treatment, we had one problem. There is one (female) leader in this treatment who chose an effort level of 20 in spite of the fact that she was sending a message asking the followers to choose 40. To analyze the impact of a leader's gender on followers' effort level while holding effort level of 40 in round 11 and the content of the message, we exclude this one group who did not choose 40 in round 11 and her group for the rest of the analysis. This leaves us with 16 groups in pre-set message treatment with 11 male leaders and 5 female leaders in Experiment 1.

effort level different from forty; how many times the message is sent, and how many words to use per message. But, if we find no differences in the pre-set message treatment, then we might be able to conclude that gender per se is not so important.[18] In that case, the free-form treatment allows us to focus on other sources of differences; for instance, whether there are systematic differences in the message or effort choices of the leaders. Further, as we explain below, we will control for message content, frequency and effort choice on the part of the leaders.

We start by showing the evolution of average effort levels in male and female led groups over the course of an entire session. Figures 3.5A and 3.5B are based on data from Experiment 1. They show what happens to effort levels of the course of the session; i.e., Round 1-20 for the pre-set and free-form message treatments respectively. Figures 3.6A and 3.6B provide the same information for the pre-set and free-form message treatments respectively in Experiment 2. In each figure, the left panel shows the evolution of average effort levels over the first 10 rounds of the session, where there is no leader. The middle panel shows the average effort levels chosen by the leaders (male or female; at the expense of being labelled sexist, we have chosen to go with blue for males and pink for females) in Rounds 11-20. The right panel shows the average effort levels for male and female led groups (once again blue for male led groups and pink for female led groups) over the course of Rounds 11-20.

[18] Another explanation is that leaders are treated differently in the pre-set message treatment and the free-form message treatment. Two female leaders choose an effort level less than 40 in round 11. In the free-from message treatment, choosing a payoff other than 40 is likely to send a bad signal to the group members. In the pre-set message treatment, leaders are aware that they will be provided a fixed message as are the other group members. They can only pass on the message provided to them by the experimenter and they only need to decide on whether to pass this information to their group members or not. On the other hand, in the free-form message treatment, leaders know that they will have to write their own messages, they are required to lead the team by using their own messages. Therefore, leaders in the free-form message treatment are more likely to treated as an actual leader than those leaders in the pre-set message treatment. In the pre-set message treatment, the credibility of the message might be attributed to the experimenter and thus independent of the gender of the participant in the leader role.

It is clear that there are no dramatic differences in average effort choices for either leader or followers. This is particularly true for Experiment 2 (with no change in leader payoff) where average leader and follower efforts are very similar. One can discern some differences in Experiment 1 (where leader payoff changes). There is a sharp downward spike in average leader effort in female led groups in the pre-set message treatment (middle panel of Figure 3.5A) but this does not seem to have had much of an adverse impact on average follower effort in those female led groups; if anything, average follower effort in female-led groups is marginally higher than that in male led groups (right panel of Figure 3.5A).

We can also see differences in the free-form message treatment of Experiment 1, where both average leader and follower effort appears to be lower in female led groups. We note two points before proceeding to look into this in greater detail. First, when we look at the average effort level of leaders and/or groups, we are working with a small number of observations. Second, as noted previously, there is path-dependency in such weak-link games, implying that follower effort choices are affected crucially by leader effort choices. So, some of the dynamics here are being driven by the small number of leader observations.[19]

[19] We also look at whether, during the first set of 10 rounds, there are differences in the average effort levels of groups that went on to have a male or female leader. This is to ascertain that there were systematic differences in the performance of these groups during the first part, which may have led to more (or fewer) female participants volunteering to lead. So, effectively, we are asking whether these groups behaved differently prior to a male or female leader being appointed. We do not find evidence of any significant differences and therefore, refrain from elaborating any further on this aspect of the study.

Figure 3.5A: Average effort choice in Rounds 11-20 in pre-set message treatment (Experiment 1)

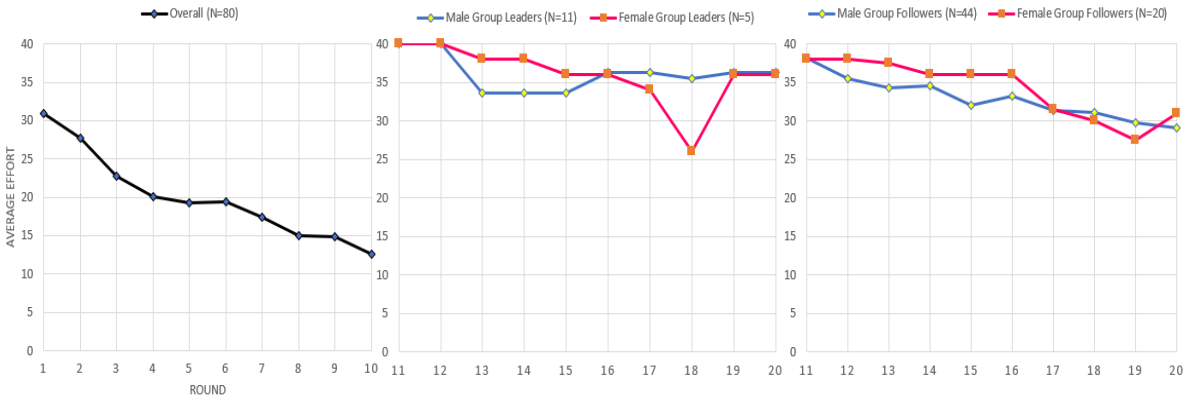


Figure 3.5B: Average effort choice Rounds 11-20 in free-form message treatment (Experiment 1)

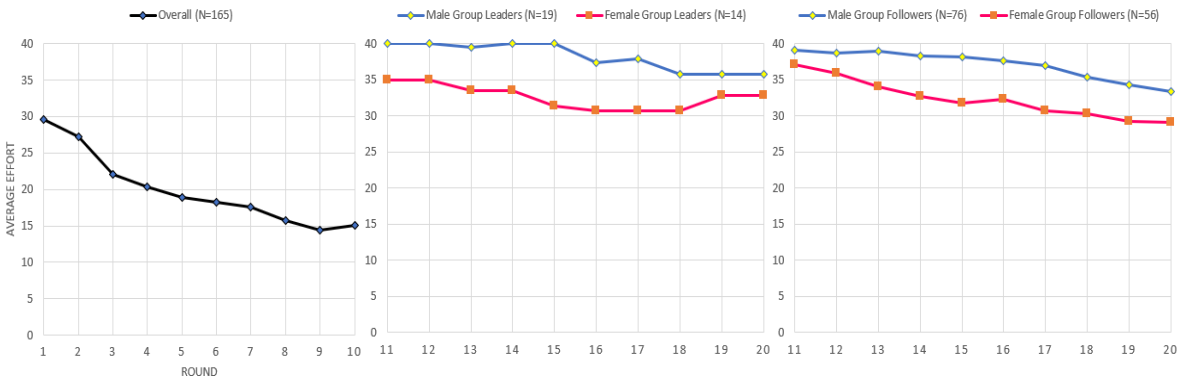


Figure 3.6A: Average effort choice in Rounds 11-20 in pre-set message treatment (Experiment 2)

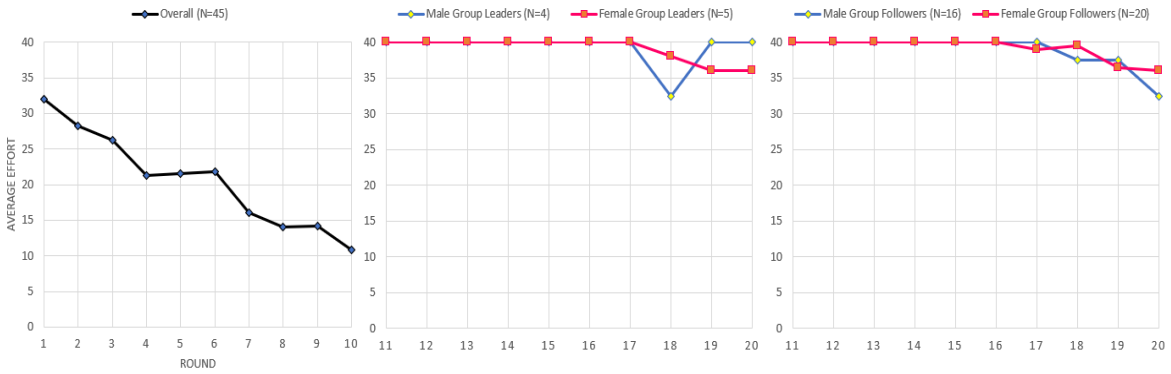
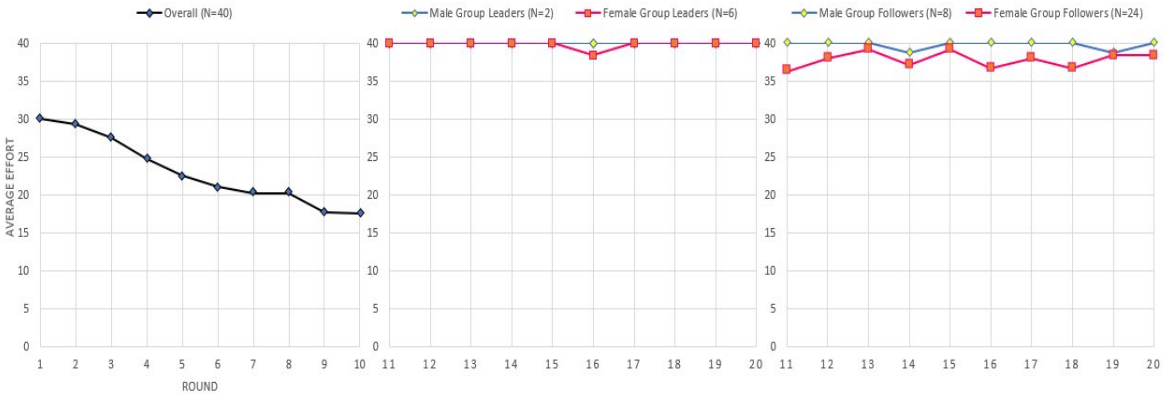


Figure 3.6B: Average effort choice in Rounds 11-20 in free-form message treatment (Experiment 2)



In order to explore any potential gender differences more rigorously, we next turn to regression analysis. In this part, we will look at two issues: first, are there differences in the effort levels of followers in male and female led groups? Second, we will explore whether there are differences in the earnings of male and female led groups. This will help establish to what extent there are differences in the performance of male and female led groups and whether such differences, if any, result is differential earnings for these groups.

3.4.2.1 Follower effort levels in male and female led groups

From Figures 3.5 and 3.6, it seems clear that there are no dramatic differences in average leader /follower effort choices in Experiment 2 (where leader payoff does not change); differences,, if any, arise primarily in Experient 1 (where leader payoff is different). So, in this part of the paper, we will confine our attention to data from Experiment 1 alone. We present detailed analogous results for Experiment 2 in the Appendix 9.

In Table 3.6, we present results of random effects ordered probit regression to examine the impact of leader's gender on follower's effort choice for both treatments with standard

errors clustered on groups.[20] The dependent variable is an individual's effort choice per round from round 11 to 20. Among the regressors, Female = 1, if the follower is female and zero otherwise; Female Leader = 1, if the group leader is female and zero othertwise. The other variables are self-explanatory. For both treatments, the coefficient for round is negative and significant which is in line with what is apparent from Figure 3.5; that there is decay in effort choice over time. The coefficients for earnings in the previous round are postive and statistsically significant suggesting that the higher the earning followers receive, the higher the likelihood of them choosing a higher effort in the following round, but the values of the coefficient are numerically small for both treatments. The coefficients for female leader are not significant for both treatments which suggest that average follower results are not significantly different between male and female led groups. We present analogous results for Experiment 2 in Table 3.A6 (and Table 3.A6*) of the Appendix 9.

[20] Given that there is a natural ordering of the effort choices ranging from zero to 40 in this game, an ordered probit regression is the appropriate way to model choices. Further, given that groups are fixed over time, it makes sense to cluster errors on groups rather than on individual subjects or at the session level.

Table 3.6: Random effects ordered probit model for follower's effort in Experiment 1 (errors clustered on the group)

Effort Choice	Pre-set	Pre-set	Free-form	Free-form
Round	-0.310***	-0.314***	-0.282***	-0.281***
	(0.061)	(0.061)	(0.060)	(0.060)
Female	-2.533*	-2.706**	-1.493	-1.472
	(1.316)	(1.275)	(1.603)	(1.545)
Female*Round	0.147**	0.152***	0.090	0.088
	(0.060)	(0.059)	(0.092)	(0.091)
Female Leader	-0.343	-0.440	-1.047	-1.016
	(0.733)	(0.772)	(0.720)	(0.759)
Lag Earning	0.005***	0.004***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Risk Averse	-0.797	-0.764	1.033***	0.953**
	(0.681)	(0.726)	(0.397)	(0.454)
Demographic Control	NO	YES	NO	YES
Number of observations	470	470	990	990
Wald χ^2	82.29	413.13	66.33	482.20
Prob > χ^2	0.000	0.000	0.000	0.000

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

In Tables 3.7 and 3.8, we take a more disaggregated look at the issue of resistance to female leadership. In Table 3.7, we present the results from random effects ordered probit model for follower effort choices in the pre-set message treatment for Experiment 1. The dependent variable is the effort choice by follower *i* from group *g* in round *t*. The regressors include three dummies: male follower with female leader, female follower with male leader and female follower with female leader. The reference category is male follower with male leader. We also control for the group *g* leader's effort in round *t* (since the followers get to see the leader's effort choice prior to making their own choices), a dummy variable for whether the leader sent a message in round *t* or not and (given the path dependency in such games) the lagged minimum effort, i.e., minimum effort in round *t-1*.

We present results from three different specifications. The first one includes only the regressors without controlling for other demographics; the second one adds demographic controls, while the third and final specification adds both demographic controls as well controls for risk preference derived from the Holt-Laury lottery choice task. (As noted earlier, we lose observations when we control for risk preferences due to inconsistent choices made by a number of participants, who get dropped.) Follower effort levels increase in leaders's effort and the lagged minimum effort, both of which facts make intuitive sense. The results do not provide evidence of any backlash against female leadership. The main thing that stands out is that female followers exert lower effort than their male counter-parts and this is true both when the leader is female as well as when the leader is male. The relevant coefficients are significant at 5% in the second and third specifications. However, male followers do not behave differently regardless of whether the team leader is male or female. The results for Experiment 2 are similar and are reported in Table 3.A7 of the Appendix 9.

Table 3.7: Random Effects Ordered Probit model for effort level under pre-set message treatment in Experiment 1 (errors clustered on the group)

	Model 1	Model 2	Model 3
Round	-0.238***	-0.240***	-0.287***
	(0.041)	(0.042)	(0.042)
Female*Round	0.089*	0.092*	0.155***
	(0.048)	(0.049)	(0.054)
Male follower with female leader	0.298	0.264	-0.137
	(0.580)	(0.633)	(0.704)
Female follower with male leader	-1.423	-1.610	-2.658**
	(1.051)	(1.092)	(1.091)
Female follower with female leader	-1.802*	-1.951*	-2.864**
	(1.051)	(1.183)	(1.161)
Leaders effort	0.057***	0.057***	0.063***
	(0.020)	(0.020)	(0.024)
Message shown	-0.097	-0.099	-0.139
	(0.156)	(0.154)	(0.251)
Lag minimum effort	0.036***	0.035***	0.040***
	(0.008)	(0.007)	(0.007)
Risk Averse	-	-	-0.470
	-	-	(0.492)
Constant 1	-3.619***	-3.800***	-4.546***
	(1.010)	(1.071)	(0.999)
Constant 2	-3.562***	-3.742***	-4.464***
	(0.993)	(1.060)	(0.970)
Constant 3	-2.736**	-2.913**	-4.006***
	(1.146)	(1.226)	(1.039)
Constant 4	-1.463	-1.639	-2.719**
	(1.101)	(1.188)	(1.166)
Log pseudolikelihood	-416.522	-415.097	-250.541
Demographic Controls	NO	YES	YES
Wald χ^2	124.83	5313.37	2018.53
Prob > χ^2	0.000	0.000	0.000
Number of observations	640	640	470
Wald test for equality of coefficients			
Male follower with female leader = Female follower with male leader	χ^2=2.85 p>χ^2 =0.09	χ^2=3.33 p>χ^2 =0.07	χ^2=4.81 p> χ^2 =0.03
Male follower with female leader = Female follower with female leader	χ^2=5.01 p>χ^2 =0.03	χ^2=5.17 p>χ^2 =0.02	χ^2=5.94 p>χ^2 =0.01
Female follower with male leader = Female follower with female leader	χ^2=0.47 p>χ^2 =0.49	χ^2=0.30 p>χ^2 =0.58	χ^2=0.09 p>χ^2 =0.76

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

In Table 3.8, we carry out a similar exercise for the free-form message treatment in Experiment 1 (where leader payoff changes). The regressors are similar to those in Table 3.7 except, here, we also control for the style of message sent by the leader. We classify messages into three different types: authoritarian, democratic and laissez-faire. We provide details on this classification exercise in Section 3.4.3 below. We provide four different specifications here: the first one includes only the main regressors of interest; the second one adds demographic controls, the third adds dummies for the message style (authoritarian/democratic/laissef-faire) and the fourth specification also controls for risk preference.

The results are similar and once again we do not see any evidence of resistance against female leadership. Female followers exert lower effort levels in general but this is true in both male and female led groups. The only other fact that stands out (in the fourth specification) is that democratic messages from leaders seem to have a positive and significant impact on follower effort choice. We conclude this section by arguing that there is no systematic evidence of backlash against female leaders. Female followers choose lower effort levels in general but this is true for both male led groups. There are no differences in male follower effort levels regardless of whether the leader is male or female.[21]

[21] The results for the same exercise for Experiment 2 are reported in Table 3.A8 of the Appendix 9.

Table 3.8: Random Effects Ordered Probit model for effort level under Free-form message treatment in Experiment 1 (errors clustered on the group)

Follower effort	Model 1	Model 2	Model 3	Model 4
Round	-0.169***	-0.165***	-0.165***	-0.181***
	(0.046)	(0.045)	(0.045)	(0.036)
Female*Round	0.058	0.053	0.055	0.025
	(0.061)	(0.061)	(0.061)	(0.081)
Male follower with female leader	-0.485	-0.480	-0.638	-0.696
	(0.469)	(0.465)	(0.485)	(0.453)
Female follower with male leader	-0.915	-0.703	-0.835	-0.736
	(1.040)	(1.017)	(1.027)	(1.360)
Female follower with female leader	-1.219	-0.981	-1.180	-0.920
	(0.992)	(0.963)	(1.006)	(1.235)
Leaders Effort	0.032***	0.033***	0.032***	0.037***
	(0.009)	(0.009)	(0.009)	(0.009)
Authoritarian Message Style	-	-	0.576**	0.536
	-	-	(0.261)	(0.346)
Democratic Message Style	-	-	0.526**	0.576**
	-	-	(0.223)	(0.243)
Laissez-Faire Message Style	-	-	0.694	0.745
	-	-	(0.453)	(0.467)
Lag minimum effort	0.036***	0.036***	0.036***	0.041***
	(0.008)	(0.009)	(0.008)	(0.010)
Risk averse	-	-	-	0.645**
	-	-	-	(0.306)
Constant 1	-3.642***	-3.726***	-3.570***	-3.365***
	(0.831)	(0.917)	(0.927)	(1.071)
Constant 2	-3.422***	-3.505***	-3.346***	-3.104***
	(0.842)	(0.934)	(0.946)	(1.087)
Constant 3	-2.857***	-2.941***	-2.774***	-2.413**
	(0.913)	(0.988)	(0.995)	(1.199)
Constant 4	-2.421***	-2.508***	-2.337**	-1.949
	(0.892)	(0.966)	(0.971)	(1.188)
Log pseudolikelihood	-551.775	-546.307	-542.730	-397.170
Demographic Controls	NO	YES	YES	YES
Wald χ^2	94.78	645.24	726.72	1181.63
Prob > χ^2	0.000	0.000	0.000	0.000
Number of observations	1320	1320	1320	990
Wald test for equality of coefficients				
Male follower with female leader = Female follower with male leader	χ^2=0.13 p=0.71	χ^2=0.04 p=0.84	χ^2=0.03 p=0.86	χ^2=0.00 p=0.98
Male follower with female leader = Female follower with female leader	χ^2=0.47 p=0.49	χ^2=0.26 p=0.61	χ^2=0.32 p=0.57	χ^2=0.03 p=0.86
Female follower with male leader = Female follower with female leader	χ^2=0.44 p=0.51	χ^2=0.41 p=0.52	χ^2=0.68 p=0.41	χ^2=0.13 p=0.72
Authoritarian Message Style = Democratic Message Style	-	-	χ^2=0.04 p=0.85	χ^2=0.02 p=0.90

3.4.2.2 Earnings in male and female groups

Figures 3.5 and 3.6 as well as the regression results presented in Tables 3.8 and 3.9 suggest that there are no systematic differences in the performance of male and female led groups in terms of average effort. This, in turn, implies that we would not expect to see differences in the average earnings of male and female led groups as well. We present results for earnings from Experiment 1, in Table 3.9. As noted above, we expect differences, if any, to arise in Experiment 1 rather than Experiment 2, since the latter shows no difference in average effort levels at all. For the sake of completeness, we also provide the earnings regression for Experiment 2 in the Appendix 9. The coefficient for the female leader dummy is not significant in any of the treatments, whether pre-set message or free-form; neither is it significant if we control for risk preference or not. We conclude that there is no evidence to suggest that male and female led groups differed in terms of their average earnings.[22]

[22] The results for the same exercise for Experiment 2 are reported in Table 3.A9 of the Appendix 9.

Table 3.9: Random Effects Model for follower's round earnings in Experiment 1 (errors clustered on the group)

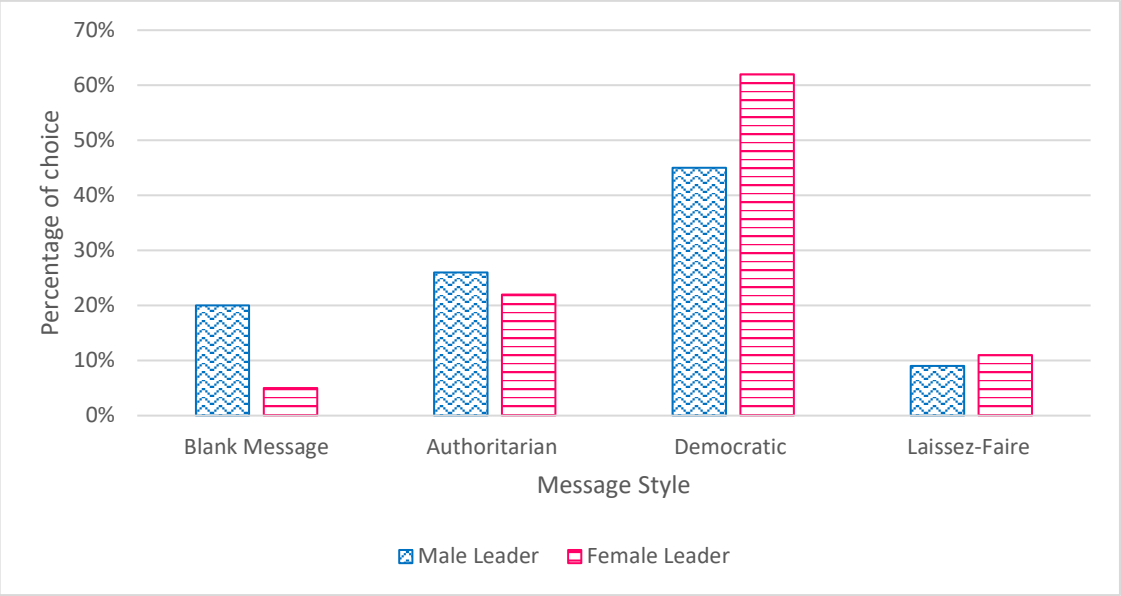
Earnings	Pre-set	Pre-set	Free-form	Free-form
Round	**-11.136*****	**-8.767*****	**-8.441*****	**-8.184*****
	(3.421)	**(2.628)**	**(2.607)**	**(2.821)**
Female	**-61.546***	-2.644	-22.629	-29.332
	(33.223)	(36.178)	(32.783)	(37.469)
Female Leader	-12.469	-16.237	-9.852	-11.777
	(13.366)	(15.259)	(14.252)	(14.189)
Female*Round	**3.995***	0.087	1.627	2.123
	(2.385)	(2.506)	(2.097)	(2.407)
Leader's Effort	1.205	1.195	**2.154*****	**1.539****
	(0.934)	(0.975)	**(0.710)**	**(0.727)**
Lag Minimum Effort	**3.768*****	**3.685*****	**3.230*****	**3.484*****
	(0.427)	**(0.509)**	**(0.623)**	**(0.561)**
Risk Averse	-	0.405	-	7.590
	-	(8.303)	-	(6.804)
Constant	**551.393*****	**521.925*****	**497.061*****	**504.101*****
	(46.791)	**(34.761)**	**(47.028)**	**(51.153)**
Wald χ^2	321.314	707.255	251.730	511.960
Prob > χ^2	0.000	0.000	0.000	0.000
Number of observations	640	470	1320	990

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

3.4.3. Leadership style and message content in free-form message treatments

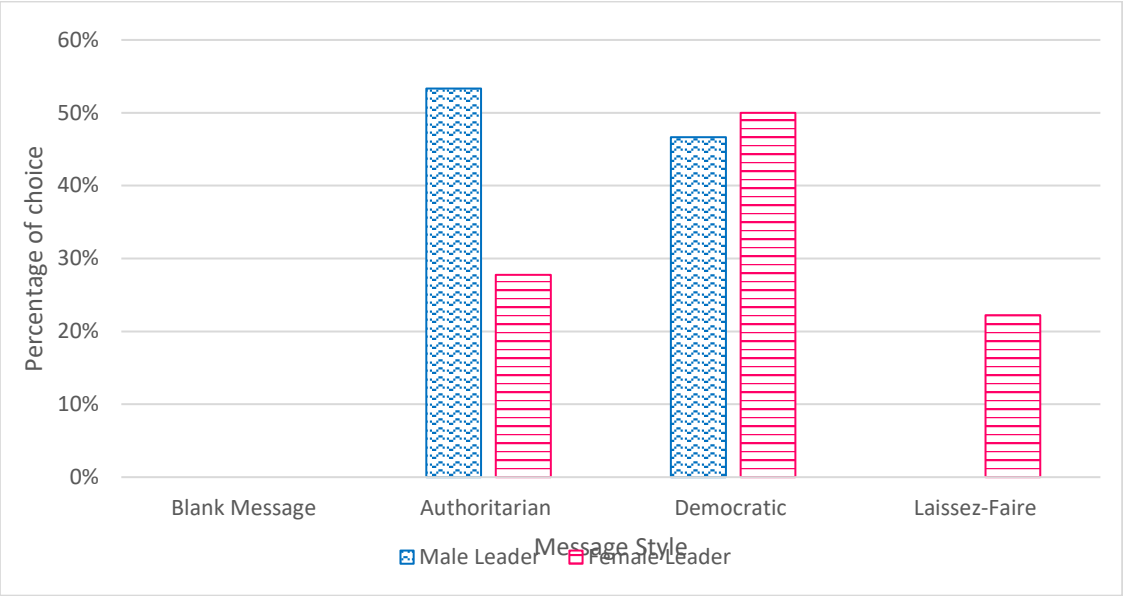
We rely on the Northouse (2012) methodology to identify the types of messages sent by the leaders. Bear in mind that the leaders have two choices for each round following Round 11; whether to send a message or not and if yes, then what type of message to send. We classify messages into four categories. "0" indicates that leaders did not send any message to their group members. "1" stands for an "authoritarian" message style, "2" for democratic and "3" for laissez faire.[23] Figure 3.7, Panels A and B show the distribution of message choices in Experiments 1 and 2. By and large there are no substantive differences here between the two genders; at most men showed a slight preference for authoritarian messages while women preferred democratic messages. But we fail to find any significant differences.

Figure 3.7A: Distribution of messages in Experiment 1



[23] An "authoritarian" message is one where leaders give orders and/or demand compliance from group members. Examples include: "*I think I've made myself clear. All choose 40 and we make the most money!*", and "*All pick 40 hours for the remaining ten rounds for maximum pay*". Style 2 is defined as "democratic", where leaders exhort followers and make suggestions/requests. Following are examples: "*Hey guys, to make the most money let's all select 40 hours. That way you and I will both earn the most profits, we all get the optimal outcome ;)",* and *"Hi guys! If you choose more than 20 that would be great!".* We categorize Style 3 as laissez faire message style where messages have no substantive content. E.g., "*You know what to do. ;) I believe in you.*", and "*McDonalds tonight!*". We use three independent coders to undertake this coding. A participant is assigned to a particular category as long as two out of the three coders assigned the same score to this participant.

Figure 3.7B: Distribution of messages in Experiment 2



In Table 3.10, we show the relative frequencies of messages sent by leaders. While there are differences here, there do not seem to be any obvious patterns. It is also the case that these are repeated observations for the same leaders rather than independent observations and for Experiment 2, we have very few observations. If we look at the average number of words in the messages sent then women, on average, wrote 9.1 words (standard error of 3.3) while men wrote 9.6 words (standard error of 4.39). Once again there do not seem to be any significant differences in the style or frequency of messages sent by our leaders.

Table 3.10: Frequency of messages sent by leaders

		Pre-set message treatment	Free-form message treatment
	Female leader	0.73 (f=10)	0.95 (f=12)
Experiment 1	Male leader	0.69 (m=16)	0.79 (m=13)
Sample proportions test		*z=0.55 p=0.58*	*z=-3.68 p<0.01*
Experiment 2	Female leader	0.98 (f=5)	0.78 (f=6)
	Male leader	0.78 (m=4)	1 (m=2)

Notes: Leaders have to send a message in Round 11 but from that point onward, leaders can choose whether to send a message or not. If the leader chooses to send a message in all 10 rounds, then this is denoted as 100%. This is normalized to 1 in the table. We average over all leaders in a particular treatment. In the case of Experiment 1, within the pre-set and free-form message treatments, we combine the gender revealed and gender not revealed treatments. E.g. 0.73 implies that in that particular treatment, female leaders sent an average of 7.3 messages out of the maximum of 10 messages possible. Given the very small number of observations in Experiment 2, we have dispensed with formal statistical tests.

3.5. Conclusion

In this study, we examine whether there is a gender difference in the willingness to lead in a weak-link game designed to simulate frequently occurring intra-organization coordination problems. Overall, we find that fewer women volunteer to be the leader compared to men. This decision depends crucially on whether the leader's gender is revealed to group members or not. When the leader's gender is revealed to followers, fewer women choose to volunteer; when it is not, the willingness gap among men and women shrinks.

We do not find significant differences in the degree of coordination success (or failure) between male and female led groups. Neither do we find much difference between effort levels chosen or the content of messages sent by male and female leaders. While we do not find significant differences in followers' perception towards male and female leaders, we do find that women are less likely volunteer for leadership positions. As noted above, a plausible conjecture is that female participants shied away from volunteering because they anticipated resistance from followers. But this resistance did not materialize in our study.

4. Gender and Employment Relations

4.1. Introduction

The well-documented "gender gap" in the workplace has been, and remains, the subject of extensive research. As we have discussed in the previous chapter, this gender gap takes in two forms. The OECD defines the gender wage gap as the difference between median earnings for males and females relative to the median earnings for males. Looking at data for 2015-2018, we find that the average gender wage gap for all OECD countries is 13.5%, however there are substantial cross-country variations. At the most unequal end we have Korea (with a difference of 35%), followed by Japan (25%), Israel (22%) and then Canada, USA and Finland (around 18%). At the other extreme, we have Belgium, Greece, Costa Rica, Denmark and Italy, all hovering around 5% with Ireland, Norway and Sweden at about 6% and New Zealand a little less than 8%.

Alongside this difference in earnings, there is also a gender gap in leadership roles: as one moves up the hierarchy of organizations, one finds fewer women. Sandberg (2013) notes, that, at the time of her writing, of the 195 independent countries in the world, only 17 were led by women. Women held just 20% of seats in parliaments globally and about 14% of executive officer positions, 17% of board seats, and constituted 18% of elected congressional officials in the US.

Early research in the area, such as Altonji and Blank (1999), attribute this phenomenon mainly to differences in human capital (including the motherhood penalty and breaks in work experience), occupational choice as well as taste-based and/or statistical discrimination. However, in recent years, a large body of experimental economics research, surveyed comprehensively in Bertrand (2011), suggests that observed gaps may also arise as a result of gender differences in psychological attributes and preferences between men and women. Such differences may include gender differences in risk preferences, in attitudes towards

competition and negotiation and in other-regarding preferences. This in turn may also have implications for gender differences in occupational choice or work-place strategies.

Babcock and Laschever (2003) and Niederle and Vesterlund (2007) argue that greater female aversion to competition may explain why one finds fewer women occupying positions of power. Bohnet (2016) provides an excellent overview of the issues involved and suggests a series of institutional changes in order to achieve parity between the genders. Our study is intended to add to this experimental economics literature by looking at whether men and women have different preferences and adopt different strategies in a game designed to simulate worker-employer relationships.

This research agenda has obvious overlaps with research in leadership; specifically whether men and women tend to adopt different styles when it comes to dealing with employees. Rosener (1990), building on concepts introduced by Burns (1978), argues that men typically tend to be "*transactional*" leaders and see job performance as a series of transactions with subordinates involving rewards for services rendered and punishments for inadequate performance. Women on the other hand are seen as being more "*transformational*", relying less on explicit rewards and punishments and more on a democratic and participative style.

Eagley and Johnson (1990) undertake a meta-analysis of 162 studies on leadership and find little difference between male and female leadership styles. They find some support for the view that women adopt a more democratic style while men tend to adopt a more authoritative style. They suggest that these differences may arise in part from the fact that women, being out-numbered by men as leaders, face greater resistance from employees and feel the need to seek greater employee input. Eagly, Karau, and Makhijani (1995) undertake a further meta-analysis which extends the analysis of leadership styles to the issue of leadership effectiveness. They report that men and women are equally effective as leaders except that men tend to be more effective in occupations that are typically defined in more

masculine terms, such as the military, while women are more effective in occupations defined in primarily feminine terms such as nursing.[24]

In this study, we intend to contribute to the study of this issue of *transformational* as opposed to *transactional* leadership using the well-known gift-exchange game paradigm. (Fehr, Gächter, and Kirchsteiger, 1997; Fehr, Kirchler, Weichbold, and Gächter, 1998; Fehr, Kirchsteiger, and Riedl, 1993, 1998, 1996; Fehr, Klein, and Schmidt, 2007.) We report results from experiments simulating an employer-employee relationship, where the interactions are mediated by two types of labour contracts. The first contract relies on mutual trust and reciprocity between the two which we refer to as a *"relational"* contract. This is designed to serve as a proxy for a *transformational* leadership style. A second contract relies on explicit penalties for the worker if found to be shirking. This is designed to resemble a more *transactional* type of leadership. [25]

In doing so, we complement results reports in Chaudhuri et al. (2015), who also looked at trust based and penalty-based contracts. However, in their paper trust was one-sided in that only the employer has the option of reposing trust on the worker and the worker can reciprocate that trust or not. Among other things, they found that when it came to penalty-based contracts, women employers tended to mitigate the "stick" of the penalty with the

[24] See Moran (1992) for a succinct overview of much of this work.

[25] Laboratory gift exchange experiments are admittedly highly stylized. But the issue of relational versus transactional approaches in the work-place is complex and field studies are subject to multiple confounding factors. This often makes it hard to distinguish which features are universal and which are inextricably tied to a particular corporate culture. As a result, Camerer (2003) and Charness and Kuhn (2011) argue that, while stylized, lab studies can provide a reasonable and tractable model of labour market interactions. The particular experiments in this paper have the added advantage that they mitigate some of the artificiality of lab experiments and provide greater context by using terms such as employer, worker, wage, effort and fines. A final caveat here is that laboratory studies typically report greater gender differences in behaviour than are found in actual field studies. Eagly and Johnson (1990) point out that lab studies where (1) people deal with strangers, typically for one-off or short-lived interactions and (2) there are fewer social cues on which to anchor behaviour tend to exacerbate gender differences. We have noted the difficulties posed by field studies. One way to think of lab studies is that they may help identify areas where such differences may arise and also establish bounds on how significant those differences may turn out to be.

"carrot" of higher rents toward workers; a strategy that led to lower earnings for the women employers in those contracts. However, in that study, the assignment to a institution was exogenous and the employers had no choice in this matter. Given the greater generosity on the part of our female employers in the context of penalty contracts, a plausible conjecture is that female employers were not comfortable with the transactional contracts and therefore, felt the need to concede greater rent to make up for imposing penalties for non-compliance.

Consequently, we make two modifications to Chaudhuri et al. (2015). First, given that most studies find women to be more reciprocal than men (see, for instance, Eckel and Grossman, 2008), we modify the relational contract to allow for *two-sided trust and reciprocity* as opposed to only reciprocity from the worker to the employer; That is, we extend the contract to allow for *ex post* reciprocity from the employer to the employee. Second, we explicitly allow the employers to choose the type of institution they wish to implement. In other words, the employer is no longer exogenously assigned to one of the two types of institutions but can actually choose over multiple rounds whether they wish to choose a relational contract (based on mutual trust) or a transactional contract (based on punishment for shirking workers).

From a purely self-interested perspective the trust-based contract should result in minimal effort and output while with an optimal fine, the penalty-based contract should elicit higher effort and output than the trust-based contract. The entire point of Fehr and his colleague's voluminous work in this area is to show that contrary to this theoretical proposition, trust-based contracts often generate a greater surplus compared to penalty-based contracts. If you are choosing a relational contract, we believe it is a more democratic way while choosing the penalty-based contract is a more transactional way, in the sense that the primary assumption behind the latter is that employees will only choose higher effort if there is a penalty.

Our *ex ante* conjecture is that women will show a preference for the relational contract. However, we do not find evidence in favour of this conjecture. Both male and female employers show an overwhelming preference for the trust and reciprocity contracts over the penalty-based contract. We do replicate the Chaudhuri et al. (2015) result that women employers tend to be more generous and concede higher rent to the workers than they have to. This tends to lower female employer earnings. Our results echo the Chaudhuri et al. (2015) results in the sense that most of our findings are null findings. We do not find evidence of strong gender differences across a range of variables. Our results corroborate and replicate the earlier results of the lack of significant gender differences. These null findings are still of interest since they suggest that the differential outcomes for men and women in the workplace need not be attributed to differences in preferences. By and large, men and women do not seem to differ much in their choice of strategies. This potentially suggests a larger role for discrimination and possibly calls for a more pro-active stance on implementing institutional design changes along the lines suggested by Bohnet (2016) in order to achieve parity between the gender.

We proceed as follows. In Section 4.2 we provide an overview of the experimental design and procedures. In Section 4.3 we present our results. We make some concluding remarks in Section 4.4.

4.2. Experimental Design and Procedures

4.2.1. Experimental Design

We apply a modified version of the principal-agent model from Fehr et al. (1997, 2007). The employer offers a wage $w \in [0.01, \dots, 10]$, and the worker exerts an integer effort level ranging from one to ten, $e \in [1, \dots, 10]$, which generates output V. Output V is a function of

worker effort e, i.e. $V = V(e)$. The value of effort monotonically increases in effort level such as $V'(e) = 1$, $V''(e) = 0$. Effort generates value for the employer but is costly to the worker where the cost is measured in monetary terms, $C = C(e)$ with $C'(e) > 0, C''(e) > 0$. We normalize the price of output to 1 so that revenue is equivalent to output, and worker's outside option is normalized to zero. Table 4.1 provides detailed parameters for the relationship between value of effort and cost of effort for each effort level. All payoffs denoted in the table represent actual monetary payoffs that participants will earn in our experiment. All payoffs are in New Zealand dollars.

Table 4.1: Output and Cost of Effort ($)

Worker Effort Level:	1	2	3	4	5	6	7	8	9	10
Employer Value of Effort:	$1.00	$2.00	$3.00	$4.00	$5.00	$6.00	$7.00	$8.00	$9.00	$10.00
Worker Cost of Effort	$0.01	$0.10	$0.20	$0.40	$0.60	$0.80	$1.00	$1.30	$1.60	$2.00

Employers and workers interact for 10 rounds. We implement two types of matching protocol: fixed matching and random re-matching between rounds. At the beginning when participants login, they are randomly assigned to the role of either an employer or a worker. These roles remain unchanged for all rounds. In the fixed matching treatment, participants are matched with the same person for the entire session of the experiment. In the random re-matching treatment, participants are randomly re-matched with a different person each round.

A fixed matching protocol simulates long term relations between a pair; this allows for signalling future intentions and reputation building. Random re-matching, on the other hand, is designed to simulate a series of one-off (or short-term) interactions that often characterize some industries such as fast-food and retail with high rates of employee turn-over. In each round, employers move first by making a contract offer. They can choose a

relational contract or a transactional contract. We expect that relational contracts will do better in the fixed-matching treatment, where there is potential to build on reputation, while it is likely that transactional contracts will be required more in the random re-matching treatment. We explain these contracts in the next section.

4.2.2. Relational contracts

This is a three-stage game if the employer chooses to offer a relational contract. In stage 1, the employer makes a contract offer to the worker and the worker chooses to accept or reject the contract. The contract contains a wage rate w, a suggested effort level e^* and a proposed bonus amount b^* from the employer. The proposed bonus is a payment between \$0 and \$10 to the worker that is specified in stage 1. The actual payment, however, is selected after the worker's effort decision is revealed and the employer can observe whether or not the suggested effort level is met. The suggested effort e^* and the proposed bonus b^* are non-binding; the workers are free to choose any level of effort and the employer can pay any amount of bonus, which can be different from the proposed bonus. The interaction here then relies on mutual trust and reciprocity between employers and workers.

In Stage 2, the worker chooses to accept or reject the contract. If the worker rejects the contract, then both employer and worker earn zero in that round. On the other hand, if the worker accepts the contract, then the worker decides an actual effort level to put in. At stage 3, the employer observes the worker's actual effort level and then choose an actual bonus amount $b \in [0, \ldots, 10]$ to pay the worker. The actual bonus payment may be any amount between \$0.00 and \$10, and may be greater than, equal to, or less than what the employer had proposed earlier in Stage 1.

Any time a contract is accepted the employer's payoff is given by the value of the output minus the sum of the wage and the bonus (if any). The worker's payoff is the wage plus the bonus (if any) minus the cost of effort. If we assume that both the employer and the worker are self-interested and wish to maximize monetary returns, then we would expect that *ex-post*, the employer has no incentive to pay any positive amount of bonus. Consequently, if we also assume that the worker knows that the employer is self-interested, the worker has no incentive to provide non-minimal effort. Anticipating this, and assuming that the employer both knows that the worker is self-interested and knows that the employer is self interested, the employer should offer the lowest possible wage. Therefore, the employer should offer a wage rate of \$0.01, the worker should choose the lowest effort level of 1, and employer should pay zero bonus. The corresponding earning for the employer would be $\pi = V(e) - w - b = \$1 - \$0.01 - \$0 = \$0.99$, and the payoff for the worker would be $u = w - C(e) - b = \$0.01 - \$0.01 - \$0 = \$0$.

On the other hand, if the employer offers a positive rent by paying the worker a premium over the cost of effort, defined by $r = w - C(e)$ and the worker reciprocates this positive rent by choosing non-minimal effort, then both the employer and the worker are better off than in the self-interested equilibrium of the game as described above. We define rent in this way because the employer only needs to offer a wage that is equal to, or slightly more than, the cost of effort so that the worker is willing to accept the contract given that the worker's outside option has been normalized to zero. Any amount more than that suggests that the employer is conceding rent (in the form of higher pay than absolutely necessary) to the worker. We use rent offered to measure the generosity of the employer's contract offer in our subsequent analysis.

4.2.3. Transactional contracts

This contract allows the employer to impose a fine to penalize workers if they are shirking. The worker is shirking if the actual effort level is less than the suggested effort level, i.e. $e < e*$. The employer must invest in monitoring technology in order to detect shirking workers. By investing in this monitoring technology, the employer incurs a fixed cost k, and can impose a fine on the worker. The fine is paid to the employer only if shirking is verified. Monitoring is not perfect and in this study the probability of being discovered while shirking is 0.33.

This interaction proceeds in three stages. In stage 1, the employer offers a wage w and chooses a suggested effort level e^*. The employer also decides whether to invest in monitoring technology at a cost of k and the level of fine f to impose on the worker. The cost of implementing the monitoring technology is fixed and equal to \$1. The level of fine ranges from 0 to 1.3 in one decimal place increments, i.e. $f \in [0, \dots, 1.3]$. The fine can only be collected only if shirking is detected, which happens with a 0.33 probability.

In the second stage, the worker chooses to accept or reject the contract after observing the employer's choices in stage 1. This includes not only the wage rate and suggested effort level, but also whether the employer invested in a monitoring technology or not, and the level of fine the employer chose to impose. If the worker accepts the contract, then the worker chooses an effort level e. If the contract is rejected by the worker, then both employer and worker earn zero.

The third stage is brief. In this stage, the verification process takes place; the monitoring technology verifies, with a probability of 0.33, whether the actual effort is below the suggested effort level. It is easy to show that, in this setting, payoff for the employer is maximized by investing in the monitoring technology, imposing the maximal fine of 1.3 units

and asking for an effort level of 4 units. The employer will offer a wage to the worker that just about compensates the worker for the cost of this effort, which is $0.4, or a little bit more. In this case, the employer will earn $2.6 and worker will earn zero. This follows from Fehr et al. (1997, 2007). We also provide a simple derivation in the Appendix 6. It is also clear that this outcome is better than the purely self-interested outcome in relational contracts. Self-interested motivations suggest that relational contracts should elicit only the smallest possible effort levels and therefore the employer should be monetarily better off with transactional (penalty-based) contracts.

4.2.4. Task and Questionnaire

Following the experiment, participants are asked to take part in the Holt and Laury (2002) lottery choice experiment. This is shown in the Appendix 4. We use this task to elicit risk preferences, which is one of the co-variates in further analysis. The switching point from option A to option B is used to measure each individual's risk preference. E.g., a risk-neutral person would choose option A for the first four rows and then switch to option B at the fifth row. A risk-loving individual would switch to option B at the fourth row or before while an individual switching to option B at the sixth row or later would be categorized as risk-averse. Participants are informed that this is a separate task. At the end of the session, the computer will randomly choose a row and will pay for either Option A or Option B, depending on what the participant chose for that specific row.

Finally, participants are asked to fill out a demographic questionnaire (See Appendix 8) used to collect information regarding participant's gender, field of study, year in the undergraduate program, age, income, whether they were born in New Zealand, and their

ethnicity. This demographic questionnaire is similar to the one Statistics New Zealand uses to collect Census data.

4.2.5. Experimental Procedures

A total of 142 participants took part in this experiment. Table 4.2 provides summary statistics for our experiment. All sessions in this experiment were conducted in the DECIDE laboratory at the University of Auckland using Veconlab[26], an online software developed by Charles Holt at the University of Virginia. Participants were recruited via an email announcement and they were students from undergraduate courses without any prior experience with the principal-agent game. We have 33 pairs in the Fixed Matching treatment and 38 pairs in the Random Matching treatment.

Table 4.2: Number of participants in different treatments

	Fixed Matching	Random Matching
Employers	N=33	N=38
	Male=19	Male=22
	Female=14	Female=16
Employees	N=33	N=38
	Male=13	Male=20
	Female=20	Female=18
Total	N=66	N=76
	Male=32	Male=42
	Female=34	Female=34

Participants are directed to the computer cubicles once they enter the lab. There are dividers between each cubicle so that each participant is separated from another and is unable to see any other participant's computer screen. This prevents any opportunities to observe the

[26] http://veconlab.econ.virginia.edu/admin.htm

decisions made by any other participants during the experiment. They are also cautioned against communicating with others. At the beginning of the experiment, the experimenter distributes the game instructions (included in the Appendix 2) to participants and reads it out loud. The instructions include a summary of the game with some numerical examples and practice questions regarding payoffs from relational and transactional contracts. Participants are given 10 to 15 minutes to read the instructions and work on the numerical practice questions. All participants' answers are checked and any questions participants have are answered. Participants are told that the experiment consists of 2 parts: first, there are 4 practice rounds to help them get familiar with the game. This is followed by 10 actual rounds, where they will be earning money. Once the actual game starts, the computer program will keep track of their total earnings for all ten rounds, and these will be shown as "cumulative earnings" on a results page. The earnings from the practice rounds are also shown on the screen but participants know that this will not be added to their total earnings.

At the end of each round, participants can observe their individual earnings for that round as well as their cumulative earnings. However, they do not have any information about the earnings of any other players, including the players they are paired with in any round. However, they can use the parameters given in the instruction to calculate their pair member's earning, should they wish to do so.

Participants would know that they are randomly assigned the role of either an employer or a worker when they log in to the Veconlab website. This role will remain unchanged for the entire duration of the experiment, that is, for the 4 practice rounds and 10 actual rounds. Each participant is also assigned a subject identification number (ID), and they are unable to learn the identity of other participants. At the completion of the 10 rounds of the game, participants are asked to take part in the Holt-Laury lottery choice game and fill out the demographic questionnaire.

Each session lasts approximately 90 minutes. At the end of the session, participants are paid their earnings in cash from the principal-agent game and from the lottery choice task, plus a show-up fee. Participants are told that their earnings are private information and they were free to leave after collecting their payment.[27] Average earnings for the 10 rounds are $23 for employers and $26 for workers, not including payment from the lottery choice experiment and the $5 show-up fee.

4.2.6. Hypotheses

Based on our review of prior findings in the literature, we propose the following hypotheses.

Hypothesis E1: Female employers are more likely to choose a higher proportion of trust-based relational contracts as opposed to punishment-based transactional contracts.

Hypothesis E2: Given existing evidence in favour of greater female generosity, female employers will concede higher rent to the workers; this will result in lower earnings for female employers.

---

[27] A common difficulty with this game is that employers often end up with negative earnings. This is true for many of the studies we have cited including the one most closely related, Chaudhuri et al. (2015). Given that taking money away from subjects is not an option, experimenters typically restrict earnings to zero in the event of negative earnings. But this limited liability creates potential confounds because once in negative territory, subjects may not pay adequate attention. There is no longer any incentive to make careful decisions at this point since they cannot earn anything less than zero. This leads to a loss of reward salience. To get around this (and any other loss aversion type arguments), we provide an additional $10 dollars to the employer without providing any further information regarding the reason for this additional endowment. The relevant paragraph in our instruction states:

> *The employers will be provided with an additional endowment of $10.00 at the start of this experiment. This amount will be added to the earnings of the employers at the end of the experiment. This is in addition to the show-up fee of $5.00 paid to every participant.*

Given that this is a lump-sum transfer paid at the conclusion of the experiment, rather than based on individual decisions, this should not distort decisions. A potential worry is that workers may suffer from a sense of inequity but as the game progresses, the workers get to see that on average the employer earnings are not very high. We believe that this mitigates any inequity concerns. In any event, one needs to trade-off between inequity concerns and concerns caused by negative earnings. We believe that the latter is a bigger worry than the former.

Hypothesis E3: Given existing evidence in favour of greater female reciprocity, female employers will either pay a higher bonus or pay the bonus more often or both compared to male employers.

Hypothesis E4: Given existing evidence in favour of greater female reciprocity, controlling for rent, we expect female workers to shirk less.

4.3. Results

In this section we report on our findings. We will report four separate results, each built around one of the hypotheses stated above.

Result 1: Female employers are no more likely to choose trust-based relational contracts than male employers. In fact, they are more likely to choose penalty-based transactional contracts under random re-matching protocol though this difference dissipates over time. Thus, there are no significant gender differences in contract choice under fixed matching protocol.

Our first hypothesis was that female employers will show a preference for the relational contract. This is not borne out in the results; If anything, female employers show a slightly greater preference for the transactional contract under random re-matching than male employers. Both men and women choose the relational contract much more frequently than the transactional contract. This is true with both fixed matching and random re-matching. Table 4.3 shows the average proportion of each contract type as well as results of non-parametric ranksum tests where we take the proportion of transactional contracts chosen by each employer over ten rounds as the unit of observation. We find that, on average, women choose transactional contract significantly more than men under random matching treatment,

but not under the fixed matching treatment. Figures 4.1A and 4.1B shows the time-series of contract choices over time.

Table 4.3: Average proportion of contracts chosen across genders

	Fixed Matching	Random Matching
Male employers	Transactional=12% (Relational=88%)	Transactional=13% (Relational=87%)
Female employers	Transactional=17% (Relational=83%)	Transactional=21% (Relational=79%)
Non-parametric Wilcoxon ranksum test z	-0.94	-2.54
p-value	0.35	0.01
Observations	Overall=33 Male=19 Female=14	Overall=38 Male=22 Female=16

Figure 4.1A: Choice of transactional contract over time under fixed matching

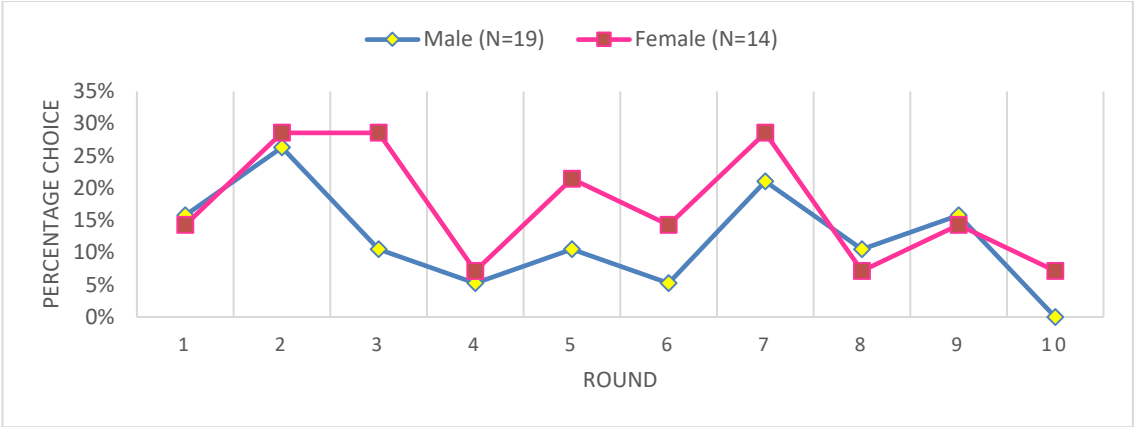
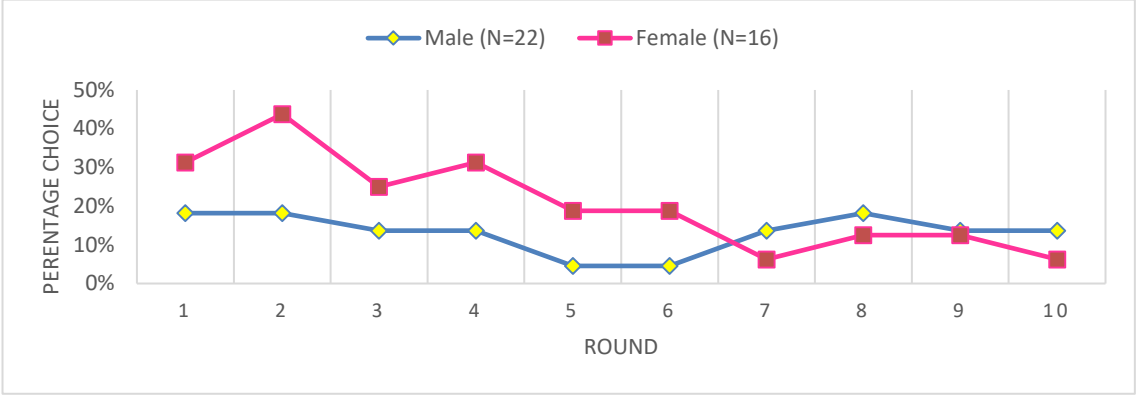


Figure 4.1B: Choice of transactional contract over time with random re-matching



In Table 4.4 we provide further corroborating evidence in the form of random effects regression results on contract choice, where we control for other relevant co-variates. The dependent variable is equal to 1 if the transactional contract is chosen, zero otherwise. The regressors include a dummy for Female (=1 if employer is female, and 0 if not); Round, an interaction term between Female and Round; how much the employer earned in the previous round and finally Lag Shirk, a dummy for whether the paired employee shirked in the previous or not. Lag Shirk =1 if the employee shirked in the previous round, and 0 otherwise.[28] When we have fixed matching, one employer-employee pair can be treated as an independent observation. Conversely, in random re-matching, the only independent observation is the session itself. Given this, in Table 4.4 and for other regressions reported below, we cluster errors at the level of individual subjects.

We present the results for fixed matching and random re-matching separately. In each case we present three separate models. The first one includes only the regressors noted above. The second one controls for risk preferences using the Holt-Laury lottery choice task and the third specification controls for demographic characteristics as well.[29] Very few of the coefficients are statistically significant for fixed matching except that of Lag Shirk, which is positive and significant. This makes intuitive sense; when the employee shirked in the previous round, the employer responded by moving to a penalty-based contract in the next round. This result is significant at 5% in the first specification but becomes only marginally significant after we control for risk preferences and demographics.

[28] As a robustness check and also for the sake of completeness, we also re-estimated all of these regressions by replacing Lag Shirk with Lag Effort, the effort exerted by the employee in the previous round. The results are unchanged and therefore, we have chosen to report the results with the Lag Shirk regressor here.

[29] We lose a few observations when we control for risk preferences and demographics. This is mostly because of participants who make inconsistent choices on the Holt-Laury lottery choice task. This is a well-known drawback of this task. When we control for risk aversion, we exclude these participants who made inconsistent choices resulting in a loss of observations. We also lose a few observations for those participants who did not complete the demographic survey.

Turning to the results for the random re-matching treatment, we find that consistent with the ranksum tests in Table 4.3, female employers are more likely than the male employers to choose a transactional contract under the random re-matching treatment. However, this difference declines over time as indicated by the negative and significant coefficient for the interaction term of female and round. It is important to understand that with random re-matching participants are not playing one another for more than one round and any effects need to be interpreted in the context of the whole session. Looking at the random re-matching protocol, we find that in the first specification the coefficients for lagged earnings and lagged shirk (whether the worker shirked in the previous round or not) are both negative and significant. This suggests that when earnings or shirking rises, employers are less likely to choose the transactional contract in the current round. This is not altogether surprising since prior work[30] in the area generally find that both effort levels and earnings are usually higher with trust-based contracts rather than penalty-based ones. This suggests that at the level of the session an increase in shirking in one round or an increase in earnings in one round leads to an increase in the choice of relational contracts. We also note that once we control for risk preferences and demographics, both these coefficients become non-significant. This suggests that some of the results in the first specification are explained partly by the interaction between gender and risk preferences and may be the result of an omitted variable bias.

[30] Fehr and Gächter (2000); Fehr, Fischbacher and Gächter (2002); Chaudhuri (2011); Nosenzo, Offerman, Sefton and van der Veen (2016).

Table 4.4: Random effect probit regression for contract choice with errors clustered on subjects.

Dependent variable: Contract Choice = 1 if transactional (penalty based) contract chosen; 0 if transformational (trust based) contract chosen

Contract Choice	Fixed Matching			Random re-matching		
	(1)	(2)	(3)	(1)	(2)	(3)
Female	0.551	0.454	0.315	**1.862*****	**2.622*****	**2.753*****
	(0.628)	(0.695)	(0.760)	**(0.737)**	**(0.815)**	**(0.945)**
Round	**-0.093***	**-0.092***	-0.091	0.013	0.076	0.074
	(0.055)	**(0.056)**	(0.056)	(0.092)	(0.111)	(0.109)
Female*Round	-0.028	-0.029	-0.030	**-0.209***	**-0.284****	**-0.284****
	(0.077)	(0.080)	(0.080)	**(0.114)**	**(0.132)**	**(0.131)**
Lag Earnings	0.009	0.002	0.001	**-0.137***	-0.020	-0.012
	(0.099)	(0.103)	(0.106)	**(0.079)**	(0.072)	(0.070)
Lag Shirk	**0756****	**0.661***	**0.661***	**-0.896****	-0.415	-0.384
	(0.337)	**(0.335)**	**(0.373)**	**(0.417)**	(0.408)	(0.403)
Risk Averse		0.285	0.154		0.217	0.077
		(0.543)	(0.523)		(0.426)	(0.523)
Constant	**-1.513*****	**-1.525*****	**-1.352***	**-1.335****	**-2.425*****	**-2.132****
	(0.565)	**(0.620)**	**(0.757)**	**(0.652)**	**(0.832)**	**(0.835)**
Demographic Control	No	No	Yes	No	No	Yes
Log likelihood	-100.078	-92.694	-89.097	-109.047	-86.479	-84.663
Wald χ^2	13.368	11.658	14.189	22.929	18.715	22.015
Prob > χ^2	0.02	0.07	0.16	0.00	0.01	0.06
Number of observations	297	270	243	342	297	297

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

The results in Table 4.4 suggest that we do not find systematic gender differences in the choice of contracts. If anything, women were marginally more likely to choose a transactional contract, compared to men. However, this difference also dissipates over time. Importantly, there is one area where we find some evidence in keeping with our initial hypothesis. If we look within the transactional contracts at the patterns of fine usage, we find that on average, women impose a fine of $1.10 compared to $1.23 imposed by men. Men tended to impose the optimal fine in 88% (46/52) of the contracts while women did so in only 65% of contracts (37/57). This is a significant difference using two-sample proportions test ($z=2.96$; $p<0.01$). However, as noted already, such transactional contracts constitute a small proportion of choices.

Result 2: Female employers offered higher rent to workers under random re-matching. This results in lower earnings for the female employers under that protocol. There are no gender differences in either rent offered or earnings under fixed matching.

This result corroborates our second hypothesis that female employers will offer higher rents. As explained above, for any contract, the employer offers a wage rate and suggests an effort level. Given that the worker's reservation wage has been normalized to zero, the employer has to pay the worker the cost of effort (or epsilon more) in order for the worker to accept the contract. Any amount higher than that implies that the worker receives rent from the employer. Figures 4.2A and 4.2B show average rent offered by male and female employers over 10 rounds in both Fixed matching treatment and Random matching treatment. We can see that female employers on average offer a higher rent in all 10 rounds compared to male employers.

Figure 4.2A: Average rent offered over 10 rounds under fixed matching

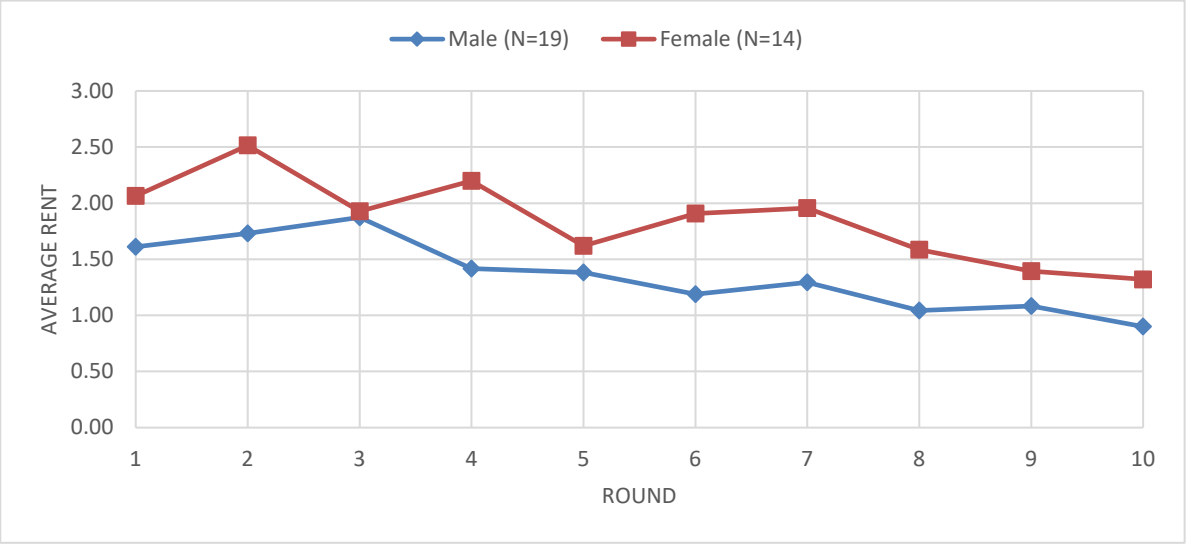
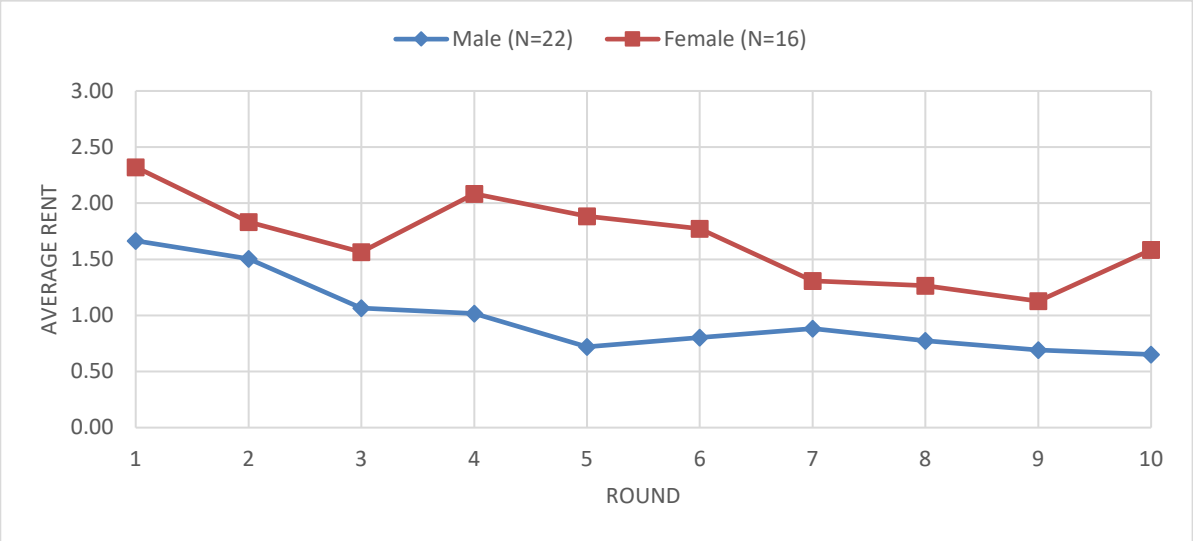


Figure 4.2B: Average rent offered over 10 rounds under random re-matching



In Table 4.5, we present results from non-parametric Wilcoxon ranksum tests for rent offered. We look separately at round 1 (which are independent observations) and at the average over 10 rounds as the units of observation. Women offer significantly higher rent in both the first round and overall with the random re-matching protocol but not under fixed matching.

Table 4.5: Non-parametric Wilcoxon Ranksum test for rent

Rent	Fixed Treatment		Random Treatment	
	Round 1	Average	Round 1	Average
Men	1.61	1.35	1.67	0.98
Women	2.06	1.85	2.32	1.67
z	-1.02	-1.39	-1.70	-1.88
p-value	0.31	0.17	*0.09*	*0.06*
Observations	Overall=33		Overall=38	
	Male=19		Male=22	
	Female=14		Female=16	

Next, in Table 4.6, we turn to a regression analysis for rent offered to examine the possible differences between male and female employers because the non-parametric test does not control for covariates. We apply a random effects model with robust standard errors clustered on individuals to account for any individual-specific components of choice behaviour. We use a random effects specification here because each decision regarding rent offered involves the specific choice behaviour of each participant.

The dependent variable is the rent offered by employer i in round t. The independent variables include (1) female (a dummy variable =1 for female and =0 for male); (2) round; (3) an interaction term between female dummy and round to detect any gender specific trends in rent over time; (4) lag cumulative earnings; (5) lag worker effort; (6) risk-averse (=1 for risk aversion and =0 otherwise); (7) a dummy variable for contract type (=1 for transactional contract, =0 for relational contract); (8) a vector of other demographic variables. The coefficient for round is negative and significant in both treatments indicating that the rent offered by employers declines over time. The female dummy is only significant under random matching treatment, consistent with our results for ranksum tests. Employers choosing the transactional contract under the fixed matching treatment are more likely to make a higher rent offer. The coefficient for transactional contract is marginally significant (at 10%) in the fixed matching protocol. This suggests that both genders offer higher rent in this type of

contract but as noted the effect is only marginal and therefore we do not elaborate on this further.

Table 4.6: Random Effects Regression on Rent Offered (errors clustered on subject)

Rent Offered	(1)	(2)	(3)	(4)
Female	0.685	0.829	**1.088****	**0.935****
	(0.603)	(0.590)	**(0.467)**	**(0.426)**
Round	**-0.107*****	**-0.107*****	**-0.053*****	**-0.053*****
	(0.032)	**(0.033)**	**(0.018)**	**(0.018)**
Female*Round	-0.008	-0.008	-0.041	-0.041
	(0.069)	(0.070)	(0.039)	(0.040)
Lag Earnings	-0.017	-0.016	0.000	0.001
	(0.030)	(0.030)	(0.019)	(0.019)
Transactional Contract	**0.507***	**0.519***	0.243	0.249
	(0.297)	**(0.298)**	(0.197)	(0.196)
Risk Averse	-0.557	-0.518	0.085	0.030
	(0.366)	(0.367)	(0.354)	(0.410)
Constant	**2.118*****	**2.051*****	**1.112*****	**1.085*****
	(0.461)	**(0.448)**	**(0.229)**	**(0.251)**
Demographic control	No	Yes	No	Yes
Wald χ^2	33.30	38.70	31.04	41.70
Prob > χ^2	0.0000	0.0000	0.0000	0.0000
Number of observations	270	270	297	297

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

In Table 4.7, we present results for random effects regressions for earnings per round. We do not find any significant gender differences in earnings under fixed matching treatment, however, we do find that that earnings are lower in (penalty-based) transactional contracts compared to the (trust-based) relational contracts.[31] With random re-matching, female employers earn less compared to male employers though we note that the interaction term between the female dummy and round is positive and significant. This suggests that over time this difference in earnings reduces. The result is not surprising given that female employers tend to make higher rent offers to the workers. It is also the case that under random re-

[31] As noted already, it is a common finding in the literature that contracts relying on mutual trust and reciprocity often lead to higher efficiency compared to penalty-based contracts. This finding is true for some of the original papers by Fehr and his co-authors, such as Fehr et al. (1997). This finding is also reported in the closely related Chaudhuri et al. (2015) study. Therefore, we refrain from elaborating further on this.

matching women tend to choose transactional contracts more often and these contracts perform worse in terms of earnings. Over time women move away from these penalty-based contracts and this results in a narrowing of the earnings gap between men and women. The coefficient of the rent variable is negative and significant under both fixed matching treatment and random matching treatment suggesting that contracts with higher rent offers resulted in lower earnings for the employer. We do not find any significant gender differences in earnings of the worker for both of our treatments.

Table 4.7: Random Effects Regression on Employer Earnings (errors clustered on subject)

	(1)	(2)	(3)	(4)
Rent	**-0.292****	**-0.219***	**-0.600*****	**-0.609*****
	(0.140)	**(0.121)**	**(0.152)**	**(0.145)**
Female	-0.293	-0.660	**-2.079*****	**-2.243*****
	(0.869)	(0.850)	**(0.624)**	**(0.622)**
Transactional Contract	**-2.434*****	**-2.583*****	**-0.752****	**-0.688***
	(0.374)	**(0.374)**	**(0.328)**	**(0.364)**
Round	-0.024	-0.020	**-0.218*****	**-0.218*****
	(0.078)	(0.076)	**(0.050)**	**(0.050)**
Female*Round	0.047	0.049	**0.261*****	**0.264*****
	(0.125)	(0.125)	**(0.090)**	**(0.090)**
Suggested Effort	**0.257*****	**0.276*****	-0.064	-0.063
	(0.072)	**(0.091)**	(0.047)	(0.046)
Risk Averse	0.075	0.189	-0.017	0.092
	(0.387)	(0.321)	(0.257)	(0.303)
Constant	0.898	0.456	**3.506*****	**3.436*****
	(0.878)	(0.771)	**(0.548)**	**(0.577)**
Demographic Control	No	Yes	No	Yes
Wald χ^2	79.07	149.42	67.23	96.51
Prob > χ^2	0.0000	0.0000	0.0000	0.0000
Number of observations	300	300	330	330

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

Result 3: Male employers tend to propose a higher bonus and pay a higher bonus on average

This result deals with our third hypothesis that within relational contracts, women will tend to be more generous by offering higher bonuses. This conjecture turns out to be incorrect. In Table 4.8, we present ranksum tests for proposed bonus and actual bonus in round 1 for both fixed matching treatment and random matching treatments. On average, men proposed $2.25 in round 1 while women proposed $1.5, and, on average men paid an actual bonus of 1.81 in round 1 while women paid an actual value of $0.75. Both the proposed bonus and the actual bonus are marginally higher (significant at 9%) for men compared to women under the fixed matching treatment. However, men pay a lower actual bonus under random matching treatment than women.

Table 4.8: Ranksum tests for proposed and actual bonus by gender and treatment in Round 1

Fixed matching treatment		Random re-matching treatment	
Proposed	Actual	Proposed	Actual
Male=2.25	Male=1.81	Male=3.28	Male=0.28
Female= 1.50	Female= 0.75	Female= 2.64	Female= 0.73
z=1.90	z=1.69	z=0.74	z=-1.66
p=0.06	p=0.09	p=0.46	p=0.10
Male=16 Female=12		Male=18 Female=11	

We also looked at the average values for the proposed bonus and the actual bonus over 10 rounds presented in Table 4.9. Male employers tend to propose a higher bonus and pay a higher bonus on average and in round 1 than female employers under the fixed matching treatment. Male employers also propose a higher bonus than female employers under the random matching treatment. We do not find any significant gender differences between the average value of actual bonus paid.

Table 4.9: Ranksum tests for average proposed bonus and average actual bonus by gender and treatment

Fixed matching treatment		Random re-matching treatment	
Proposed	actual	proposed	actual
Male=2.67	Male=1.58	Male=3.84	Male=0.37
Female= 2.01	Female= 0.73	Female= 2.59	Female= 0.35
z=1.99	z=2.00	z=1.77	z=0.02
p=0.05	p=0.05	p=0.08	p=0.99
Male=19 Female=14		Male=21 Female=16	

Result 4: There is no significant gender difference in shirking under fixed matching treatment. However, we find that female workers are more likely to shirk under the random matching treatment, but that this trend disappears over time.

This result has to do with our fourth and final hypothesis suggesting that female employees, being more reciprocal, in general, will tend to shirk less, controlling for the level of rent offered by the employer. We use a random effects probit regression model with shirking as the dependent variable. We define the dummy variable shirking = 1 if actual effort is less than suggested effort and otherwise shirking = 0. We look at the relational contract and the transactional contract separately. Under the relational contract, our independent variables include (1) the rent offered by the employer, which can be viewed as an intrinsic incentive for the workers to choose the suggested effort level; (2) female; (3) round; (4) an interaction term between the female dummy and the round; (5) the proposed bonus; (6) an interaction term between the female dummy and the proposed bonus; (7) lag earnings; (8) risk-averse (=1 for risk aversion and =0 otherwise); (9) a vector of other demographic variables. The results are presented in Table 4.10.

Table 4.10: Random Effects probit regression on shirking under Relational Contract (standard errors clustered on the subject)

Dependent variable: Shirking = 1 if actual effort < suggested effort; 0 otherwise

	Fixed matching	Random re-matching
Rent	-0.192	-0.035
	(0.173)	(0.082)
Female	-0.119	0.553
	(0.683)	(0.680)
Round	**0.108***	0.079
	(0.058)	(0.054)
Female*Round	-0.006	**-0.182****
	(0.079)	**(0.079)**
Proposed Bonus	**-0.443*****	-0.003
	(0.144)	(0.065)
Female*Proposed Bonus	**0.323***	-0.010
	(0.185)	(0.100)
Lag Earnings	0.021	0.025
	(0.107)	(0.058)
Risk Averse	-0.597	-0.331
	(0.410)	(0.244)
Constant	-0.188	0.032
	(1.143)	(0.770)
Demographic Control	Yes	Yes
Wald χ^2	35.30	52.02
Prob > χ^2	0.0000	0.000
Number of observations	206	262

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

Under relational contracts, the amount of rent offered does not seem to have a significant effect on the decision to shirk or not. However, the amount of bonus offered by the employer does make a difference and higher bonuses promised (even though this is akin to cheap-talk) leads to lower shirking. Though, we do not find any significant gender differences in the level of shirking under the relational contract with either fixed matching or random re-matching.

Under a transactional contract, our independent variables include: (1) the rent offered; (2) extrinsic incentive;[32] (3) female; (4) round; (5) an interaction term between the female dummy and the round; (6) an interaction term between the female dummy and extrinsic; (7) values of the fine imposed; (8) risk-averse (=1 for risk aversion and =0 otherwise); (9) lag cumulative earnings;. The results are presented in Table 4.11. Here we have combined the data for the fixed matching and random re-matching protocols because, as noted earlier, the majority of the time the employers chose the relational contract. This implies that there are few observations for the transactional contract and if we further divide this by the matching protocol then we have very few observations from which to estimate.

As with the relational contract, rent does not seem to have an effect on shirking. However, for the transactional contracts, the coefficient for extrinsic incentive is negative and significant which suggests that workers do pay attention to the payoff differential from not shirking as opposed to shirking. When the payoff differential increases, workers responded by lowering their likelihood of shirking. The coefficient on the female dummy is negative and significant indicating that female workers are more likely to shirk compared to male workers. Though this effect decreases over time.

[32] In the transactional contract treatment, workers face a trade-off between shirking or not. If a worker shirks then he/she incurs a lower cost of effort but may be found out and therefore penalized. On the other hand, not shirking means incurring a higher effort cost but no possibility of being fined. The extrinsic incentive ($U^{ns} - U^{s}$) variable is defined as the expected payoff to the worker from not shirking minus the expected payoff when the worker does shirk. The value of this variable depends on the wage offered and the suggested effort and follows from the theoretical model presented in the Appendix 6. Suffice it to say that in the context of the transactional contracts, the rent acts as an intrinsic incentive to provide effort while this other variable provides a proxy for the extrinsic incentive to provide effort.

Table 4.11: Random Effects probit regression on shirking under Transactional Contract

	Fixed matching and Random re-matching
Rent	-0.094
	(0.126)
Extrinsic	**-1.460****
	(0.597)
Female	**2.057****
	(0.943)
Round	0.190
	(0.150)
Female*Round	**-0.291****
	(0.133)
Female*Extrinsic	1.190
	(0.733)
Fine	0.705
	(0.518)
Risk Averse	-0.255
	(0.403)
Lag Cumulative Earnings	-0.017
	(0.036)
Constant	-1.224
	(1.368)
Demographic Control	Yes
Wald	23.13
	0.0266
Observations	72

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

4.4. Concluding remarks

We designed this study to test the proposition that in terms of the work-place, female leaders tend to be more transformational in the sense of relying more on mutual trust and reciprocity based relationships, compared to male leaders, who rely more on explicit "carrots" and "sticks". Our views in this regard were formed on the basis of earlier work looking at gender differences in the context of labour markets. We conjectured that such differential preferences on the part of men and women may have interesting and interpretable implications for their relative success in the workplace. By and large, we fail to find dramatic differences in behaviour. Contrary to our *priori* hypothesis, women showed a slight preference for transactional contracts. When it comes to worker behaviour, women actually tend to shirk a little more. However, these differences are neither major nor permanent.

We do find some evidence that regardless of contract type, female employers offer higher rent to the workers. This is in line with prior findings regarding greater female generosity. Offering higher rents leads to lower the earnings for those female employers, though, the differences are not stark. Overall, we conclude that when it comes to work-place strategies, there are no dramatic differences in the strategies employed by men and women.

Overall this suggests that the differential outcomes for men and women observed in the workplace both in terms of the gender wage gap and the leadership gap possibly do not arise due to dramatic differences in psychological preferences or work-place strategies. Therefore, as pointed out by Bohnet (2016), such a result suggests a requirement for more pro-active work-place practices and institutional changes that work toward ensuring greater gender parity.

5. Gender and Trust

5.1. Introduction

In *The Theory of Moral Sentiments*, Adam Smith states that "in civilized society [man] stands at all times in need of the cooperation and assistance of great multitudes, while his whole life is scarce sufficient to gain the friendship of a few persons" (p.26). In many cases, we need to cooperate with others including those whom we do not know. Economic transactions are generally associated with trust behaviour, which includes trusting others, being trusted, and reciprocating others' trust.

Trust is essential in many situations. Trust and trustworthiness are robustly related to life satisfaction and well-being. Individuals with a higher level of trust and reciprocity enjoy high social capital (Helliwell, 2006). A higher level of average employee trust within organizations is positively related to financial performance and labour productivity (Brown, Gray, McHardy, and Taylor, 2015). Trust is considered important to social function and economic growth (Knack and Keefer, 1997; Algan and Cahuc, 2010). Social scientists have typically relied on survey responses to measure trust, such as the standard trust questions from the World Values Survey and the General Social Survey, and other types of questions. Recently, the trust game, also known as the investment game, introduced by Berg, Dickhaut, and McCabe (1995), has been widely used to measure trust and reciprocity. This game is a two-person game in which the "sender" or "proposer" decides to transfer some, none or all his or her $10 experimental payment to a "receiver" or "responder". Any amount transferred gets tripled by the experimenter before received by the responder. The responder then decides whether to keep the entire tripled amount or transfer some of it back to the proposer. The game ends after the responder's decision.

Applying backward induction and assuming no post-game consequences, the responder has no incentive to return any money. Anticipating that, the proposer should not transfer any money. This is the unique subgame-perfect Nash equilibrium of the game. However, both players are better off if the proposer decides to send some money, and the responder decides to return an amount higher than the amount sent by the proposer. The level of trust is typically measured by the amount sent from the proposer, while reciprocity is measured by the percentage returned by the responder.

A few studies applied the trust game to explore the issue of gender and trust behaviour. Appendix 7 contains a selected list of papers that explore such differences using gift-exchange games along the lines of Berg et al. (1995) or variants thereof. When we refer to a game as "discrete" we mean that trustors in that game had a choice between two choices, loosely speaking to either repose trust in the trustee or not. When we refer to a game as "continuous" we mean that– much like in the original Berg et al. (1995) paper – trustors in that game had several choices, for instance, a number of levels of how much to send out of an initial endowment of $10.

Cookson (2000, p 56-57) writes: "Experimental economists typically devote a great deal of effort into investigating complex variations in strategy sets … in order to test competing theories. By contrast, rather less time and effort is spent on investigating simple variations in how those strategy sets are described to subjects". Cookson (2000) argues that the "framing effect" where one applying different ways to describe the same choice problem could lead to changes in behaviour. Levati, Miettinen, and Rai (2010) argue the importance of possible mismatch between experimenters and participants. The behaviour of participants could be influenced by various contextual features. It is essential that the participant's perception of the experimental games coincides with the model that the experimenter has in mind. Samuelson (2005) also argues the importance of the framing effect. He writes: "Despite

an experimenter's best efforts to ensure that subjects understand what they are dealing with, including careful presentations, questions, and preliminary quizzes, it is not clear when we can be confident that the subjects' models match the experimenter's". The interpretation of experimental data depends crucially on how we imagine participant perceives the game.

Previous studies from the trust game generally report either that there are no significant gender differences or that men exhibit more trust than women while women reciprocate more than men. Given that most of the studies focused on one-shot games, one possible explanation is the misunderstanding or misinterpretation between participants and experimenters due to there being little scope for the participants to learn the game. Given that most studies use neutral and context-free language, there may well be cognitive demand effects. The interpretation of the task by participants may be different from that intended by the experimenter. Furthermore, it is conceivable that such cognitive effects themselves may be gender dependent. In this chapter, I will examine whether and how gender differences influence the cognitive demand effect.

In our experiment five different treatments are used to elucidate the underlying differences in preferences that may not be captured by context-free instructions. We study a repeated version of the Berg et al., (1995) trust game with random re-matching. In our first treatment, which we refer to as the *private knowledge* treatment, participants are only provided with context-neutral written instructions. In our second treatment, which we refer to as the *common knowledge* treatment, participants are provided with the same context-neutral written instruction as in the *private knowledge* treatment, but here, the instructions are also read out loud to the participants by the experimenter prior to the start of the session. The intention here is to create common knowledge in the sense that everyone has heard the instruction, and everyone knows everyone else has heard the same instruction. We regard the *private knowledge* and *common knowledge* treatments as our control treatments. The reason

for considering both treatments as control treatments is that prior studies of trust and reciprocity generally use one of these treatments.[33]

Our three context treatments are the *context neutral* treatment, the c*ontext-loaded A* treatment and the c*ontext-loaded B* treatment. In all three context treatments, participants receive the same written instructions as in our control treatments, which are also read out loud. However, two additional paragraphs are added to the instruction. A detailed explanation of those treatments is provided in the section describing our experimental design and the instructions are included in Appendix 3.

The difference between our three *context* treatments is as follows. In our *Context-neutral* treatment, only context-free language is used in the additional instruction as follows:

One way to think about this situation is as follows: the second mover has no incentive to send any money back to the first mover because the round ends immediately after that. Anticipating that, the first mover should hang on to his $10.00 and send nothing to the second mover. This means they will both end the round with $10.00 each.

But suppose the first mover decides to transfer $10.00 to the second mover. Then the second mover will get $30.00. If the second mover sends back an amount more than $10.00 then it is easy to see that both the first mover and the second mover can make more money than if they simply hung on to their $10.00 in each and every round.

For two *context* loaded treatments, we use the language "trust" and "trustworthy" and apply the above two paragraphs in a different order. Our additional instruction contains both a negative and a positive message. Ex-ante, there is no compelling reason to think that

[33] See Appendix 7; in Appendix 7 we provide a list of prior studies and whether they used only written instructions, or both provided written instructions and read the instructions out loud.

participants will prioritize one message over the other. The additional instruction used in *context-loaded* A is as follows:

But the second mover has no incentive to send any money back to the first mover because the round ends immediately after that. Anticipating that, the first mover should hang on to his $10.00 and send nothing to the second mover. This means they will both end the round with $10.00 each.

One way to think about this situation is as follows: suppose the first mover decides to trust the second mover by sending $10.00. Then the second mover will get $30.00. If the second mover behaves in a trustworthy manner and sends back an amount more than $10.00 then it is easy to see that both the first mover and the second mover can make more money than if they simply hung on to their $10.00 in each and every round.

The *context-loaded* B treatment is essentially the same as *context-loaded* A, but the ordering of two above paragraphs is reversed. In *context-loaded* B, we present the trust and reciprocity based explanation first and the self-interest-based explanation (resulting in the Nash equilibrium outcome) second.

Aggregating over individuals, we find interesting and interpretable differences across the three treatments. In Chaudhuri, Li, and Paichayontvijit (2016), we highlight the importance of context. The level of trust and reciprocity are higher when additional context is provided in the instruction, even if this context is presented using neutral language. There is a sharp increase in the levels of trust and reciprocity in moving from *private* or *common*

knowledge to *context neutral*. However, there are no significant differences between the *context neutral* and the two *context-loaded* treatments.[34]

We exclude a detailed discussion of these treatment differences in the current study and instead focus on the issue of gender differences.[35] In the *private knowledge* and *common knowledge* treatment, we do not find any significant gender differences in trust, and men appear to be more reciprocal than women in the *common knowledge* treatment. In our three *context* treatments, we do find that both gender explicit more trust and being more reciprocal compared to our control treatments, but we do not find any significant gender differences within our context treatments. Our results suggest that the availability of an explicit context leads to a significant increase in trust and reciprocity for both men and women.

This chapter proceeds as follows. We describe our experimental design in Section 5.2. We present our results in Section 5.3 and make some concluding remarks in Section 5.4.

5.2. Experimental design and procedures

5.2.1. Experimental Design

In our experiments, participants play 10 rounds of a trust game (first studied by Berg et al (1995) in a one-shot setting) with participants being randomly re-matched prior to the beginning of each round. This "stranger" matching protocol retains the nature of one-shot games while allowing us to study the dynamics of decision making. This design makes our results more comparable with previous studies which have concentrated on one-shot games

[34] We report these treatment differences in Chaudhuri, Li, and Paichayontvijit (2016), which complements this current analysis.
[35] We lost demographic questionnaires which contain gender information in one session of *context-loaded* A treatment. We ran another session of Context-loaded A and used the new session in this chapter. This indicates that we have the same number of sessions in this chapter and in Chaudhuri, Li and Paichayontvijit (2016). However, the data in the *context-loaded* A treatment is different in one session.

as it allows learning and gathering experience while it removes the incentive for reputation building associated with fixed matching in repeated games.

At the beginning of each round each participant is endowed with ten experimental dollars—which is equivalent to $1 NZ dollar. In each round, the proposer has to decide how to allocate his or her endowment, i.e. how much money to pass to the responder out of his or her endowment and how much to keep. The proposer can send any amount between zero to ten, including decimal numbers. However, the majority chooses whole numbers. If the proposer chooses to send zero, then that round ends immediately after the proposer's decision. Both proposer and responder will earn 10 experimental dollars for that round. Any non-zero amount passed gets tripled before it is received by the responder, who then decides how much to keep and how much to pass back to the proposer. For example, if the proposer decides to send a strictly positive value, X, then the responder would receive an amount of $3X$. If the responder chooses to return nothing to the proposer, then earnings for the proposer would be $10 - X$ and the responder would end up with $10 + 3X$. If the responder decides to transfer a positive value to the proposer, say Y, then the proposer would end up with $10 - X + Y$ and the responder would earn $10 + 3X - Y$. At the end of each round, each matched pair is able to see the amount transferred between them and their earnings for that round on their computer screens. The round ends after the responder's decision. The game proceeds in the exact same way for all ten rounds.

Given that each participant is re-matched with another person in each round, we can use backward induction to solve for the equilibrium outcome. Since the game ends immediately after the responder's decision and there is no incentive to build any reputation, a self-interested responder would want to maximize his own payoff by keeping every dollar he received and never return a positive value back to the proposer. Anticipating the responder's decision, then proposer would send nothing to the responder in the first place.

Therefore, both proposer and responder would end up with 10 experimental dollars in each round. However, if the proposer chooses to send any positive value, the responder would receive a tripled value, and if the responder returns any value higher than the value transferred by the proposer, then both parties can be better off than with their initial endowment. The amount transferred by the proposer could be used to measure the amount of trust the proposer has towards the responder and the proportion of the tripled amount returned would measure reciprocity.

5.2.2. Treatments

We apply five different treatments to a ten-fold replication of a trust game with random re-matching. We explore those five treatments in detail in the following sections.

5.2.1. Private Knowledge Treatment

The first one is *private knowledge* treatment. Under this treatment, participants are only provided with written instructions of the game using neutral, *context*-free, language. Words such as trust and reciprocity are not used. Participants read the instruction on their own and communication is not allowed. Instructions for all treatments are provided in Appendix 3. A summary of prior studies that provide only written instructions and those that both provide written instruction and also read the same instructions out loud is included in Appendix 7.

5.2.2 Common Knowledge Treatment

The second treatment is the *common knowledge* treatment. Participants were provided with the same written instructions as in the *Private knowledge* treatment, but the experimenter also read the instruction out loud to all participants before round 1 started. The idea of reading out the instructions is to make sure everyone has heard the instruction, and everyone also knows that everyone else has heard the instruction. The *common knowledge* treatment is designed to

reduce the uncertainty regarding whether all participants have paid attention to the instructions; in other words, to remove the lack of public knowledge. Previous studies suggest that reading the message out loud does improve the common perception of said message.[36] If the absence of extra returns to trust is a result of uncertainty regarding whether all participants have understood the game and is not by virtue of understanding the game itself, we would expect a higher trust and reciprocity level under the *common knowledge* treatment than under the *private knowledge* treatment.

5.2.3 Context-Neutral Treatment

The third treatment is the *context neutral* treatment. Participants were provided the same written instructions with two additional paragraphs and these instructions were also read out loud by the experimenter. The two additional paragraphs were as follows:

One way to think about this situation is as follows: the receiver has no incentive to send any money back to the sender because the round ends immediately after that. Anticipating that, the sender should hang on to his $10.00 and send nothing to the receiver. This means they will both end the round with $10.00 each.

But suppose the sender decides to transfer $10.00 to the receiver. Then the receiver will get $30.00. If the receiver sends back an amount more than $10.00 then it is easy to see that both the sender and the receiver can make more money than if they simply hung on to their $10.00 in each and every round.

The additional paragraphs are introduced here in order to remove any uncertainties raised because not every participant may be convinced that every other participant has

[36] See Chaudhuri and Sbai (2011); Chaudhuri, Schotter and Sopher (2009); Bangun, Chaudhuri, Prak and Zhou (2006)

interpreted the game in the same way. Even if every participant is convinced that everyone else has heard the instructions read out loud by the experimenter in the *common knowledge* treatment, this does not mean that every participant would understand the experimenter's purpose and draw the same conclusion. This treatment makes it more likely that the experimenter's interpretation of the game and the participants' interpretation of the game are the same and that each participant believes that the other participants have the same interpretation.

5.2.4 Context-Loaded A Treatment

The fourth treatment is the *context-loaded A* treatment. The instructions are similar to those in the *context neutral* treatment except that here we explicitly use the words "trust" and "trustworthiness" in the second paragraph. The idea is to make the motivations explicit to the participants. The additional two paragraphs are:

One way to think about this situation is as follows: the receiver has no incentive to send any money back to the sender because the round ends immediately after that. Anticipating that, the sender should hang on to his $10.00 and send nothing to the receiver. This means they will both end the round with $10.00 each.

But suppose the sender decides to trust the receiver by sending $10.00. Then the receiver will get $30.00. If the receiver behaves in a trustworthy manner and sends back an amount more than $10.00 then it is easy to see that both the sender and the receiver can make more money than if they simply hung on to their $10.00 in each and every round.

5.2.5 Context-Loaded B Treatment

The fifth and last treatment is the *context-loaded* B. The additional two explanation paragraphs are identical to the *context-loaded* A treatment but here we interchange the order of the two paragraphs. In the *context-loaded* A treatment, the explanation of the sub-game

perfect equilibrium comes first while the trust and reciprocity based explanation comes second. In the *context-loaded* B treatment, this order is reversed to investigate whether presentation order makes any differences to participants' decision making.

5.2.6 Summary of Treatments

The design of the three *context* treatments is to remove "cognitive" demand effects which may arise because (1) participants incorrectly interpret the task, or (2) there is at least a mismatch between the expected outcome that the experimenter wishes to achieve and how the participants perceive the same task. One issue associated with our three *context* treatments is the experimenter induced (or social) demand effects. This may occur as we remove cognitive demand effects by explaining the game. However, we do not believe this is a concern because our instructions contain both a negative message stating that a strategically self-interested sender should not send any money and a positive message indicating that both senders and receivers are better off if they trust and are trustworthy. We introduce the A and B versions of the *context-loaded* treatment to check whether the order of the explanations made any difference to the participants' decision making. We implement the *context neutral* treatment which leaves out any loaded words such as trust or reciprocity, and the two *context-loaded* treatments to remove "cognitive" demand effects while carefully limiting any "social" demand effects.

In the *private knowledge* treatment, participants can suffer both lack of common knowledge by not reading aloud the instructions and the lack of context by not receiving an explanation of the strategic imperatives of the game. In the *common knowledge* treatment, the instructions are read out by the experimenter to limit the extent to which participants may be uncertain as to whether the other participants were aware of the nature of the game. In the three *context* treatments instructions are read out, and additional paragraphs providing an explanation of the strategic imperatives of the game are both included.

We use the *private knowledge* and the *common knowledge* treatments as our control treatments. If there is a significant difference in behaviour between the *private knowledge* treatment and the *common knowledge* treatment, then we would conclude that there is uncertainty about the nature of the game or, at least about the knowledge of others about the nature of the game, that participants are concerned whether other participants have correctly comprehended the true incentives in the game which may lead to a lower level of trust. On the other hand, if there is a significant difference between the results in the control treatments and the results in the *context* treatments, then we could conclude that trust is not being sufficiently reciprocated because of a failure to understand the incentives inherent in the game. Lastly, by comparing between the three *context* treatments, we seek to find whether participants are successful in interpreting the task the same way as the experimenter, by including or excluding the explicit use of words like "trust" and "trustworthiness".

The following table provides a summary view of our experimental design.

Table 5.1 Summary of experimental design for the Trust Game

	Private Knowledge	*Common Knowledge*	*Context Neutral*	*Context-Loaded* A	*Context-Loaded* B
Written Instructions	YES	YES	YES	YES	YES
Instructions Read	NO	YES	YES	YES	YES
Additional Message	NO	NO	YES	YES	YES
Non-Neutral Words	NO	NO	NO	YES	YES

5.3.3 Experimental Procedure

A total of 380 students took part in this experiment, with 190 proposers and 190 responders. All sessions in this experiment were conducted in the DECIDE laboratory at the University of Auckland using the trust game available in the Veconlab website developed by Professor

Charles Holt at the University of Virginia[37]. Participants were recruited via an email announcement, and they were students from undergraduate courses without any prior knowledge related to the trust game. Students are told that they can expect to earn around $25 dollars (including the show-up fee) in a session which lasts approximately 90 minutes. Participants can only sign up for one session to avoid obtain repeated observations of the same participant.

The experimenter advised participants to sit in any available cubicle with a computer once they entered the lab. There are dividers between each cubicle so that each participant is separated from another and is unable to see any other participant's computer screen. This prevents any opportunities to observe the decisions made by any other participants during the experiment.

At the beginning of the experiment, the experimenters distribute game instructions to each participant and read it out loud (except for the *private knowledge* treatment). Participants are informed of the random re-matching protocol and that they are unable to learn the identity of the person they are matched with at any time. Participants were given 5-10 minutes to read the instructions and ask any questions they might have. The experimenters then distribute the log-in instructions and help participants to log in to the website. Once participants are logged on to the website, they are each assigned with a game ID number and they are randomly assigned to the role of either a proposer or a responder. These roles remain unchanged for the entire session. Participants are handed a set of demographic questionnaires at the end of the game. The demographic questionnaire is used to collect information regarding participant's gender, the field of study, year in the undergraduate program, age, income, whether they were

[37] http://vecoblab.econ.virginia.edu/admin.htm

born in New Zealand, and their ethnicity. Each session lasts approximately an hour. At the end of the session, participants were paid their earnings in cash from the trust game plus a $5 show-up fee. Participants are told that their earnings are private information and they are free to leave after collecting their payment. The average earnings are $20 for proposers and $21 for responders, excluding the $5 show-up fee.

5.3.4 Hypotheses

Our main hypotheses are listed as follows:

(1) Males would display higher levels of trust in the *private knowledge* treatment and the *common knowledge* treatment than females, but not in the three *context* treatments.

(2) Females would display a higher level of reciprocity in the *private knowledge* treatment and the *common knowledge* treatment than males, but not in the three *context* treatments.

According to previous research on trust game, men are more trusting than women, while women are more reciprocal than men. We expect the same result would arise in our in *Private knowledge* treatment and *common knowledge* treatment. We do not expect any gender differences to occur in three *context* treatments that provide a detailed explanation of the contingencies of the game. However, if there exists a significant difference between male and female within the three *context* treatments, then with a degree of confidence, we can attribute this to the underlying differences in gender preferences rather than a result of task mismatch between the experimenter and the participants.

We can look at differences in behaviour in two ways, differences in the behaviour of all participants between different treatments and, for each treatment, differences in the behaviour of females and males. In the next section, we will explore gender differences within and between our five treatments. We have looked at the differences in behaviour between the

different treatments with the genders aggregated in a different paper (Chaudhuri et al. 2019). We find that both the trust level and the reciprocity level are higher in the three *Context* treatments, than in the *private knowledge* treatment and the *common knowledge* treatment. This result indicates that the average behaviour of the two genders is sensitive to the way the instructions are framed. The issue with looking at the treatments, is that we do not know if the two genders responded differentially to them; E.g. one treatment may result in lower trust by men and higher trust by women but this difference will not be apparent when looking at treatment effects only. In this chapter, we are looking at gender differences within and across treatments.

In total there are 380 participants, 190 assigned to each role. Altogether there are 82 participants in the *private knowledge* treatment, 78 participants in the *common knowledge* treatment, 68 participants in the *context neutral* treatment, 82 participants in the *context-loaded* A treatment, and 70 participants in the *context-loaded* B treatment. Table 5.2 summarizes information for all five treatments regarding the number of participants in one role and gender composition.

Table 5.2: Details of experimental design

	Private Knowledge	*Common Knowledge*	*Context Neutral*	*Context-* loaded A	*Context-* loaded B
Number of Sessions	4	4	4	4	4
Participants per session	[20, 20, 20, 22]	[16, 24, 18, 20]	[18, 16, 18, 16]	[18, 24, 20, 20]	[16, 16, 20, 18]
Total Participants	82	78	68	82	70
Senders	41	39	34	41	35
	Male = 22	Male=25	Male=18	Male=24	Male=15
	Female = 19	Female=14	Female=16	Female=17	Female=20
Receivers	41	39	34	41	35
	Male=25	Male=25	Male=18	Male=24	Male=18
	Female=16	Female=14	Female=16	Female=17	Female=17

5.4 Results for Gender Differences

This section focuses on gender differences in trust and reciprocity within and across five treatments. Trust is measured by the amount transferred by the sender and reciprocity is measured by the proportion of the amount received returned to the sender by the receiver.

5.4.1 Gender differences in Trust

Result 1: Both men and women are more trusting in the context treatments compared to the control treatments

Table 5.3: Summary Statistics for Senders

Average amount sent	*Private Knowledge*	*Common Knowledge*	*Context Neutral*	*Context-loaded* A	*Context-loaded* B
Total	2.56	4.39	6.15	5.42	6.75
Male	2.71	4.74	6.16	5.99	6.66
Female	2.37	3.77	6.14	4.62	6.82

Table 5.3 provides summary statistics for proposers regarding the average amount transferred in the five treatments. It shows that on average, proposers in the three *context* treatments transferred the most (above 50% of their initial $10 endowment), followed by those in the *common knowledge* treatment and then those in the *private knowledge* treatment. Proposers in the *context-loaded B* treatment passed around $6.75, proposers in the *context neutral* treatment passed $6.15, and proposers in the *context-loaded A* passed $5.42. This pattern holds for both male and female proposers.

Figure 5.1: Average amount passed by males over 10 rounds

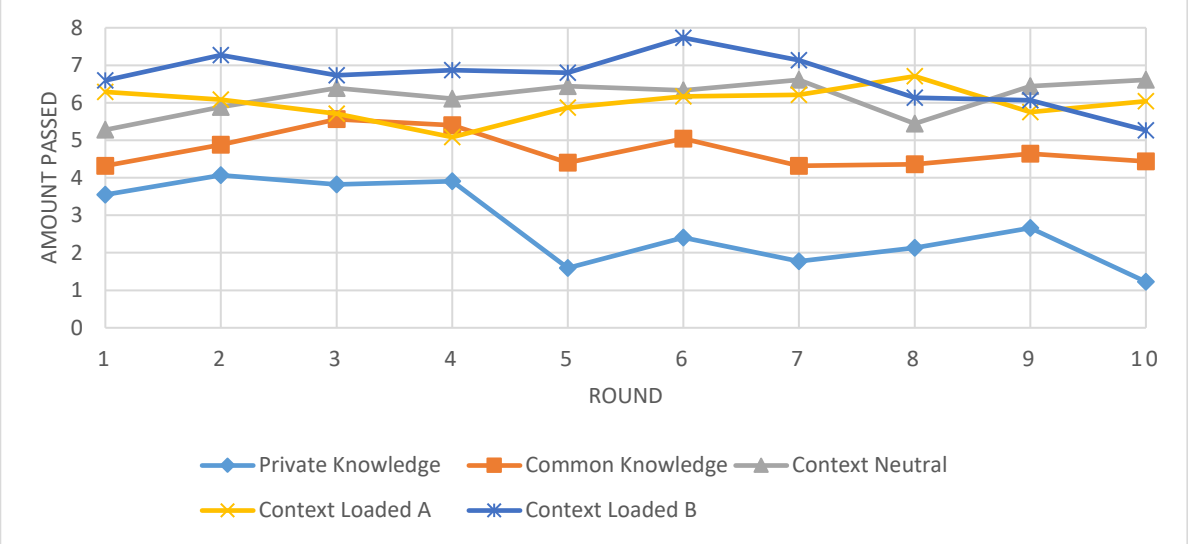


Figure 5.2 Average amount passed by females over 10 rounds

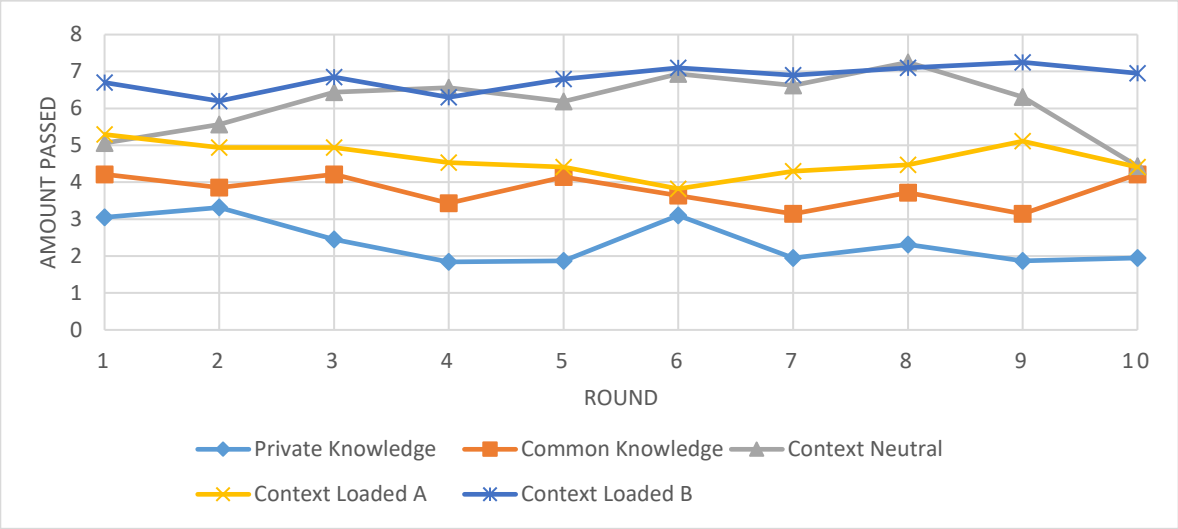


Figure 5.1 shows the average amount transferred by males over 10 rounds across five treatments while Figure 5.2 shows the average amount transferred by females. For males, there is a considerable increase when we move from the *private knowledge* treatment to the *common knowledge* treatment whereas for females the differences are less pronounced. However, on average both genders transferred more in all three *context* treatments compared to the *private knowledge* treatment and the *common knowledge* treatment.

Considering the situation where senders choose to transfer all $10 dollar to the receiver, around 10% male participants sent everything in the *private knowledge* treatment, 27% in the *common knowledge* treatment, 38% in the *context neutral* treatment, 41% in *context-loaded* A treatment and 33% in the *context-loaded* B treatment. For female participants, the corresponding proportions are 10% in the *private knowledge* treatment, 14% in the *common knowledge* treatment, 28% in the *context neutral* treatment, 25% in the *context-loaded* A treatment and 31% in the *context-loaded* B treatment. A Kruskal-Wallis test comparing the amount transferred by males in round 1 across the five treatments rejects the null hypothesis that these observations are drawn from the same distribution ($p=0.026$), and the same test comfortably rejects the same null hypothesis for females ($p<0.01$).

Since participants are randomly paired for each round, within-session observations are not independent. Therefore, we look at the first round decision only as these are not contaminated by any learning or repeated game effects and constitute independent observations. Table 5.4A summarizes observation details between genders across treatments. We use The Pair-wise Wilcoxon rankusm tests in Table 5.4.

Table 5.4: Pair-wise Wilcoxon Ranksum tests for amount sent by each participant in round 1 between treatments for males and females

Round 1		*Common Knowledge*	*Context Neutral*	*Context-loaded* A	*Context-loaded* B
Private Knowledge	Male (N=22)	z=0.89 p=0.37	**z=1.18** **p=0.07**	**z=2.30** **p=0.02**	**z=2.58** **p=0.01**
	Female (N=19)	z=1.33 p=0.19	**z=2.30** **p=0.02**	**z=1.97** **p=0.05**	**z=3.85** **p=0.00**
Common Knowledge	Male (N=25)		z=1.118 p=0.26	**z=1.83** **p=0.07**	**z=2.20** **p=0.03**
	Female (N=14)		z=0.901 p=0.37	z=0.82 p=0.41	**z=2.47** **p=0.01**
Context Neutral	Male (N=18)			z=-0.90 p=0.37	z=1.20 p=0.23
	Female (N=16)			z=-0.46 p=0.65	**z=1.96** **p=0.05**
Context-loaded A	Male (N=24)				z=0.08 p=0.94
	Female (N=17)				z =1.27 p=0.21

In Table 5.4 we present the outcomes of pairwise non-parametric ranksum tests comparing the amount transferred in round 1 only across different treatments for each gender. The results show that for both men and women, the amount transferred in round 1 are significantly higher in all three *context* treatments compared to the *private knowledge* treatment. Men in the *context-loaded* A and *context-loaded* B treatments transferred a higher amount than men in the *common knowledge* treatment. For women, the amount transferred in the *context-loaded* B treatment is significantly higher than that in the *private knowledge*, *common knowledge*, and *context neutral* treatments, but not with that in the *context-loaded* A treatment. Overall, from Table 5.4, we find clear evidence that both men and women

transferred more in round 1 under the three context treatments than in the two control treatments.[38]

[38] As discussed, we used round 1 values for the amount sent because participants are randomly paired for each round, therefore within-session observations are not independent. However, we could compare gender differences in the amount sent between male and female by looking at the average values. We have included the table 5.4A in the Appendix 10.

Table 5.5: Random effects regression results for amount sent disaggregated by gender with standard errors clustered on the subject

Dependent variable: Amount Passed

Amount Passed	(1)	(2)	(3)
Round	**-0.049***	-0.048	**-0.056***
	(0.029)	(0.029)	**(0.031)**
Female *Private knowledge*	-0.343	-0.403	-0.447
	(0.755)	(0.798)	(0.825)
Male *Common knowledge*	**2.022*****	**1.913****	**1.962****
	(0.778)	**(0.866)**	**(0.886)**
Female *Common knowledge*	1.057	1.093	1.022
	(0.845)	(0.861)	(0.860)
Male *Context neutral*	**3.441*****	**3.194*****	**3.119*****
	(0.887)	**(0.958)**	**(0.945)**
Female *Context neutral*	**3.423*****	**3.388*****	**3.354*****
	(0.811)	**(0.872)**	**(0.879)**
Male *Context-loaded* A	**3.277*****	**3.097*****	**3.032*****
	(0.817)	**(0.889)**	**(0.891)**
Female *Context-loaded* A	**1.909****	**1.993****	**1.843****
	(0.901)	**(0.942)**	**(0.934)**
Male *Context-loaded* B	**3.946*****	**3.696*****	**3.499*****
	(0.757)	**(0.878)**	**(0.880)**
Female *Context-loaded* B	**4.101*****	**3.905*****	**3.859*****
	(0.705)	**(0.773)**	**(0.784)**
Lag earning			**0.120*****
			(0.015)
Constant	**2.986*****	**3.276*****	**1.919***
	(0.525)	**(0.980)**	**(1.037)**
Demographic controls	NO	YES	YES
Wald χ^2	80.85	97.95	186.59
Prob > χ^2	0.0000	0.0000	0.0000
Number of observations	1900	1900	1710
Wald test for equality of coefficients			
Male *Common* = Female *Common*	$\chi^2 = 1.11$ p = 0.29	$\chi^2 = 0.87$ p =0.35	$\chi^2 = 1.12$ p = 0.29
Male *Context* N = Female *Context* N	$\chi^2 = 0.00$ p =0.99	$\chi^2 = 0.04$ p = 0.84	$\chi^2 = 0.06$ p =0.81
Male *Context* A = Female *Context* A	$\chi^2 = 1.87$ p =0.17	$\chi^2 = 1.24$ p =0.27	$\chi^2 =1.49$ p =0.22
Male *Context* B = Female *Context* B	$\chi^2 = 0.04$ p =0.84	$\chi^2 = 0.06$ p =0.80	$\chi^2 =0.19$ p =0.67

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

Table 5.5 provides further support for Result 1, that both men and women are more trusting in the *context* treatments than in the control treatments, with the results from random effects regressions with standard errors clustered on the subject. We choose to cluster on the subject level because we only have 4 sessions in each treatment. Nevertheless, we provide the regression results with standard errors clustered on the session level in Appendix 10 (table 5.5A).[39] These latter results are consistent with the standard errors clustered on the subject level.

Model (1) of Table 5.5 includes variables for round and dummies interacting gender with treatments. The reference category here is males in the *private* knowledge treatment. The round variable shows that amount transferred decreases over time. Compared to men in the *private knowledge* treatment, both men and women transferred more in all three *context* treatments. Men transfer more in the *common knowledge* treatment than men in the *private knowledge* treatment, but the result is not significant for women. A Wald test comparing the equality of gender-treatment dummies shows that there is no significant difference between men and women regarding the amount transferred in the *common knowledge* treatment nor in any of the *context* treatments.

Model (2) shows the results of the random effects model with demographic controls. The outcomes with respect to gender treatment interaction variables are qualitatively very similar to the first model. A Wald test fails to reject the null hypothesis which suggests that there is no gender difference in the amount transferred in each of the treatments.

[39] One could argue that the amount send is bounded by zero from below and ten from above, and there is a significant proportion of choices at the lower and upper bounds. For robustness check, we include the random effects Tobit model in the Appendix 10 (Table 5.5B).

Model (3) of Table 5.5 includes earnings from the previous round under the assumption that the previous round earnings may impact the senders' decisions on how much to transfer in the current round. The outcome is consistent with the first two models.

5.4.2 Gender differences in Trust for differences in earnings

Table 5.6 shows the average earning broken up by gender in each treatment. We use non-parametric tests to compare earnings between men and women within each treatment, and the result shows that men on average earn less than women (at a 6% significance level) under the *private knowledge* treatment, but we do not find any significant differences in the other four treatments.

Table 5.6: Average earnings across treatments for proposers

Treatment	Gender	Average earnings
Private knowledge	Male	9
	Female	9.52
Common Knowledge	Male	10.16
	Female	9.99
Context Neutral	Male	11.27
	Female	11.27
Context-loaded A	Male	10.52
	Female	10.27
Context-loaded B	Male	11.57
	Female	10.38

Table 5.7 shows the random effects regressions (with earnings per round as the dependent variable and the same dummies interacting gender with each treatment) with and without demographic controls show both genders earn more in the *common knowledge* treatment and the three *context* treatments than in the reference category of males in the *private knowledge* treatment, but there is no significant difference between the two genders within each treatment.

Table 5.7: Random Effect Model for Earnings across treatments for proposers with standard errors clustered on the subject

Dependent variable: Earnings per round

Earnings per round	(1)	(2)
Round	**-0.084****	**-0.083****
	(0.036)	**(0.036)**
Female *Private knowledge*	0.518	0.507
	(0.373)	(0.383)
Male *Common knowledge*	**1.158*****	**1.107*****
	(0.390)	**(0.387)**
Female *Common knowledge*	**0.983****	**0.971****
	(0.386)	**(0.390)**
Male *Context neutral*	**2.270*****	**2.145*****
	(0.482)	**(0.450)**
Female *Context neutral*	**2.266*****	**2.206*****
	(0.496)	**(0.498)**
Male *Context-loaded* A	**1.518****	**1.419****
	(0.604)	**(0.581)**
Female *Context-loaded* A	**1.268****	**1.327****
	(0.561)	**(0.542)**
Male *Context-loaded* B	**2.564*****	**2.434*****
	(0.661)	**(0.667)**
Female *Context-loaded* B	**1.378*****	**1.258*****
	(0.499)	**(0.479)**
Constant	**9.463*****	**9.828*****
	(0.260)	**(0.540)**
Demographic controls	NO	YES
Wald χ^2	58.282	62.72
Prob > χ^2	0.0000	0.0000
Number of observations	1900	1900
Wald test for equality of coefficients		
Male *Common* = Female *Common*	$\chi^2 = 0.13$ p = 0.72	$\chi^2 = 0.07$ p = 0.79
Male *Context* N = Female *Context* N	$\chi^2 = 0.00$ p =0.99	$\chi^2 = 0.01$ p =0.92
Male *Context* A = Female *Context* A	$\chi^2 = 0.10$ p =0.75	$\chi^2 = 0.01$ p =0.90
Male *Context* B = Female *Context* B	$\chi^2 = 2.25$ p =0.13	$\chi^2 = 2.19$ p =0.14

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

Instead of finding a gender difference in earnings in each treatment, we look at whether men are more likely to identify the payoff-maximizing transfer in this game. Figure 5.3 shows average earnings from each possible amount transferred, which ranges from 0 to 10. We aggregate over all five treatments and look at which amount passed generate a net positive return. For example, senders who sent $1 dollar received back $0.82 back on average. The average earning from sending all $10 is $11.44, which is a return of 14.4%. Figure 5.4 shows the distribution of the amount passed by gender for each of the five treatments.

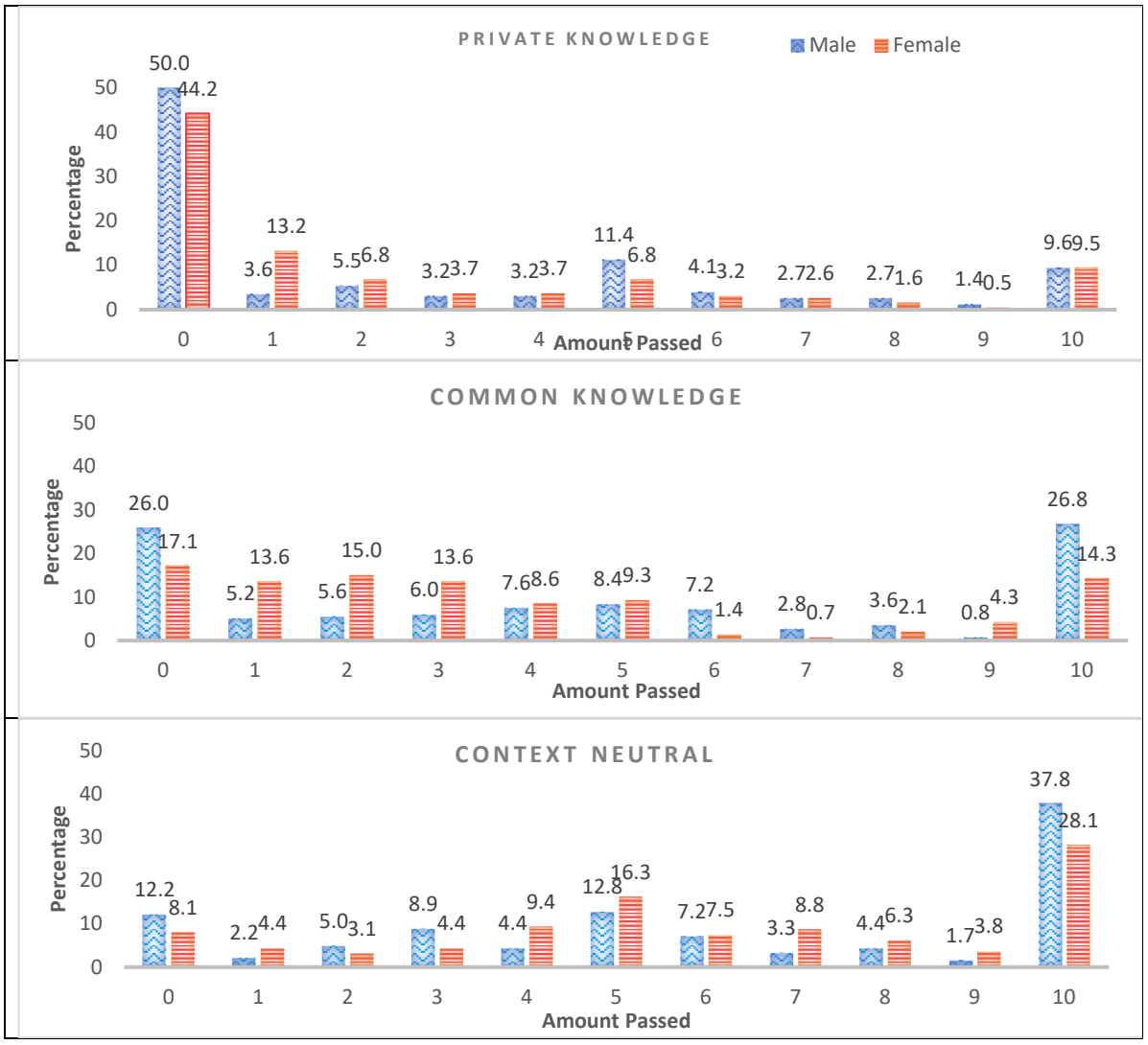
Figure 5.3: Average earnings for each possible amount passed

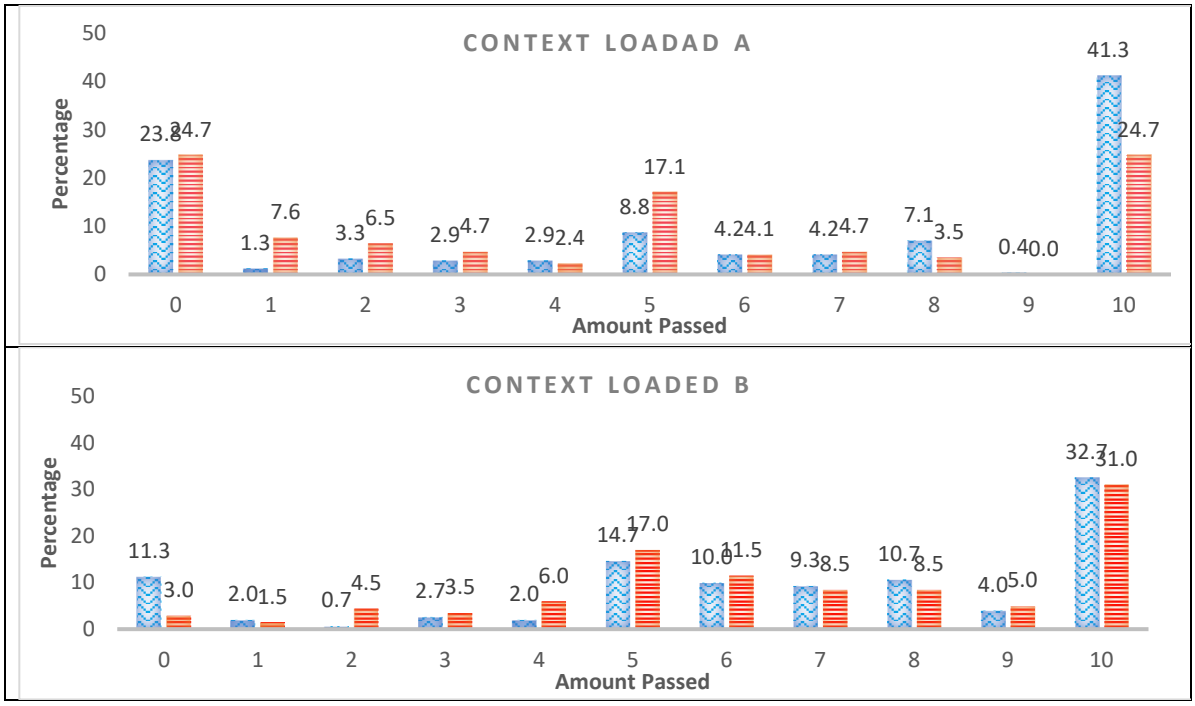
Result 2: Men show a higher propensity to choose the pay-off maximizing transfer.

Given that senders are randomly re-matched with receivers over the ten rounds and make only ten decisions during the course of the game, it is possible that ten rounds do not provide sufficient evidence to learn and find the actual pay-off maximizing transfer. Consequently, in this study, although on average male participants show a higher propensity to trust, we do not find any significant gender differences in earnings, possibly because subjects did not choose the payoff-maximizing transfer amount of $10.

However, the evidence provided in Figures 5.4 suggests that on average, men exhibit a much greater propensity for transferring all $10, which generates the highest net returns. Although in the instructions of the three *context* treatments we have used the choice of $10 as an illustration, it is possible that choice of $10 could be as a result of an experimenter demand effect. In the *common knowledge* treatments where we do not include the additional instruction of using $10 as an example, 27% male participants choose to send all $10 while only 14% female participants have chosen the same but there is almost no gender difference in the *private knowledge* treatment. Nonetheless, it would be interesting to find out if any possible demand effect is more pronounced for males than females.

Figure 5.4: Distribution of amount passed broken up by gender for treatments





5.4.3 Gender differences in Receiver Behaviour

Result 3: Both men and women return more in the context treatments, but we do not have any evidence suggesting that women are more reciprocal than men in context treatments.

Table 5.8 provides a summary statistic for receivers between the two genders across the five treatments. On average, men returned 19% in the *private knowledge* treatment, 39% in the *common knowledge* treatment, 40% in the *context neutral* treatment, 36% in the *context-loaded* A treatment, and 35% in the *context-loaded B* treatment. On average, women returned 31% in the *private knowledge* treatment, 19% in the *common knowledge* treatment, 36% in the *context neutral* treatment, 34% in the *context-loaded* A treatment, and 39% in *the context-loaded* B treatment. Figure 5.5 shows the percentage returned over the ten rounds for males in each treatment, while Figure 5.6 shows the same information for females.

Table 5.8: Average Percentage returned by receivers in each treatment

Percentage Returned	*Private Knowledge*	*Common Knowledge*	*Context Neutral*	*Context-loaded* A	*Context-loaded* B
Total	23.9	31.6	38.2	34.7	36.8
Male	19.0	39.2	40.1	35.5	35.2
Female	30.9	18.8	36.0	33.5	38.5

Figure 5.5: Average percentage returned by male

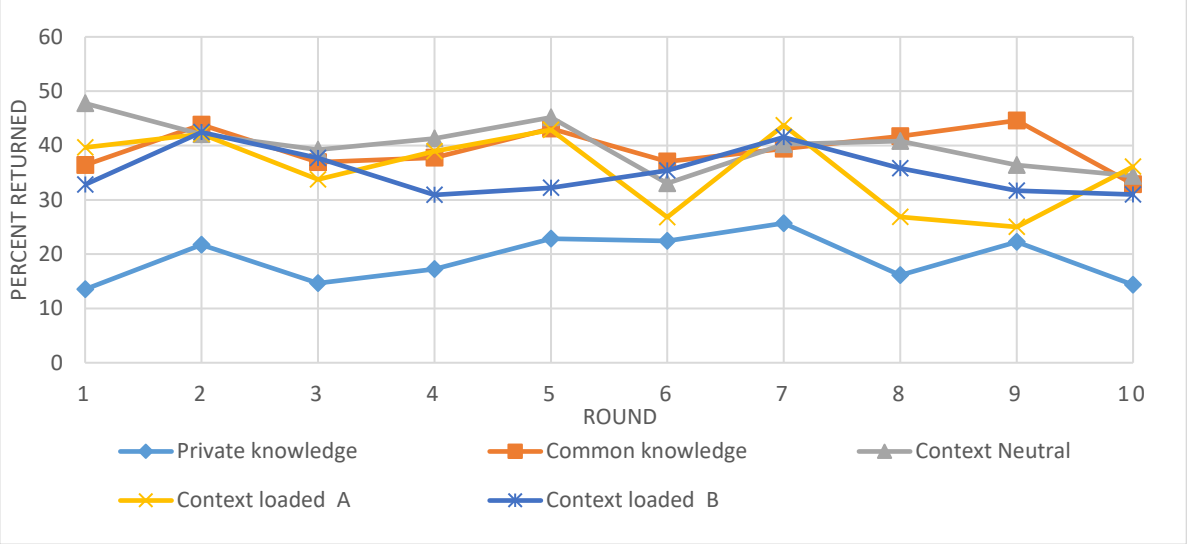
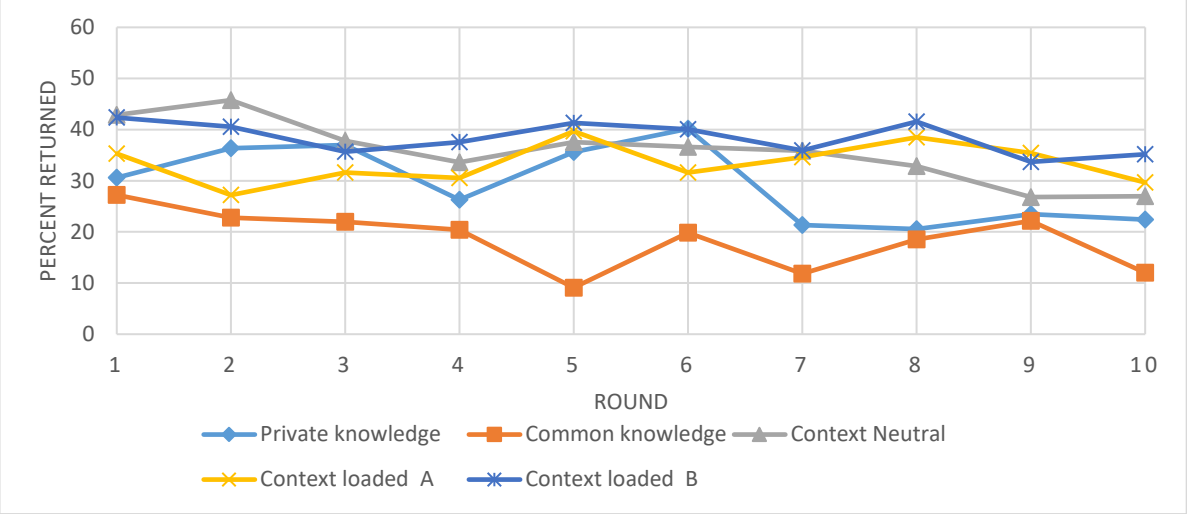


Figure 5.6: Average percentage returned by female



As we did for the amount transferred, we look at pair-wise ranksum tests for the percentage returned in round 1 of each treatment as these are the only independent observations. We exclude observations where the amount returned is zero since there is no decision regarding percentage return could be made by those receivers. Table 5.9 provides the non-parametric ranksum tests for percentage returned by males and females in round 1 across treatments. We have 168 observations in round 1 of the five treatments. There are 18 males and 14 females in the *private knowledge* treatment, 23 males and 12 females in the *common knowledge* treatment, 17 males and 15 females in the *context neutral* treatment, 20 males and 15 females in the *context-loaded A* treatment, and 17 males and 17 females in the *context-loaded B* treatment. Men show a higher degree of reciprocity in the *common knowledge* treatment and in all three *context* treatments than men in the *private knowledge* treatment. Comparing women in different treatments, we only find that women in the *context neutral* treatment show higher returns compared to women in the *private knowledge* or *common knowledge* treatments. We find no evidence suggesting that men and women show a different degree of reciprocity across the three *context* treatments.[40]

[40] We include another table in Appendix 10 using average values.

Table 5.9: Pair-wise Wilcoxon ranksum tests for percentage returned in round 1 by gender between treatments

Round 1		*Common Knowledge*	*Context Neutral*	*Context*-loaded A	*Context*-loaded B
Private Knowledge	Male (N=18)	**z=2.97 p=0.00**	**z=3.84 p=0.00**	**z=3.03 p=0.00**	*z=2.24 p=0.03*
	Female (N=14)	z= -0.26 p= 0.79	*z= 1.69 p= 0.09*	z=0.64 p=0.52	z=1.52 p=0.13
Common Knowledge	Male (N=23)		z= 1.390 p=0.1646	z=-0.52 p=0.61	z= -0.28 p=0.78
	Female (N=12)		*z= 1.98 p=0.05*	z=1.01 p=0.31	z= 1.85 p=0.06
Context Neutral	Male (N=17)			z= -0.72 p=0.47	z= -1.60 p=0.11
	Female (N=15)			z= -1.354 p=0.1757	z=-0.06 p=0.95
Context-loaded A	Male (N=20)				z= -0.77 p=0.44
	Female (N=15)				z =1.20 p= 0.23

Comparing the percentage returned in round 1 by male and female within each treatment using the non-parametric ranksum test, we only find that women returned significantly more than men in the Private knowledge treatment (z=-2.014, p>|z|=0.04, number of male=126, number of female=90), but not in the other four treatments.

In Table 5.10 we use a random effects regression analysis with errors clustered on individuals to compare the percentage returned in the five treatments. The first model only includes variable round and dummies interacting gender and treatment with males in the *private knowledge* being the reference category. Demographic controls are added in the second model, while the third model controls for lagged values of earnings. The coefficient of round in all three models shows that the proportion returned declines over time. The significance level for gender dummies is very similar in all three models.

Both men and women are more reciprocal in all three *context* treatments compared to men in the *private knowledge* treatment and that men in the *common* knowledge treatment are more reciprocal as well. However, there is no significant evidence that suggests other differences. The Wald tests comparing the various coefficients only show that men are more reciprocal than women in the *common* knowledge treatment at a 1% significance level. We do not find any significant differences between men and women in any of the three *context* treatments.

Table 5.10: Random effects regression results for percentage returned broken up by gender

Dependent variable: Percentage returned

Percentage returned	(1)	(2)	(3)
Round	**-0.855*****	**-0.862*****	**-0.796*****
	(0.179)	**(0.180)**	**(0.204)**
Female *Private knowledge*	8.021	7.998	6.039
	(7.402)	(7.087)	(7.086)
Male *Common knowledge*	**16.190****	**16.713*****	**14.942****
	(6.538)	**(6.242)**	**(6.206)**
Female *Common knowledge*	-2.402	-2.370	-4.772
	(6.255)	(6.111)	(6.063)
Male *Context Neutral*	**18.613*****	**17.304*****	**14.887****
	(5.979)	**(6.082)**	**(5.848)**
Female *Context Neutral*	**16.875*****	**16.419****	**14.228****
	(6.523)	**(6.670)**	**(6.540)**
Male *Context loaded* A	**13.737****	**14.450****	**14.201****
	(6.519)	**(6.325)**	**(6.228)**
Female *Context loaded* A	**12.949***	**12.704****	**11.759***
	(6.840)	**(6.358)**	**(6.334)**
Male *Context loaded* B	**14.132****	**13.770***	**13.111***
	(7.166)	**(7.124)**	**(6.821)**
Female *Context loaded* B	**17.847*****	**19.581*****	**17.925*****
	(5.897)	**(5.780)**	**(5.750)**
Lag earning			**-0.356*****
			(0.076)
Constant	**25.169*****	**24.980*****	**34.475*****
	(4.764)	**(6.570)**	**(6.909)**
Demographic controls	NO	YES	YES
Wald χ^2	54.31	66.32	99.45
Prob > χ^2	0.0000	0.0000	0.0000
Number of observations	1460	1460	1270
Wald test for equality of coefficients			
Male *Common* = Female *Common*	**χ^2 = 9.40** **p = 0.00**	**χ^2 = 9.99** **p = 0.00**	**χ^2 = 11.57** **p = 0.00**
Male *Context* N = Female *Context* N	χ^2 = 0.09 p = 0.76	χ^2 = 0.02 p = 0.89	χ^2 = 0.01 p = 0.91
Male *Context* A = Female *Context* A	χ^2 = 0.01 p = 0.91	χ^2 = 0.07 p = 0.79	χ^2 = 0.16 p = 0.69
Male *Context* B = Female *Context* B	χ^2 = 0.34 p = 0.56	χ^2 = 0.81 p = 0.37	χ^2 = 0.65 p = 0.42

*Notes: Standard errors in parentheses; ***, ** and* denote significance at 1%, 5% and 10% respectively.*

5.5 Conclusion

This chapter investigates gender differences in trust and reciprocity. Our findings show that both men and women are more trusting and trustworthy in the *context* treatments than in the *private* knowledge treatment or the *common* knowledge treatment. Looking at the gender differences within our Context treatments could provide a useful comparison given that the gender differences could exist due to strategic uncertainty.

Our study examines a between-subjects experiment with five different treatments to investigate if gender differences will emerge after the underlying incentives are clearly provided to them. Consistent with our hypotheses, both genders show a higher level of trust when the game is explained to them.

We also find that both genders show a higher level of reciprocity in the three *context* treatments than in the two control treatments. We do not find any gender differences within the three *context* treatments. For both proposers and receivers, the level of trust and reciprocity decline over time. We do not observe an increase in the level of trust nor the level of reciprocity from women in the *common* knowledge treatment than in the *private* knowledge treatment. However, we do find that men trust more and reciprocate more in the *common* knowledge treatment than in the *private* knowledge treatment. A possible reason is that men and women do not respond differently when we remove the uncertainty about the nature of the game and how the other participants view the game, but they do respond similarly when we enhance the comprehension of the game.

We find that men exhibit more trust and reciprocate more than women when we remove the uncertainty about the nature of the game and how the other participants view the game. Contrary to many previous studies, we find that men are significantly more reciprocal than women when that uncertainty is eliminated. Although we do not find gender differences

in reciprocal behaviour within our three *context* treatments, the level of trust and reciprocity exhibited by genders in the three *context* treatments are significantly higher than that of men and women in the two control treatments. We do not find any evidence that cognitive demand effects differ by gender. The provision of context does not seem to make a difference, but the strategic uncertainty does play a role in the trust game. When we remove the uncertainty, men are more trusting and trustworthy. Therefore, our results show that there are no significant gender differences in how men and women interpret this game.

6. Conclusion

6.1. Summary

This thesis studies the gender differences in leadership roles from both the supply-side and the demand side of the labour market. Our four primary research questions are: (1) are there systematic gender differences in the willingness to lead; (2) are there differences in the perception of female leaders compared to male leaders; (3) do men and women behave differently, both as employers and as workers; and (4) are there systematic gender differences in cognitive demand effects in a trust game.

These four research questions are addressed in three different experiments. The first experiment looks at our first two research questions using the turnaround game. The third research question is addressed using the principle agent game. The last research question is addressed using the trust game.

We design two experiments in our turnaround game, one has the same payoff matrix for both the leader and the follower, while the other has different payoff matrices for the leader and the follower. We implement four different treatments and within these treatments we manipulate the message type and information regarding the leader's gender. This generates a 2x2 protocol: (1) Leader's gender not revealed; pre-set message; (2) Leader's gender revealed; pre-set message; (3) Leader's gender not revealed; free-form message and (4) Leader's gender revealed; free-form message. Overall, we find that women are less likely to volunteer for leadership positions compared to men, and there are no significant differences in the followers' perception towards male and female leaders.

We use a principal agent game to examine the gender differences in behaviour of both the employer and the employee. We design two treatments, one treatment has fixed pair for the entire game while in the other treatment, participants are randomly re-matched with a

different person in each round. The employer has two contract choices, the "relational contract" is based on mutual trust and reciprocity while the "transactional contract" relies on extrinsic motivations in the form of fines if the worker is found to be shirking. We do not find significant differences between the genders, other than the fact that women tend to offer more generous contract terms.

Our last research question is addressed using the trust game with random re-matching protocol. Our study examines a between-subjects experiment with five different treatments to investigate if gender differences will emerge after the underlying incentives are clearly explained to them. Our main finding is that both genders show a higher level of trust when the game is explained to them. However, we do not find any evidence that cognitive demand effects differ by gender.

6.2. Implications for policy

Our study has important policy implications for gender gaps in the labour market. There is increasing need for diversity programs in both private and public sectors. However, it remains a challenge to promote more women to participate in leadership roles.

Our research focus on gender differences in both the supply side factors and the demand side factors in the labour market. While we do not find significant differences in followers' perception towards male and female leaders, we do find that women are less likely to sort themselves into leadership positions. Consequently, it may not be enough to increase mentoring and training programmes in organizations if women are reluctant to sort themselves into leadership positions in the first place. Women may continue having difficulty in succeeding at the upper levels of organization and government even if there is no biased

perception towards them. Therefore, policies like gender quotas are likely to increase female representation in senior roles.

It is important that women are willing to take senior roles. Despite the mixed findings regarding the effectiveness of gender quotas, our findings do suggest that the leadership gap in the workplace may be partially due to supply-side factors with fewer women volunteering to lead.

6.3. Limitations and Directions for Future Research

Our research findings come from laboratory experiments. Our results are limited by the design of our treatments and the number of observations, which restrains our analyses. One limitation of our study is that we provide a pre-determined message to the leaders in our pre-set message treatment of the turnaround game. A set of messages to the leaders could allow us to study whether there are gender differences in the choice made by leaders. Our intension is to analyse if the follower's perception towards the leader is different,

Another limitation is that all our participants are university students who have not experienced real leadership roles in organizations and government. Their leadership styles or strategies may not reflect the real-world situation. Future research could usefully be carried out using field experiments.

It is entirely possible that the lack of resistance toward female leaders is due to the fact that this study was carried out in New Zealand; the first country to endow women with the right to vote back in 1893. New Zealand also has a long history of strong female leaders including three female Prime Ministers in recent times, and as noted above, the gender wage gap in New Zealand is small and lower than the OECD average. However, while this may explain why we do not find evidence of backlash against female leadership, it is harder to

explain why female participants are so much more reluctant to volunteer for leadership roles and why this reluctance is exacerbated in the gender revealed treatments. It seems difficult to explain this in any way other than to appeal to internalized norms whereby women anticipate greater resistance based on their socialization and lived experiences. So, a potential lesson of this study is that the gender gap in leadership may arise due to a greater female reluctance to assume leadership roles but to the extent that these results may extend to other countries and cultures. It is possible that female leaders are over-estimating the degree of resistance to female leadership.

Appendix 1. Instructions for Chapter 3

A1.1. General instructions

Instructions

General instructions

Welcome. The University of Auckland has provided funding in order to conduct this research. The instructions are simple. If you follow them closely and make appropriate decisions, you may make an appreciable amount of money. For this experiment all earnings are denoted in experimental dollars. At the end of the session you will be paid your earnings in cash at the rate of 400 experimental dollars = NZ $1. This money is in addition to the $5 show-up fee that you get.

The experiment will be conducted using computers. Please do not talk at any point during the experiment. If you have any questions then please raise your hand and one of us will come to you to answer it.

Specific instructions

This experiment consists two parts. In each part, there will be 10 rounds. We will give you the instructions to the first set of ten rounds now. We will give you the instructions to the next set of ten rounds at the end of the first ten rounds. At the completion of the 20 rounds of the game, we will ask you fill out two questionnaires.

Instructions for Part 1: first 10 rounds

One way to think about this experiment is that you are an employee of an organization. You will be part of a group consisting of 5 employees, that is you and four other employees. However, you will not learn the identity of the other people in your group in any round. The composition of these groups will remain unchanged for the entire time. This means that you will be interacting with the same four other people during the whole experiment.

In each round every employee will decide how many hours to work. The number of hours you may choose to work are {0, 10, 20, 30, or 40}. Your earnings in each round will depend on the number of hours that you pick and the smallest number of hours chosen by any participant in your group, including your own choice of hours.

Table 1 below tells you the potential payoffs you may receive. The earnings in each round may be found by looking across from the value you choose on the left hand side of the table and down from the smallest value chosen by any participant from the top of the table. For example, if you choose to work 20 hours while the minimum number of hours chosen in your group is 10 (that is, there is at least one person who has chosen to work for 10 hours and no one has chosen 0 hours), then you will earn 400 experimental dollars for that round.

If you choose to work 30 hours and the smallest value chosen is 20 then you will earn 450 experimental dollars for that round. If you choose to work 40 hours and the smallest value chosen is 0, then you will earn 200 experimental dollars for that round.

Table 1:

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Your choice of hours	0	400	400	400	400	400
	10	350	450	450	450	450
	20	300	400	500	500	500
	30	250	350	450	550	550
	40	200	300	400	500	600

Notice: If someone chooses a particular value X as the number of hours and the others in the group all choose to work for at least X hours or more, then the smallest number of work hours chosen in the group is X. For example, if the number of hours chosen by the five members of a group are 0, 10, 20, 30 and 40, then the minimum number of hours is 0. Alternatively, if the number of hours chosen by the five members of a group are 10, 20, 20, 30 and 30, then the minimum number of hours is 10.

As noted above, the experiment will consist of 20 rounds. After we have finished reading the instructions you will proceed to play the first 10 rounds of this game.

We will stop at the end of the 10^{th} round. At that point we will give you further instructions about how the next ten rounds (Rounds 11 – 20) will work.

Please do NOT continue on to the 11^{th} round of this game until asked by the experimenter to do so.

We will pay you your earnings from the experiment at the end of the session. You are free to go once you have been paid. Your earnings are private information and we encourage you to keep this information private. If at any point you have any questions or problems, please raise your hand and ask for assistance.

A1.2. Gender Revealed Pre-set Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you are chosen as the leader, then the experimenter will provide you with a message that you can send to the other members of the group. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. More importantly, the Leader is now paid differently. The leader's payoff now depends on the hours of work chosen by the leader and also the minimum number of hours chosen by someone in the group, *including the leader*. This is shown in the payoff table below – Table 2. The larger the minimum number of hours chosen in the group, the higher is the leader's payoff.

The payoff table for the employees remains unchanged and is shown in Table 1.

Table 2: The leader's payoff is summarized in the following table:

Leader's Payoff	
Minimum hours chosen by employee	Earnings
0	100
10	300
20	500
30	700
40	900

But if the leader's choice of hours exceeds that of the workers, then there is an additional cost to the leader of the form Cost = 2(Leader's choice of hours – Minimum hours chosen by someone in the group).

For instance, if the leader chooses 20 hours and the minimum chosen is also 20 hours, that is, no one chooses to work for less than 20 hours then the difference between the leader's choice and the minimum choice is zero. In this case, the leader earns 500 – 2(20-20) = 500. On the other hand, if the leader chooses 40 hours while the minimum hours chosen in the group is 0, then the leaders' payoff is 100– 2(40-0) = 100 – 80 = 20.

For your convenience, Table 3 incorporates the information in Table 2 plus any additional cost incurred when leader's choice is larger than the minimum number of hours chosen in the group.

Table 3:

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Leader's choice of hours	0	100	100	100	100	100
	10	80	300	300	300	300
	20	60	280	500	500	500
	30	40	260	480	700	700
	40	20	240	460	680	900

Once the leader has been chosen and leader's message has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or not. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours, the leader's gender and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.3. Gender Not Revealed Pre-set Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you are chosen as the leader, then the experimenter will provide you with a message that you can send to the other members of the group. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer. If no one has volunteered to be the leader then the leader will also be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. More importantly, the Leader is now paid differently. The leader's payoff now depends on the hours of work chosen by the leader and also the minimum number of hours chosen by someone in the group, *including the leader*. This is shown in the payoff table below – Table 2. The larger the minimum number of hours chosen in the group, the higher is the leader's payoff.

The payoff table for the employees remains unchanged and is shown in Table 1.

Table 2: The leader's payoff is summarized in the following table:

Leader's Payoff	
Minimum hours chosen by employee	Earnings
0	100
10	300
20	500
30	700
40	900

But if the leader's choice of hours exceeds that of the workers, then there is an additional cost to the leader of the form Cost = 2(Leader's choice of hours – Minimum hours chosen by someone in the group).

For instance, if the leader chooses 20 hours and the minimum chosen is also 20 hours, that is, no one chooses to work for less than 20 hours then the difference between the leader's choice and the minimum choice is zero. In this case, the leader earns 500 – 2(20-20) = 500. On the other hand, if the leader chooses 40 hours while the minimum hours chosen in the group is 0, then the leaders' payoff is 100– 2(40-0) = 100 – 80 = 20.

For your convenience, Table 3 incorporates the information in Table 2 plus any additional cost incurred when leader's choice is larger than the minimum number of hours chosen in the group.

Table 3:

Leader's choice of hours	Minimum of hours chosen by other members of the group				
	0	10	20	30	40
0	100	100	100	100	100
10	80	300	300	300	300
20	60	280	500	500	500
30	40	260	480	700	700
40	20	240	460	680	900

Once the leader has been chosen and leader's message has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or not. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.4. Gender Revealed Free-form Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you choose to be a leader by saying "Yes" then you will be asked to type a message. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer. If no one has volunteered to be the leader then the leader will also be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. More importantly, the Leader is now paid differently. The leader's payoff now depends on the hours of work chosen by the leader and also the minimum number of hours chosen by someone in the group, *including the leader*. This is shown in the payoff table below – Table 2. The larger the minimum number of hours chosen in the group, the higher is the leader's payoff.

The payoff table for the employees remains unchanged and is shown in Table 1.

Table 2: The leader's payoff is summarized in the following table:

Leader's Payoff	
Minimum hours chosen by employee	Earnings
0	100
10	300
20	500
30	700
40	900

But if the leader's choice of hours exceeds that of the workers, then there is an additional cost to the leader of the form Cost = 2(Leader's choice of hours – Minimum hours chosen by someone in the group).

For instance, if the leader chooses 20 hours and the minimum chosen is also 20 hours, that is, no one chooses to work for less than 20 hours then the difference between the leader's choice and the minimum choice is zero. In this case, the leader earns 500 – 2(20-20) = 500. On the other hand, if the leader chooses 40 hours while the minimum hours chosen in the group is 0, then the leaders' payoff is 100– 2(40-0) = 100 – 80 = 20.

For your convenience, Table 3 incorporates the information in Table 2 plus any additional cost incurred when leader's choice is larger than the minimum number of hours chosen in the group.

Table 3:

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Leader's choice of hours	0	100	100	100	100	100
	10	80	300	300	300	300
	20	60	280	500	500	500
	30	40	260	480	700	700
	40	20	240	460	680	900

Once the leader has been chosen and the message written by the leader has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or write a different message. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours, the leader's gender and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.5. Gender Not Revealed Free-form Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you choose to be a leader by saying "Yes" then you will be asked to type a message. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer. If no one has volunteered to be the leader then the leader will also be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. More importantly, the Leader is now paid differently. The leader's payoff now depends on the hours of work chosen by the leader and also the minimum number of hours chosen by someone in the group, *including the leader*. This is shown in the payoff table below – Table 2. The larger the minimum number of hours chosen in the group, the higher is the leader's payoff.

The payoff table for the employees remains unchanged and is shown in Table 1.

Table 2: The leader's payoff is summarized in the following table:

Leader's Payoff	
Minimum hours chosen by employee	Earnings
0	100
10	300
20	500
30	700
40	900

But if the leader's choice of hours exceeds that of the workers, then there is an additional cost to the leader of the form Cost = 2(Leader's choice of hours – Minimum hours chosen by someone in the group).

For instance, if the leader chooses 20 hours and the minimum chosen is also 20 hours, that is, no one chooses to work for less than 20 hours then the difference between the leader's choice and the minimum choice is zero. In this case, the leader earns 500 – 2(20-20) = 500. On the other hand, if the leader chooses 40 hours while the minimum hours chosen in the group is 0, then the leaders' payoff is 100– 2(40-0) = 100 – 80 = 20.

For your convenience, Table 3 incorporates the information in Table 2 plus any additional cost incurred when leader's choice is larger than the minimum number of hours chosen in the group.

Table 3:

		Minimum of hours chosen by other members of the group				
		0	10	20	30	40
Leader's hours	0	100	100	100	100	100
	10	80	300	300	300	300
	20	60	280	500	500	500
	30	40	260	480	700	700
	40	20	240	460	680	900

Once the leader has been chosen and the message written by the leader has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or write a different message. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.6. General instructions for the Same Payoff treatment

Instructions

General instructions

Welcome. The University of Auckland has provided funding in order to conduct this research. The instructions are simple. If you follow them closely and make appropriate decisions, you may make an appreciable amount of money. For this experiment all earnings are denoted in experimental dollars. At the end of the session you will be paid your earnings in cash at the rate of 400 experimental dollars = NZ $1. This money is in addition to the $5 show-up fee that you get.

The experiment will be conducted using computers. Please do not talk at any point during the experiment. If you have any questions then please raise your hand and one of us will come to you to answer it.

Specific instructions

This experiment consists two parts. In each part, there will be 10 rounds. We will give you the instructions to the first set of ten rounds now. We will give you the instructions to the next set of ten rounds at the end of the first ten rounds. At the completion of the 20 rounds of the game, we will ask you fill out two questionnaires.

Instructions for Part 1: first 10 rounds

One way to think about this experiment is that you are an employee of an organization. You will be part of a group consisting of 5 employees, that is you and four other employees. However, you will not learn the identity of the other people in your group in any round. The composition of these groups will remain unchanged for the entire time. This means that you will be interacting with the same four other people during the whole experiment.

In each round every employee will decide how many hours to work. The number of hours you may choose to work are {0, 10, 20, 30, or 40}. Your earnings in each round will depend on the number of hours that you pick and the smallest number of hours chosen by any participant in your group, including your own choice of hours.

Table 1 below tells you the potential payoffs you may receive. The earnings in each round may be found by looking across from the value you choose on the left hand side of the table and down from the smallest value chosen by any participant from the top of the table. For example, if you choose to work 20 hours while the minimum number of hours chosen in your group is 10 (that is, there is at least one person who has chosen to work for 10 hours and no one has chosen 0 hours), then you will earn 400 experimental dollars for that round. If you choose to work 30 hours and the smallest value chosen is 20 then you will earn 450 experimental dollars for that round. If you choose to work 40 hours and the smallest value chosen is 0, then you will earn 200 experimental dollars for that round.

Table 1:

Your choice of hours	Minimum of hours chosen by other members of the group				
	0	10	20	30	40
0	400	400	400	400	400
10	350	450	450	450	450
20	300	400	500	500	500
30	250	350	450	550	550
40	200	300	400	500	600

Notice: If someone chooses a particular value X as the number of hours and the others in the group all choose to work for at least X hours or more, then the smallest number of work hours chosen in the group is X. For example, if the number of hours chosen by the five members of a group are 0, 10, 20, 30 and 40, then the minimum number of hours is 0. Alternatively, if the number of hours chosen by the five members of a group are 10, 20, 20, 30 and 30, then the minimum number of hours is 10.

As noted above, the experiment will consist of 20 rounds. After we have finished reading the instructions you will proceed to play the first 10 rounds of this game.

We will stop at the end of the 10th round. At that point we will give you further instructions about how the next ten rounds (Rounds 11 – 20) will work.

Please do NOT continue on to the 11th round of this game until asked by the experimenter to do so.

We will pay you your earnings from the experiment at the end of the session. You are free to go once you have been paid. Your earnings are private information and we encourage you to keep this information private. If at any point you have any questions or problems, please raise your hand and ask for assistance.

A1.7. Gender Revealed Pre-set Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you are chosen as the leader, then the experimenter will provide you with a message that you can send to the other members of the group. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. The payoff table for both leaders and employees remain unchanged and as shown in Table 1.

Once the leader has been chosen and leader's message has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or not. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours, the leader's gender and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.8. Gender Revealed Free-form Message treatment

Instructions for Part 2: Rounds 11 through 20

Before round 11 starts, you have a chance to volunteer to be a leader for your group. You will be asked if you wish to be a leader and you will have to respond by choosing "Yes" or "No." If you choose to be a leader by saying "Yes" then you will be asked to type a message. This message will be shown to the other 4 group members. If more than one member of your group volunteer to be the leader then the leader will be chosen at random by the computer. If no one has volunteered to be the leader then the leader will also be chosen at random by the computer.

There are still five people in each group, except now four of them are employees and one of them is the leader. The payoff table for leaders and employees remain unchanged as shown in Table 1.

Once the leader has been chosen and the message written by the leader has been shared with the four employees, then the game proceeds as follows. In each round the leader moves first by choosing the number of hours. The leader can also choose to re-send the message sent before or write a different message. Once the leader has chosen the number of hours, the information regarding the leader's choice of hours, the leader's gender and the content of the leader's message will be revealed to the group members. The employees will then choose their hours simultaneously, i.e., each employee will make a choice without knowing the choice of the other employees.

Your earnings will be denoted in experimental dollars as before.

Your $5 show up fee and earnings for the first 10 rounds are unaffected.

Are there any questions? If at any point you have any questions or problems, please raise your hand and ask for assistance.

We will tell you when we are ready to proceed.

A1.9. Post-experiment questionnaire for leader's perception

Participant ID:

Please answer ALL of the questions below by selecting <u>one</u> of the <u>five</u> options listed below each question. Please put a circle around your preferred answer.

Question 1: Females do not possess good leadership quality.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Question 2: Because leadership is viewed as a masculine trait, females will not be viewed as strong leaders

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Question 3: Females are too emotional to lead effectively

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Question 4: Females are capable of performing effectively in any leadership position

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Question 5: Male leaders connect with the public better than female leaders

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Appendix 2. Instructions for Chapter 4

A2.1. Random re-matching treatment

Instructions for the Experiment

WELCOME.

This is a study in economic decision making. The University of Auckland has provided funding to conduct this research. If you follow the instructions and make appropriate decisions you might earn a considerable amount of money. The experiment will be conducted using computers. In a minute we will give you the instructions for logging in to the experiment.

Please do not talk at any point during the experiment. If you have any questions then please raise your hand and one of us will come to you to answer it.

You will take part in two separate experiments. We will provide the instruction for experiment 1 first. Instructions for experiment 2 will be provided after experiment 1 is completed.

General instructions for experiment 1

The experiment will consist of a number of rounds. In each round you will be matched with another person. The person you are matched with will change from one round to the next, so that you are not playing with the same person for more than one round at a time. The decisions you and the other person make will determine your earnings. One person in each matched pair will have the role of an "employer" who fixes a wage, and the other will be a "worker" who makes an effort decision, after seeing the employer's wage for that round.

You will be randomly assigned to the role of either an employer or a worker. Your role will remain unchanged for all rounds. The employer will begin each round by selecting a wage between (and including) $0.01 and $10.00. The worker will see the wage offered by the employer and then choose an effort that may be any number between (and including) 1 and 10.

Effort is costly for the worker but generates value for the employer. The following table shows these values and costs:

Worker Effort Level:	1	2	3	4	5	6	7	8	9	10
Employer Value of Effort:	$1.00	$2.00	$3.00	$4.00	$5.00	$6.00	$7.00	$8.00	$9.00	$10.00
Worker Cost of Effort	$0.01	$0.10	$0.20	$0.40	$0.60	$0.80	$1.00	$1.30	$1.60	$2.00

Specific instructions for experiment 1

In each round, the employer must decide whether to specify a **bonus** or a **penalty**:

- **Bonus**: This is a payment that is proposed in the event that the effort is no lower than the suggested effort level. The proposed bonus is **non-binding** in the sense that the actual bonus is selected after the worker's effort is observed by the employer, and it may be any amount between 0.00 and 10, irrespective of the worker's effort, and irrespective of what was promised.
- **Penalty**: If this option is selected, the employer must decide whether or not to incur a monitoring cost of $1.00 to be able to specify a penalty that must be paid by the worker if the effort falls short of the suggested level **AND** if this shortcoming can be verified by a third party. The verification process is random, and verification can only be accomplished with probability of 0.33; otherwise no penalty is paid.
  Note: even though the worker's effort is observed by the employer, the penalty cannot be collected unless the effort is verified by a third party.

In each round, the employer will begin by offering a proposed employment contract which specifies:

- ***Wage***, a fixed payment between (and including) $0.01 and $10.00, which covers the worker's cost associated with the suggested effort level.
- A ***Suggested Effort*** level made to the worker, which is between 1 and 10.
- Either a ***Proposed Bonus***, a payment between $0.00 and $10.00 to the worker that is specified in advance. The actual payment, however, is selected after the worker's effort decision is revealed and the employer can observe whether or not the suggested effort level is met. The **actual** bonus payment may be any amount between $0.00 and $10.00, and may be greater than, equal to, or less than the **proposed** bonus amount.
- Or a ***Penalty Provision,*** a pre-specified payment between $0.00 and $1.30. By incurring a monitoring cost of $1.00, the employer can specify a penalty that may be assessed if a third party can verify that the worker's effort falls short of the suggested effort level. Note that this will only be implemented if the worker's effort is below the suggested level and non-compliance can be verified by a third party, which occurs with probability 0.33.

Employer Contract Earnings: If the contract is accepted, the employer earns the difference between the **value of effort** and the **wage** paid to the worker. If the employer specified a bonus, then it is subtracted from the employer's earnings. If the employer specified a penalty, then a monitoring cost of $1.00 is subtracted, and the penalty amount is only added to the employer's earnings if the effort shortcoming can be verified by a third party, which occurs with probability of 0.33.

Worker Contract Earnings: If the contract is accepted, the worker earns the difference between **wage** and the cost of that worker's effort. In addition, the penalty payment (if any) is subtracted from the worker's earnings or the bonus (if any) is added to the worker's earnings.

If the contract is not accepted by the worker, both parties have contract earnings of $0.00.

The actual experiment consists of ten rounds where you can earn money. But before starting the actual game, we will ask you to play 4 practise rounds to get you familiar with the game. Once the

actual game starts, the computer program will keep track of your total earnings for all ten rounds, and these will be shown as "cumulative earnings" on a results page. However, the earnings from the practise rounds will be shown on the screen but they will not be added to your total earnings. This total earning is in addition to the $5.00 show-up fee.

The employers will be provided with an additional endowment of $10.00 at the start of this experiment. This amount will be added to the earnings of the employers at the end of the experiment. This is in addition to the show-up fee of $5.00 paid to every participant.

Below we provide some **numerical examples**:

Example 1:

The employer offers a wage of $5.00, suggests an effort level of 9 and incurs the monitoring cost of $1.00 in order to specify a penalty of $1.30. This penalty must be paid by the worker to the employer if the worker's effort falls below the suggest level (of 9) and this fact is verified by a third party (this occurs with probability of 0.33). The worker accepts the contract and puts in the suggest effort level of 9. The earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \$9.00 - \$5.00 - \$1.00 = \$3.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \$5.00 - \$1.60 = \$3.40$$

Example 2:

The employer offers a wage of $2.00, suggests an effort level of 6 and incurs the monitoring cost of $1.00 in order to specify a penalty of $1.30. This penalty must be paid by the worker to the employer if the worker's effort falls below the suggest level (of 6) and this fact is verified by a third party (this occurs with probability of 0.33). The worker accepts the contract but puts in an effort level of 4 (which is below the suggest effort of 6).

If the effort shortcomings are verified by a third party (this occurs with a probability of 0.33), the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost + penalty$$

$$= \$4.00 - \$2.00 - \$1.00 + \$1.30 = \$2.30$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort - penalty$$

$$= \$2.00 - \$0.40 - \$1.30 = \$0.30$$

If the shortcomings are NOT verified by a third party, the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \$4.00 - \$2.00 - \$1.00 = \$1.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \$2.00 - \$0.40 = \$1.60$$

Example 3:

The employer offers a wage of $5.00, suggests an effort level of 9 and proposes a bonus of $2.00. The worker accepts the contract and puts in the effort level of 9. By observing the worker's effort level of 9, the employer then decides to pay the actual bonus of $2.00. The earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \$9.00 - \$5.00 - \$2.00 = \$2.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \$5.00 - \$1.60 + \$2.00 = \$5.40$$

On the other hand, if the employer only pays a bonus of $1, then the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \$9.00 - \$5.00 - \$1.00 = \$3.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \$5.00 - \$1.60 + \$1.00 = \$4.40$$

You will be able to read the specific instructions once more after you log in to the computer. Before we start the practise rounds, we will ask you to answer some questions to make sure that you understand how the game works. Please do not proceed to round 1 till instructed to do so.

Practise Questions

Question 1

Employer asks for 5 units of output, offers wage of $2, and invests $1.00 in monitoring technology in order to specify a penalty of $1.30. Worker accepts the contract and puts an effort level of 3.

(a) If the shortcomings are NOT verified by a third party, the earnings in this situation will be:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

(b) If the effort shortcomings are verified by a third party (this occurs with a probability of 0.33):

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost + penalty$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort - penalty$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

Question 2

Employer asks for 8 units of effort, offers wage of $3, and proposes a bonus of $2.00. The worker accepts the contract and puts an effort level of 8.

(a) The employer then pays an actual bonus of $2.50.

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

(b) If employer pays an actual bonus of $1.00:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

A2.2. Fixed matching treatment

Instructions for the Experiment

WELCOME.

This is a study in economic decision making. The University of Auckland has provided funding to conduct this research. If you follow the instructions and make appropriate decisions you might earn a considerable amount of money. The experiment will be conducted using computers. In a minute we will give you the instructions for logging in to the experiment.

Please do not talk at any point during the experiment. If you have any questions then please raise your hand and one of us will come to you to answer it.

You will take part in two separate experiments. We will provide the instruction for experiment 1 first. Instructions for experiment 2 will be provided after experiment 1 is completed.

General instructions for experiment 1

The experiment will consist of a number of rounds. You will be matched with the same person in all rounds. The decisions you and the other person make will determine your earnings. One person in each matched pair will have the role of an "employer" who fixes a wage, and the other will be a "worker" who makes an effort decision, after seeing the employer's wage for that round.

You will be randomly assigned to the role of either an employer or a worker. Your role will remain unchanged for all rounds. The employer will begin each round by selecting a wage between (and including) $0.01 and $10.00. The worker will see the wage offered by the employer and then choose an effort that may be any number between (and including) 1 and 10.

Effort is costly for the worker but generates value for the employer. The following table shows these values and costs:

Worker Effort Level:	1	2	3	4	5	6	7	8	9	10
Employer Value of Effort:	$1.00	$2.00	$3.00	$4.00	$5.00	$6.00	$7.00	$8.00	$9.00	$10.00
Worker Cost of Effort	$0.01	$0.10	$0.20	$0.40	$0.60	$0.80	$1.00	$1.30	$1.60	$2.00

Specific instructions for experiment 1

In each round, the employer must decide whether to specify a **bonus** or a **penalty**:

- **Bonus**: This is a payment that is proposed in the event that the effort is no lower than the suggested effort level. The proposed bonus is **non-binding** in the sense that the actual bonus is selected after the worker's effort is observed by the employer, and it may be any amount between 0.00 and 10, irrespective of the worker's effort, and irrespective of what was promised.
- **Penalty**: If this option is selected, the employer must decide whether or not to incur a monitoring cost of $1.00 to be able to specify a penalty that must be paid by the worker if the effort falls short of the suggested level **AND** if this shortcoming can be verified by a third party. The verification process is random, and verification can only be accomplished with probability of 0.33; otherwise no penalty is paid.
  Note: even though the worker's effort is observed by the employer, the penalty cannot be collected unless the effort is verified by a third party.

In each round, the employer will begin by offering a proposed employment contract which specifies:

- ***Wage***, a fixed payment between (and including) $0.01 and $10.00, which covers the worker's cost associated with the suggested effort level.
- A ***Suggested Effort*** level made to the worker, which is between 1 and 10.
- Either a ***Proposed Bonus***, a payment between $0.00 and $10.00 to the worker that is specified in advance. The actual payment, however, is selected after the worker's effort decision is revealed and the employer can observe whether or not the suggested effort level is met. The **actual** bonus payment may be any amount between $0.00 and $10.00, and may be greater than, equal to, or less than the **proposed** bonus amount.
- Or a ***Penalty Provision,*** a pre-specified payment between $0.00 and $1.30. By incurring a monitoring cost of $1.00, the employer can specify a penalty that may be assessed if a third party can verify that the worker's effort falls short of the suggested effort level. Note that this will only be implemented if the worker's effort is below the suggested level and non-compliance can be verified by a third party, which occurs with probability 0.33.

Employer Contract Earnings: If the contract is accepted, the employer earns the difference between the **value of effort** and the **wage** paid to the worker. If the employer specified a bonus, then it is subtracted from the employer's earnings. If the employer specified a penalty, then a monitoring cost of $1.00 is subtracted, and the penalty amount is only added to the employer's earnings if the effort shortcoming can be verified by a third party, which occurs with probability of 0.33.

Worker Contract Earnings: If the contract is accepted, the worker earns the difference between **wage** and the cost of that worker's effort. In addition, the penalty payment (if any) is subtracted from the worker's earnings or the bonus (if any) is added to the worker's earnings.

If the contract is not accepted by the worker, both parties have contract earnings of $0.00.

The actual experiment consists of ten rounds where you can earn money. But before starting the actual game, we will ask you to play 4 practise rounds to get you familiar with the game. Once the actual game starts, the computer program will keep track of your total earnings for all ten rounds, and these will be shown as "cumulative earnings" on a results page. However, the earnings from the practise rounds will be shown on the screen but will not be added to your total earnings. This total earning is in addition to the $5.00 show-up fee.

The employers will be provided with an additional endowment of $10.00 at the start of this experiment. This amount will be added to the earnings of the employers at the end of the experiment. This is in addition to the show-up fee of $5.00 paid to every participant.

Below we provide some **numerical examples**:

Example 1:

The employer offers a wage of $5.00, suggests an effort level of 9 and incurs the monitoring cost of $1.00 in order to specify a penalty of $1.30. This penalty must be paid by the worker to the employer if the worker's effort falls below the suggest level (of 9) and this fact is verified by a third party (this occurs with probability of 0.33). The worker accepts the contract and puts in the suggest effort level of 9. The earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \$9.00 - \$5.00 - \$1.00 = \$3.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \$5.00 - \$1.60 = \$3.40$$

Example 2:

The employer offers a wage of $2.00, suggests an effort level of 6 and incurs the monitoring cost of $1.00 in order to specify a penalty of $1.30. This penalty must be paid by the worker to the employer if the worker's effort falls below the suggest level (of 6) and this fact is verified by a third party (this occurs with probability of 0.33). The worker accepts the contract but puts in an effort level of 4 (which is below the suggest effort of 6).

If the effort shortcomings are verified by a third party (this occurs with a probability of 0.33), the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost + penalty$$

$$= \$4.00 - \$2.00 - \$1.00 + \$1.30 = \$2.30$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort - penalty$$

$$= \$2.00 - \$0.40 - \$1.30 = \$0.30$$

If the shortcomings are NOT verified by a third party, the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \$4.00 - \$2.00 - \$1.00 = \$1.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \$2.00 - \$0.40 = \$1.60$$

Example 3:

The employer offers a wage of $5.00, suggests an effort level of 9 and proposes a bonus of $2.00. The worker accepts the contract and puts in the effort level of 9. By observing the worker's effort level of 9, the employer then decides to pay the actual bonus of $2.00. The earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \$9.00 - \$5.00 - \$2.00 = \$2.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \$5.00 - \$1.60 + \$2.00 = \$5.40$$

On the other hand, if the employer only pays a bonus of $1, then the earnings in this situation will be as follows:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \$9.00 - \$5.00 - \$1.00 = \$3.00$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \$5.00 - \$1.60 + \$1.00 = \$4.40$$

You will be able to read the specific instructions once more after you log in to the computer. Before we start the practise rounds, we will ask you to answer some questions to make sure that you understand how the game works. Please do not proceed to round 1 till instructed to do so.

Practise Questions

Question 1

Employer asks for 5 units of output, offers wage of $2, and invests $1.00 in monitoring technology in order to specify a penalty of $1.30. Worker accepts the contract and puts an effort level of 3.

(c) If the shortcomings are NOT verified by a third party, the earnings in this situation will be:

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort$$

$$= \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

(d) If the effort shortcomings are verified by a third party (this occurs with a probability of 0.33):

$$Employer\ earnings = employer\ value\ of\ effort - wage - monitoring\ cost + penalty$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort - penalty$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

Question 2

Employer asks for 8 units of effort, offers wage of $3, and proposes a bonus of $2.00. The worker accepts the contract and puts an effort level of 8.

(c) The employer then pays an actual bonus of $2.50.

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

(d) If employer pays an actual bonus of $1.00:

$$Employer\ earnings = employer\ value\ of\ effort - wage - bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ - \_\_\_\_\_$$

$$= \_\_\_\_\_$$

$$Worker\ earnings = wage - worker\ cost\ of\ effort + bonus$$

$$= \_\_\_\_\_ - \_\_\_\_\_ + \_\_\_\_\_$$

$$= \_\_\_\_\_$$

A2.3. Login Instructions for Veconlab

Login instructions

- Login to the computer (using your username and password).
- Check that you are logged in to your Net Account.
- Open Internet Explorer.
- Enter the following web address and press enter:

  http://veconlab.econ.virginia.edu/login.htm.
- The "Veconlab Participant Login Screen" screen should be displayed.

  Click on 'Login'.
- The 'Veconlab: Enter Session Name' screen should be displayed.

  Enter the Session Name: aicXX. Click on 'Submit'. (*You will be told what two numbers to enter in the place of XX. This will be written down on the white-board in from the lab.*)
- The 'Veconlab Participant Login' screen should be displayed.

  Fill in the boxes. Click on 'Continue'.
- The computer will assign you a Participant ID Number. Please write down your ID number and Password at the top of the page of your instructions in the space provided. It is important that you remember the password! This password will help us to go back and retrieve your data should you happen to close the browser window by mistake during the session.
- Please follow the instructions displayed on screen.

Appendix 3. Instructions for the Chapter 5

Participant ID:
Password:

Instructions

General Instructions

Welcome. The University of Auckland has provided funding in order to conduct this research. The instructions are simple. If you follow them closely and make appropriate decisions, you may make an appreciable amount of money. For this experiment all earnings are denoted in experimental dollars. At the end of the session you will be paid your earnings in cash at the rate of 10 experimental dollars = NZ $1. This money is in addition to the $5 show-up fee that you get.

The experiment will be conducted using computers. In a minute we will give you the instructions for logging in to the experiment.

Please do not talk at any point during the experiment. If you have any questions then please raise your hand and one of us will come to you to answer it.

Specific instructions

This experiment consists of ten rounds. In each round, each of you will be matched with another participant in the room. The person that you are matched with will change from one round to the next, i.e., you will not be matched with the same person for all rounds. You will not learn the identity of the person you are matched with at any time.

One of you will be designated as the first mover and the other one as the second mover. The decisions that you and the other person make will determine the amounts earned by each of you.

At the beginning of each round both the first mover and the second mover will have $10.00.

The first mover will decide how much money (if any) out of his $10.00 to pass on to the second mover and how much (if any) to keep. Any amount passed gets tripled (multiplied by 3) before it is received by the second mover, who then decides how much (if any) to keep and how much (if any) to pass back to the first mover.

The first mover earns the amount kept initially (out of his $10.00) plus any amount that the second mover passes back.

The second mover keeps any amount not returned. (This is in addition to the $10.00 given to the second mover at the beginning of the round). Any amount returned by the second mover is not tripled. The game ends after the second mover's decision.

You will be able to read the specific instructions once more after you log in to the computer.

Context Neutral Treatment—additional Instruction included:

One way to think about this situation is as follows: the second mover has no incentive to send any money back to the first mover because the round ends immediately after that. Anticipating that, the first mover should hang on to his $10.00 and send nothing to the second mover. This means they will both end the round with $10.00 each.

But suppose the first mover decides to transfer $10.00 to the second mover. Then the second mover will get $30.00. If the second mover sends back an amount more than $10.00 then it is easy to see that both the first mover and the second mover can make more money than if they simply hung on to their $10.00 in each and every round.

Context loaded A Treatment—additional Instruction included:

One way to think about this situation is as follows: the second mover has no incentive to send any money back to the first mover because the round ends immediately after that. Anticipating that, the first mover should hang on to his $10.00 and send nothing to the second mover. This means they will both end the round with $10.00 each.

But suppose the first mover decides to trust the second mover by sending $10.00. Then the second mover will get $30.00. If the second mover behaves in a trustworthy manner and sends back an amount more than $10.00 then it is easy to see that both the first mover and the second mover can make more money than if they simply hung on to their $10.00 in each and every round.

Context loaded B Treatment—additional Instruction included:

One way to think about this situation is as follows: suppose the first mover decides to trust the second mover by sending $10.00. Then the second mover will get $30.00. If the second mover behaves in a trustworthy manner and sends back an amount more than $10.00 then it is easy to see that both the first mover and the second mover can make more money than if they simply hung on to their $10.00 in each and every round.

But the second mover has no incentive to send any money back to the first mover because the round ends immediately after that. Anticipating that, the first mover should hang on to his $10.00 and send nothing to the second mover. This means they will both end the round with $10.00 each.

Appendix 4. Post-questionnaire—Holt-Laury Task /Experiment 2

This experiment consists of ten paired lottery-choice, you are required to choose between option A and option B for each of these ten pairs.

The following table shows these options and their probabilities and payoffs:

Lottery number	Option A	Option B	YOUR CHOICE
1	1/10 of \$6.00, 9/10 of \$4.80	1/10 of \$11.55, 9/10 of \$0.30	
2	2/10 of \$6.00, 8/10 of \$4.80	2/10 of \$11.55, 8/10 of \$0.30	
3	3/10 of \$6.00, 7/10 of \$4.80	3/10 of \$11.55, 7/10 of \$0.30	
4	4/10 of \$6.00, 6/10 of \$4.80	4/10 of \$11.55, 6/10 of \$0.30	
5	5/10 of \$6.00, 5/10 of \$4.80	5/10 of \$11.55, 5/10 of \$0.30	
6	6/10 of \$6.00, 4/10 of \$4.80	6/10 of \$11.55, 4/10 of \$0.30	
7	7/10 of \$6.00, 3/10 of \$4.80	7/10 of \$11.55, 3/10 of \$0.30	
8	8/10 of \$6.00, 2/10 of \$4.80	8/10 of \$11.55, 2/10 of \$0.30	
9	9/10 of \$6.00, 1/10 of \$4.80	9/10 of \$11.55, 1/10 of \$0.30	
10	10/10 of \$6.00, 0/10 of \$4.80	10/10 of \$11.55, 0/10 of \$0.30	

Please choose either Option A or Option B for those 10 lottery pairs. The computer will randomly select a lottery number after everyone has made their decisions.

For example, if your die outcome is 6 and you choose Option A, then your earning is:

6	**6/10 of \$6.00, 4/10 of \$4.80**	6/10 of \$11.55, 4/10 of \$0.30	A

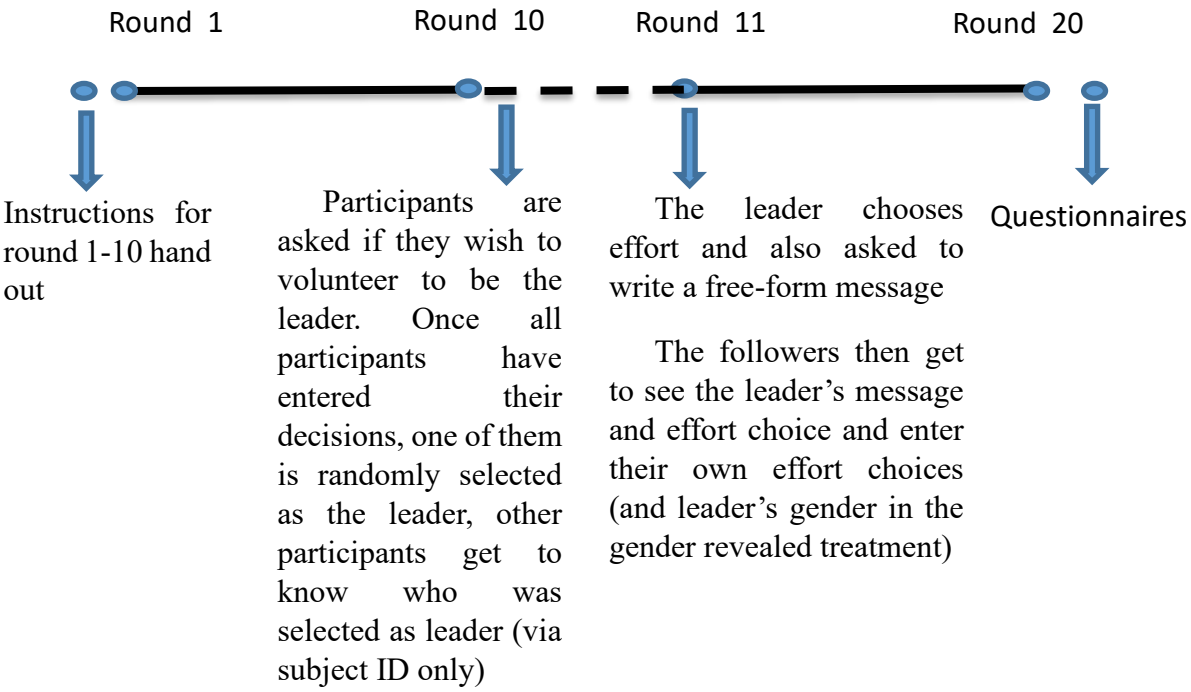
$$6/10 * \$6.00 \ + \ 4/10 * \$4.80 = \$3.60 + \$1.92 = \$5.52$$

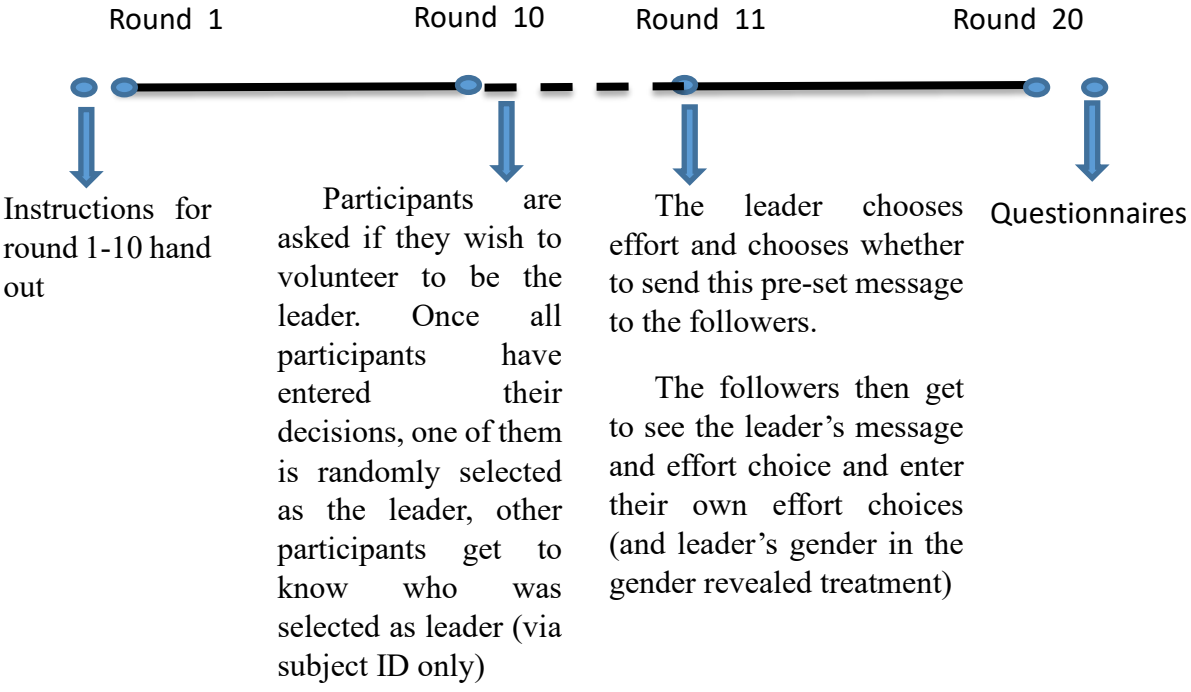
If your die outcome is 2 and you choose Option B, then your earning is:

2	2/10 of \$6.00, 8/10 of \$4.80	**2/10 of \$11.55, 8/10 of \$0.30**	B

$$2/10 * \$11.55 \ + \ 8/10 * \$0.30 = \$2.31 + \$0.24 = \$2.55$$

Appendix 5. Additional information—Chapter 3





Appendix 6. Additional information—Chapter 4

We define e^s as the effort level for the worker when the worker shirks, and e^{ns} as the effort level for the worker when the worker does not shirk. Earning for the employer is denoted as π^s when the worker shirks, and as π^{ns} when the worker does not shirk. We further define u^s as the payoff to the worker then the worker shirks, and u^{ns} as the payoff to the worker when the worker does not shirk. Recall that the verification process is random and with a probability p.

$$\pi^{ns} = V(e^{ns}) - w - k \quad if \quad e \geq e^* \tag{3}$$

$$\pi^s = p[V(e^s) - w + f] + (1-p)[V(e^s) - w] - k \quad if \quad e \geq e^*$$

$$\text{Or } \pi^s = V(e^s) - w - k + pf \quad if \quad e < e^* \tag{4}$$

$$u^{ns} = w - C(e^{ns}) \quad if \quad e \geq e^* \tag{5}$$

$$u^s = p[w - C(e^s) - f] + (1-p)[w - C(e^s)] \quad if \quad e < e^*$$

$$\text{Or } u^s = w - C(e^s) - pf \quad if \quad e < e^* \tag{6}$$

Applying the same assumption of monetary payoff maximization, we would expect, *ex-post*, worker would only provide non-minimal effort if the employer imposes a fine such that the expected utility gained from not shirking is greater than or equal to the utility of shirking, i.e. $u^{ns} \geq u^s$. Consequently, we can get the following non-shirking condition from equation (5) and (6):

$$pf \geq C(e^{ns}) - C(e^s) \tag{7}$$

The non-shirking condition expresses that the worker will choose not to shirk if the cost of providing the extra effort, $C(e^{ns}) - C(e^s)$, is less than or equal to the expected fine, pf. We can get the upper bond on enforceable effort $e^{ns} = C^{-1}[pf + C(e^s)]$ by rearranging

the non-shirking condition. Therefore, given the parameters in our experiment, the employer will impose the maximum amount of fine, $\bar{f} = 1.3$ by incurring a verification cost, $k = 1$, and demand an effort level of $e^{ns} = 4$[41]. Subsequently, the employer will offer the minimum wage to the worker in a way that the worker would be indifferent between accepting and rejecting the contract, i.e. $w = C(4) = 0.4$. Under this optimal contract case, employer will earn $\pi = V(e) - w - k = 4 - 0.4 - 1 = \$2.60$ and worker will earn $u = w - C(e) = 0.4 - 0.4 = \mathbf{0}$.

[41] $e^{ns} = C^{-1}[(0.3)(1.3) + 0.01] = C^{-1}(0.4) = 4$

Appendix 7. Additional information—Chapter 5

Study	Design	Instruction Type	Trust	Reciprocity
Ashraf et al. (2006)	One-shot / Continuous	W & R	M = F	F = M
Bellemare and Kroger (2007)	One-shot / Continuous	W	M = F	F = M
Bohnet (2007)	One-shot / Continuous	W	M = F	F < M
Buchan et al. (2008)	One-shot / Continuous	W	M > F	F > M
Chaudhuri and Sbai (2011)	Repeated ten rounds with random re-matching	W	M = F	F > M
Chaudhuri and Gangadharan (2007)	One-shot / Continuous / Switch roles	W & R	M > F	F = M
Clark and Sefton (2001)	One-shot / Continuous	W & R	M = F	F = M
Cox (2002)	One-shot / Continuous	W & R	M = F	F < M
Cox and Deck (2006)	One-shot / Discrete	W	na	F < M, F= M or F > M (dependent on context)
Croson and Buchan (1999)	One-shot / Continuous	W	M = F	F > M
Eckel and Wilson (2004)	One-shot / Semi-Discrete	W & R	M > F Written information F > M Photo	F = M F = M
Eckel and Wilson (2001)	One-shot / Continuous	W	M = F	na
Innocenti and Pazienza (2006)	One-shot / Continuous	W & R	M > F	F = M
Schwieren and Sutter (2004)	One-shot / Continuous	Unclear	M = F	F > M
Kanagaretnam et al. (2005)	One-shot / Continuous/ switch roles	W & R	M = F	F = M
Slonim (2004)	One-shot / Continuous	Unclear	M = F With selection M > F Without selection	F = M F = M

Appendix 8. Post-experiment: Survey Questionnaire

Please answer ALL of the questions on this survey as accurately as you can.

1. What is your Age? _____

2. What is your Gender? _____

3. What is your field of study at the University? _____

4. Are you an Undergraduate Student (which year) or a Postgraduate Student? Circle one.

   UG (Year _____)

   PG

5. What is your after-tax weekly income from ALL SOURCES (including salary, allowances & scholarships)? Circle one of the options below.

   Less than $250

   Between $250 and $750

   Between $750 and $1,250

   More than $1,250

6. Were you born in New Zealand? YES NO

7. Which ethnic group do you belong to? Circle one of the options below.

   New Zealand European

   Maori

   Samoan

   Cook Island Maori

   Niuean

   Chinese

   Indian Subcontinent (including Pakistan and Bangladesh)

   Other (Please specify) _____

Appendix 9. Additional Results—Chapter 3

A9.1. Table 3.A6

Table 3.A6: Random effects ordered probit model for follower's effort in Experiment 2 (cluster on the group)

Dependent variable: Choice of effort level by follower

Effort Choice	Pre-set	Pre-set	Free-form	Free-form
Round	-0.594***	-1.013***	-0.030	0.012
	(0.107)	(0.156)	(0.058)	(0.072)
Female	-4.694**	-12.999***	-0.320	2.109*
	(1.974)	(2.891)	(1.728)	(1.230)
Female*Round	0.555***	1.011***	0.051	0.034
	(0.130)	(0.156)	(0.100)	(0.106)
Female Leader	-0.259	0.279	-1.087	-8.236***
	(0.446)	(0.296)	(0.722)	(0.911)
Lag Earning	0.001	-0.003*	0.008***	0.007**
	(0.002)	(0.002)	(0.001)	(0.003)
Risk averse	-1.105*	-1.208**	-0.026	2.625***
	(0.589)	(0.531)	(0.461)	(0.328)
Demographic Control	NO	YES	NO	YES
Number of observations	270	270	220	220
Wald χ^2	2232.72	-	157.51	-
Prob > χ^2	0.000	-	0.000	-

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

Notes: In experiment 2, the majority of followers choose effort of 40, therefore the random effects ordered probit model is not a good fit. This table is included to counter-part Table 3.6 that reports the result for Experiment 1 in the chapter

*A9.2. Table 3.A6**

Table 3.A6*: Random effects model for follower's effort in Experiment 2 (cluster on the group)

Dependent variable: Choice of effort level by follower

Effort Choice	Pre-set	Pre-set	Free-form	Free-form
Round	-1.046***	-1.048***	-0.045	-0.017
	(0.250)	(0.257)	(0.040)	(0.060)
Female	-12.066***	-11.965***	1.089	1.220
	(3.311)	(3.595)	(4.022)	(3.825)
Female*Round	0.874***	0.883***	0.007	0.009
	(0.224)	(0.231)	(0.186)	(0.186)
Female Leader	-0.010	-0.128	-1.840	-1.442
	(0.644)	(0.508)	(1.403)	(0.917)
Lag Earning	0.022	0.021***	0.020***	0.018***
	(0.005)	(0.005)	(0.007)	(0.007)
Risk averse	-1.527	-2.270*	1.120	2.222
	(1.240)	(1.234)	(1.340)	(2.405)
Demographic Control	NO	YES	NO	YES
Number of observations	270	270	220	220
Wald χ^2	126.70	-	1414.69	-
Prob > χ^2	0.000	-	0.000	-

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

A9.3. Table 3.A7

Table 3.A7: Random Effects model for effort level under Pre-set message treatment in Experiment 2 (cluster on the group

Follower's Effort Choice	(1)	(2)	(3)	(4)
Round	-0.707***	-0.865**	-0.795***	-0.800***
	(0.269)	(0.350)	(0.283)	(0.293)
Female*Round	0.538***	0.674***	0.627***	0.639***
	(0.199)	(0.262)	(0.217)	(0.228)
Male follower with female leader	0.527	0.678	0.500	0.565
	(0.558)	(0.589)	(0.381)	(0.632)
Female follower with male leader	-6.913***	-8.817**	-8.345***	-8.021**
	(2.676)	(3.652)	(3.165)	(3.455)
Female follower with female leader	-7.021**	-9.120**	-8.581***	-8.312**
	(2.730)	(3.842)	(3.257)	(4.097)
Message shown	-	-	1.331	1.412
	-	-	(1.855)	(1.960)
Leaders Effort	0.335**	0.343***	0.365***	0.352***
	(0.140)	(0.121)	(0.103)	(0.112)
Lag minimum effort	0.097**	0.123**	0.117**	0.113**
	(0.048)	(0.062)	(0.054)	(0.052)
Risk averse	-	-	-1.065	**-1.618*** **
	-	-	(0.772)	**(0.628)**
Constant	32.620***	34.099***	31.462***	31.645***
	(7.619)	(7.978)	(6.120)	(7.090)
Demographic Controls	NO	NO	NO	YES
Wald χ^2	411.699	6088.546	-	-
Prob > χ^2	0.000	0.000	-	-
Number of observations	360	270	270	270
Wald test for equality of coefficients				
Male follower with female leader = Female follower with male leader	Wald χ^2=6.82 Prob > χ^2=0.01	Wald χ^2=5.90 Prob > χ^2=0.02	Wald χ^2=7.13 Prob > χ^2=0.01	Wald χ^2=6.80 Prob > χ^2=0.01
Male follower with female leader = Female follower with female leader	Wald χ^2=6.76 Prob > χ^2=0.01	Wald χ^2=5.77 Prob > χ^2=0.02	Wald χ^2=7.25 Prob > χ^2=0.01	Wald χ^2=5.20 Prob > χ^2=0.02
Female follower with male leader = Female follower with female leader	Wald χ^2=0.66 Prob > χ^2=0.42	Wald χ^2=0.33 Prob > χ^2=0.57	Wald χ^2=0.10 Prob > χ^2=0.75	Wald χ^2=0.09 Prob > χ^2=0.77

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

A9.4. Table 3.A8

Table 3.A8: Random Effects model for effort level under Free-form message treatment in Experiment 2

Follower's Effort Choice	(1)	(2)	(3)	(4)	(3)	(4)
Round	-0.170	-0.187	-0.046	-0.170**	-0.187*	-0.046
	(0.120)	(0.123)	(0.123)	(0.081)	(0.105)	(0.080)
Female*Round	-0.224	-0.244	-0.029	-0.224	-0.244	-0.029
	(0.158)	(0.157)	(0.165)	(0.310)	(0.320)	(0.170)
Male follower with female leader	-1.163	-1.568	-1.817	-1.163	-1.568	-1.817
	(0.708)	(0.975)	(1.214)	(1.035)	(1.622)	(1.463)
Female follower with male leader	3.810	4.123	1.418	3.810	4.123	1.418
	(2.686)	(2.672)	(3.219)	(4.851)	(5.005)	(3.407)
Female follower with female leader	2.554	2.602	-0.187	2.554	2.602	-0.187
	(2.539)	(2.569)	(2.722)	(4.518)	(4.915)	(3.695)
Leaders Effort	1.454***	1.389***	1.266***	1.454***	1.389***	1.266***
	(0.208)	(0.206)	(0.253)	(0.184)	(0.200)	(0.098)
Lag minimum effort	0.221***	0.211***	0.122***	0.221***	0.211***	0.122**
	(0.023)	(0.024)	(0.026)	(0.083)	(0.080)	(0.060)
Authoritarian Message Style	-	-1.296	-0.720	-	-1.296	-0.720
	-	(0.945)	(0.961)	-	(1.375)	(1.084)
Democratic Message Style	-	-1.617**	-1.139	-	-1.617	-1.139
	-	(0.708)	(0.843)	-	(1.583)	(1.615)
Laissez-Faire Message Style	-	-0.204	-0.660	-	-0.204	-0.660
	-	(0.955)	(0.989)	-	(0.918)	(1.063)
Risk averse	-	-	0.930	-	-	0.930
	-	-	(0.799)	-	-	(1.167)
Constant	-23.892***	-19.370**	-14.574	-23.892***	-19.370**	-14.574***
	(8.491)	(8.556)	(10.433)	(5.565)	(7.980)	(3.323)
Demographic Controls	NO	NO	NO	YES	YES	YES
Wald χ^2	178.528	179.173	71.621	-	-	-
Prob > χ^2	0.000	0.000	0.000	-	-	-
Number of observations	320	320	220	320	320	220

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

Notes: There were very few choices low effort choices in this treatment in Experiment 2. As a result, we could not run a random effects ordered probit regression. Therefore, we have chosen to report results from a random effects regression instead.

A9.5. Table 3.A9

Table 3.A9: Random Effects Model for follower's round earnings in Experiment 2 Pre-set Message treatment (clustered on the group)

Earnings	Pre-set	Pre-set	Free-form	Free-form
Round	-14.057***	-13.771***	-2.895	-1.991
	(5.377)	(4.926)	(2.077)	(2.151)
Female	-101.771**	-129.919**	-24.028	-4.512
	(50.985)	(56.826)	(24.984)	(16.012)
Female Leader	10.786	5.074	-7.176	-11.577
	(9.750)	(7.312)	(11.919)	(12.459)
Female*Round	7.601**	9.610**	1.023	0.346
	(3.659)	(4.025)	(1.214)	(0.868)
Leader's Effort	9.481***	9.231***	3.440	2.747
	(1.302)	(1.351)	(2.880)	(1.824)
Lag Minimum Effort	2.018**	1.927**	2.507**	1.643*
	(0.976)	(0.917)	(1.128)	(0.915)
Risk Averse	-	-1.476	-	-2.298
	-	(6.555)	-	(2.766)
Constant	336.368***	348.933***	409.076***	456.839***
	(94.971)	(95.688)	(108.984)	(75.124)
Wald χ^2	2911.998	10893.538	901.805	2359.411
Prob > χ^2	0.000	0.000	0.000	0.000
Number of observations	360	270	320	220

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively.*

Appendix 10. Additional Results—Chapter 5

A10.1. Table 5.4A

Table 5.4A: Pair-wise Wilcoxon Ranksum tests for the average amount sent by each participant across 10 rounds between treatments for males and females

Average		*Common Knowledge*	*Context Neutral*	*Context-* loaded A	*Context-* loaded B
Private Knowledge	Male (N=22)	***z= 2.21*** ***p= 0.03***	**z= 3.24** **p= 0.00**	**z= 3.41** **p= 0.00**	**z= 3.76** **p= 0.00**
	Female (N=19)	**z= 1.73** **p= 0.08**	**z= 3.38** **p= 0.00**	**z= 2.14** **p= 0.03**	**z= 4.25** **p= 0.00**
Common Knowledge	Male (N=25)		z= 1.44 p= 0.15	z= 1.39 p= 0.16	**z= 2.08** **p= 0.04**
	Female (N=14)		**z= 2.25** **p= 0.02**	z= 0.60 p= 0.55	**z= 2.91** **p= 0.00**
Context Neutral	Male (N=18)			z= -0.04 p= 0.97	z= 0.36 p= 0.72
	Female (N=16)			z= -1.48 p= 0.14	z= 0.69 p= 0.49
Context- loaded A	Male (N=24)				z= 0.55 p= 0.58
	Female (N=17)				**z= 2.17** **p= 0.03**

A10.2. Table 5.5A

Table 5.5A: Random effects regression results for amount sent disaggregated by gender with standard errors clustered on the session

Amount Passed	(1)	(2)	(3)
Round	-0.049	-0.048	-0.056
	(0.037)	(0.036)	(0.042)
Female Private knowledge	-0.343	-0.403	-0.447
	(1.252)	(1.422)	(1.454)
Male Common knowledge	2.022**	1.913*	1.962*
	(0.931)	(1.154)	(1.155)
Female Common knowledge	1.057	1.093	1.022
	(1.307)	(1.344)	(1.349)
Male *Context* N	3.441***	3.194***	3.119***
	(0.871)	(1.061)	(1.057)
Female *Context* N	3.423***	3.388**	3.354**
	(1.265)	(1.369)	(1.392)
Male *Context* A	3.277***	3.097**	3.032**
	(1.150)	(1.310)	(1.224)
Female *Context* A	1.909**	1.993**	1.843*
	(0.852)	(1.000)	(0.999)
Male *Context* B	3.946***	3.696***	3.499***
	(0.906)	(1.055)	(1.077)
Female *Context* B	4.101***	3.905***	3.859***
	(0.971)	(1.108)	(1.116)
Lag earning		-0.195	-0.054
		(0.519)	(0.530)
Constant		0.664	0.773
		(0.601)	(0.588)
Demographic controls	NO	YES	YES
Wald χ^2	242.786	783.097	1343.491
Prob > χ^2	0.0000	0.0000	0.0000
Number of observations	1900	1900	1710
Wald test for equality of coefficients			
Male Common = Female Common	$\chi^2 = 0.99$ p = 0.32	$\chi^2 = 0.78$ p =0.38	$\chi^2 = 0.93$ p = 0.34
Male *Context* N = Female *Context* N	$\chi^2 = 0.00$ p =0.99	$\chi^2 = 0.02$ p = 0.89	$\chi^2 = 0.03$ p =0.86
Male *Context* A = Female *Context* A	$\chi^2 = 1.30$ p =0.25	$\chi^2 = 0.84$ p =0.36	$\chi^2 =1.25$ p =0.26
Male *Context* B = Female *Context* B	$\chi^2 = 0.03$ p =0.86	$\chi^2 = 0.06$ p =0.81	$\chi^2 =0.13$ p =0.71

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively*

A10.3. Table 5.5B

Table 5.5B: Random Effects Tobit model with bootstrap standard errors clustering on the subject

Amount Passed	(1)	(2)	(3)
Round	-0.134**	-0.129***	-0.146**
	(0.056)	(0.042)	(0.065)
Female Private knowledge	-0.293	-0.746	-0.959
	(1.595)	(1.782)	(2.231)
Male Common knowledge	4.057***	4.166***	4.746**
	(1.532)	(1.455)	(2.365)
Female Common knowledge	2.827**	2.689	2.689
	(1.128)	(1.837)	(1.821)
Male *Context* N	6.721***	5.933***	6.173***
	(1.737)	(1.581)	(1.947)
Female *Context* N	6.328***	5.870***	6.272***
	(1.916)	(2.179)	(2.371)
Male *Context* A	6.690***	6.149***	6.215***
	(1.431)	(1.996)	(2.156)
Female *Context* A	3.808**	3.366*	3.383
	(1.600)	(1.986)	(2.322)
Male *Context* B	7.246***	6.412***	6.190***
	(1.583)	(2.383)	(2.231)
Female *Context* B	7.866***	7.185***	7.274***
	(1.324)	(1.500)	(2.107)
Lag earning			0.244***
			(0.029)
Constant	1.503*	3.327**	1.088
	(0.867)	(1.554)	(1.871)
Demographic controls	NO	YES	YES
Log likelihood	-3497.5836	-3490.079	-3006.831
Wald χ^2	84.72	794.03	645.96
Prob > χ^2	0.0000	0.0000	0.0000
Number of observations	1900	1900	1710
Number left censored	440	440	418
Number uncensored	969	969	843
Number right censored	491	491	449
Wald test for equality of coefficients			
Male Common = Female Common	$\chi^2 = 0.56$ p = 0.46	$\chi^2 = 0.86$ p =0.35	$\chi^2 = 1.68$ p = 0.19
Male *Context* N = Female *Context* N	$\chi^2 = 0.04$ p =0.84	$\chi^2 = 0.00$ p = 0.98	$\chi^2 = 0.00$ p =0.97
Male *Context* A = Female *Context* A	$\chi^2 = 2.07$ p =0.15	$\chi^2 = 1.73$ p =0.19	$\chi^2 =2.07$ p =0.16
Male *Context* B = Female *Context* B	$\chi^2 = 0.19$ p =0.66	$\chi^2 = 0.24$ p =0.63	$\chi^2 =0.35$ p =0.56

*Notes: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% respectively*

A10.4. Table 5.9A

Table 5.9A: Pair-wise Wilcoxon ranksum tests for average percentage returned by gender between treatments

Average		*Common Knowledge*	*Context Neutral*	*Context-* loaded A	*Context-* loaded B
Private Knowledge	Male (N=18)	**z=2.78 p=0.01**	**z=3.12 p=0.00**	**z=2.59 p=0.01**	*z=1.85 p=0.06*
	Female (N=14)	z=-0.95 p=0.34	z=1.18 p=0.24	z=0.48 p=0.63	z=1.43 p=0.15
Common Knowledge	Male (N=23)		z= 1.27 p=0.20	z=-0.32 p=0.75	z= -0.06 p=0.96
	Female (N=12)		**z= 2.56 p=0.01**	*z= 1.88 p=0.06*	**z= 3.01 p=0.00**
Context Neutral	Male (N=17)			z= -0.58 p=0.56	z= -0.98 p=0.33
	Female (N=15)			z= -0.54 p=0.59	z=0.13 p=0.89
Context- loaded A	Male (N=20)				z= - 0.24 p=0.81
	Female (N=15)				z =0.77 p= 0.44

7. References

Algan, Y., & Cahuc, P. (2010). Inherited Trust and Growth. *The American Economic Review, 100*, 2060-2092.

Altonji, J. G., & Blank, R. M. (1999). Race and Gender in the Labor Market. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3C, pp. 3143-3259). New York: Elsevier.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society. *Review of Economics and Statistics, 95*(4), 1438-1443.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2018). On the cultural basis of gender differences in negotiation. *Experimental Economics, 21*, 757-778.

Arch, E. C. (1993). Risk-Taking: A motivational Basis for Sex Differences. *Psychological Reports, 73*(3), 6-11.

Babcock, L., & Laschever, S. (2003). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton, NJ: Princeton University Press.

Babcock, L., Pecalde, M., Vesterlund , L., & Weingart, L. (2017). Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *American Economic Review, 107(3)*, 714-47.

Bangun, L., Chaudhuri, A., Prak, P., & Zhou, C. (2006). Common and almost common knowledge of credible assignments in a coordination game. *Economics Bulletin, 3*, 1-10.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior, 10*, 122-142.

Bertrand, M. (2011). New Perspectives on Gender. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4B, pp. 1545-1592). New York: Elsevier.

Bertrand, M., Goldin, C., & Katz, L. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics , 2*, 228-55.

Blau, F., & Kahn, L. (2000). Gender Differences in Pay. *Journal of Economic Perspectives, 14*, 75-99.

Blau, F., & Kahn, L. (2006). The U.S. Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review, 60*, 45-66.

Blau, F., & Kahn, L. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature, 55 (3)*, 789-865.

Bohnet, I. (2016). *What works: gender equality by design.* Cambridge: Harward University Press.

Bohnet, I., van Green, A., & Bazerman, M. (2016). When Performance Trumps Gender Bias: Joint vs. Separate Evaluation. *Management Science, 62*(5), 1225-1234.

Bolton, G. E., & Katok, E. (1995). An Experimental Test for Gender Differences in Beneficent Behavior. *Economics Letters, 48*(3-4), 287-292.

Brands, R., Menges, J., & Kilduff, M. (2015). The Leader-in-Social-Network Schema: Perceptions of Network Structure Affect Gendered Attributions of Charisma. *Organization Science, 26*(4), 941-1261.

Brandts, J., & Cooper, D. (2006). A Change Would Do You Good ... An Experimental Study on How to Overcome Coordination Failure in Organizations. *American Economic Association, 96*(3), 669-693.

Brandts, J., & Cooper, D. (2007). It's What You Say, Not What You Pay: An Experimental Study of Manager-Employee Relationships in Overcoming Coordination Failure. *Journal of the European Economic Association, 5*(6), 1223-1268.

Brinig, M. F. (1995). Does Mediation Systematically Disadvantage Women? *William & Mary Journal of Women and the Law, 2*(1), 1-34.

Brown, A. (2014, August 7). *Perceptions about women bosses improve, but gap remains*. Retrieved from Pew Research Center: http://www.pewresearch.org/fact-tank/2014/08/07/perceptions-about-women-leaders-improve-but-gap-remains/

Brown, S., Gray, D., McHardy, J., & Taylor, K. (2015). Employee trust and workplace performance. *Journal of Economic Behavior & Organization, 116*, 361-378.

Buchan, N., Croson, R., & Solnick, S. (2008). Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game. *Journal of Economic Behavior & Organization, 68*(3-4), 466-476.

Burns, J. M. (1978). *Leadership*. New York: Harper and Row.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Carpenter, J. P. (2007a). The Demand for Punishment. *Journal of Economic Behavior & Organization, 62*(4), 522-542.

Carpenter, J. P. (2007b). Punishing Free-Riders: How Group Size affects Mutual Monitoring and the Provision of Public Goods. *Games and Economic Behavior, 60*(1), 31-51.

Carter, N., & Silva, C. (2010, May 26). *Pipeline's Broken Promise*. Retrieved from Catalyst: https://www.catalyst.org/wp-content/uploads/2019/01/Pipelines_Broken_Promise_Final_021710.pdf

Catalyst. (2018, July 30). *Women in Management*. Retrieved from Catalyst: http://www.catalyst.org/knowledge/women-management#footnote22_g8owtq6

Charness, G., & Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior and Organization, 83*(1), 50-58.

Charness, G., & Kuhn, P. (2011). Lab Labor: What Can Labor Economists Learn from the Lab? In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4A, pp. 229-330). New York: Elsevier.

Chaudhuri, A. (2011). Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature. *Experimental Economics, 14*(1), 47-83.

Chaudhuri, A., Cruickshank, A., & Sbai, E. (2015). Gender differences in personnel management: Some experimental evidence. *Journal of Behavioral and Experimental Economics, 58*, 20-32.

Chaudhuri, A., & Gangadharan, L. (2007). An Experimental Analysis of Trust and Trustworthiness. *Southern Economic Journal, 73*(4), 959-985.

Chaudhuri, A., Li, Y., & Paichayontvijit, T. (2016). What's in a frame? Goal framing, trust and reciprocity. *Journal of Economic Psychology, 57*, 117-135.

Chaudhuri, A., Paichayontvijit, T., & Shen, L. (2013). Gender Differences in Trust and Trustworthiness: Individuals, Single Sex and Mixed Sex Groups. *Journal of Economic Psychology, 34*, 181-194.

Chaudhuri, A., & Sbai, E. (2011). Gender Differences in Trust and Reciprocity in Repeated Gift Exchange Games. *New Zealand Economic Papers, 45*(1-2), 81-95.

Chaudhuri, A., Schotter, A., & Sopher, B. (2009). Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Effort Games with Private, Almost Common and Common Knowledge of Advice. *Economic Journal, 119*(534), 91-122.

Coffman, K. (2014a). Gender Differences in Willingness to Guess. *Management Science, 60*(2), 434-448.

Coffman, K. (2014b). Evidence on Self-Stereotyping and the Contribution of Ideas. *The Quarterly Journal of Economics, 129*(4), 1625-1660.

Cox, J. C., & Deck, C. A. (2006). When are Women more Generous than Men? *Economic Inquiry, 44*(4), 587-598.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature, 47*(2), 448-474.

Cuadrado, I., Navas, M., Molero, F., Ferrer, E., & Morales, J. (2012). Gender Differences in Leadership Styles as a Function of Leader and Subordinates' Sex and Type of Organization. *Journal of Applied Social Psychology, 42*, 3083-3113.

Dittrich, M. (2015). Gender differences in trust and reciprocity: evidence from a large-scale experiment with heterogeneous subjects. *Applied Economics, 47*(36), 3825-3838.

Dutta-Gupta, N. D., Poulsen, A., & Villeval, M. C. (2013). Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry, 51*, 816-835.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association, 9*(3), 522-550.

Eagly, A., & Carli, L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly, 14*, 807-834.

Eagly, A., & Johannesen-Schmidt, M. (2001). *Journal of Social Issues, 57*, 781-797.

Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, Transactional, and Laissez-Faire Leadership Styles: A Meta-Analysis Comparing Women and Men. *Psychological Bulletin, 129*(4), 569-591.

Eagly, A. H., & Johnson, B. T. (1990). Gender and Leadership Style: A Meta-Analysis. *Psychological Bulletin, 108*(2), 233-256.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the Effectiveness of Leaders: A Meta-Analysis. *Psychological Bulletin, 117*(1), 125-145.

Eagly, A. H., & Karau, S. J. (2002). Role Congruity Theory of Prejudice Toward Female Leaders. *Psychological Review, 109*, 573-598.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the Evaluation of Leaders: A Meta-Analysis. *Psychological Bulletin, 111*, 3-22.

Eckel, C. C., & Grossman, P. J. (1996). The Relative Price of Fairness: Gender Differences in a Punishment Game. *Journal of Economic Behavior & Organization, 30*(2), 143-158.

Eckel, C. C., & Grossman, P. J. (2002). Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk. *Evolution and Human Behavior, 23*(4), 281-295.

Eckel, C. C., & Grossman, P. J. (2008a). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (Vol. 1, pp. 509-519). New York: Elsevier.

Eckel, C. C., & Grossman, P. J. (2008b). Forecasting Risk Attitudes: An Experimental Study using Actual and Forecast Gamble Choices. *Journal of Economic Behavior & Organization, 68*(1), 1-17.

Exley, C., & Kessler, J. (2019). The Gender Gap in Self-Promotion. *NBER Working Paper*.

Egas, M., & Riedl, A. (2008). The Economics of Altruistic Punishment and the Maintenance of Cooperation. *Proceedings of the Royal Society B:, 275*, 871-878.

Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making, 5*(3), 159-163.

Exley, C., Niederle, M., & Vesterlund, L. (2020). Knowing When to Ask: The Cost of Leadning in. *Journal of Political Economy, 128(3),* 816-854.

Fehr-Duda, H., de Gennaro, M., & Schubert, R. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision, 60*(2-3), 283-313.

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review, 90*(4), 980-994.

Fehr, E., & Gächter, S. (2002). Altruistic Punishment in Humans. *Nature, 415*, 137-140.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica, 65*(4), 833-860.

Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998). When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets. *Journal of Labor Economics, 16*(2), 324-351.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quarterly Journal of Economics, 108*(2), 437-459.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1996). Involuntary Unemployment and Non-Compensating Wage Differentials in an Experimental Labour Market. *Economic Journal, 106*, 106-121.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review, 42*(1), 1-34.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and Contract Design. *Econometrica, 75*(1), 121-154.

Fehr, E., & Rockenbach, B. (2003). Detrimental Effects of Sanctions on Human Altruism. *Nature, 422*, 137-140.

Fernandez-Mateo, I., & Fernandez, R. (2016). Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs. *Management Science, 62*(12), 3636-3655.

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

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European Economic Review, 90*, 302-325.

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender Differences in Competition: Evidence From a Matrilineal and Patriarchal Society. *Econometrica, 77*(5), 1637-1664.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics, 118*(3), 1049-1074.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review, 104*(4), 1091-1119. Retrieved from https://doi.org/10.1257/aer.104.4.1091

Goldin, C., Katz, L., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives, 20 (4)*, 133-56.

Grossman, P., Eckel, C., Komai, M., & Zhan, W. (2019). It pays to be a man: Rewards for leaders n a coordination game. *Journal of Economic Behavior and Organzation*, 197-215.

Grossman, P., Komai, M., & Jensen, J. (2015). Leadership and gender in groups: An Experiment. *Canadian Journal of Economics, 48*, 368-388.

Harbaugh, W. T., Krause, K., & Vesterlund, L. (2002). Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses. *Experimental Economics, 5*(1), 53-84.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking Measured Risk Aversion to Individual Characteristics. *Kyklos, 55*(1), 3-26.

Haselhuhn, M., Kennedy, J., Kray, L., Van Zant, A., & Schweitzer, M. (2015). Gender differences in trust dynamics: Women trust more than men following a trust violation. *Journal of Experimental Social Psychology, 56*, 104-109.

Helliwell, J. (2006). Well-being, Social Capital and Public Policy: What's New? *The Economic Journal, 116*, C34-C45.

Heilman, M. (2012). Gender sereotypes and workplace bias. *Research in Organizational Behavior, 32*, 113-135.

Holt, C. (2009). University of Virginia: Veconlab. Online experiments available at http://veconlab.econ.virginia.edu/.

Holt, C., & Laury, S. (2002). Risk Aversion and Incentive Effects. *American Economic Review, 92*(5), 1644-1655.

Hudgens, G. A., & Fatkin, L. T. (1985). Sex Differences in Risk Taking: Repeated Sessions on a Computer-Simulated Task. *Journal of Psychology: Interdisciplinary and Applied, 119*(3), 197-206.

Ingram, B. L., & Berger, S. E. (1977). Sex-Role Orientation, Defensiveness, and Competitiveness in Women. *Journal of Conflict Resolution, 21*(3), 501-518.

Jena, A. B., Olenski, A. R., & Blumenthal, D. M. (2016). Sex differences in physician salary in U.S. public medical schools. *JAMA Intern Medicine, 176*(9), 1294-1304. doi:10.1001/jamainternmed.2016.3284

Jianakoplos, N. A., & Bernasek, A. (1998). Are Women More Risk Averse? *Economic Inquiry, 36*(4), 620-630.

Knack, S., & Keefer, P. (1997). Does Social Capital have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics*, 1251-1288.

Knez, M., & Camerer, C. (1994). Creating Expectational Assets in the Laboratory: Coordination in 'Weakest-Link' Games. *Strategic Management Journal, 15*, 101-119.

Knez, M., & Camerer, C. (2000). Increasing Cooperation in Prisoner's Dilemmas by Establishing a Precedent of Efficiency in Coordination Games. *Organizatinal Behavior and Human Decision Processes, 82*(2), 194-216.

Koch, A., D'Mello, S., & Sackett, P. (2015). A Meta-Analysis of Gender Stereotypes and Bias in Experimental Simulations of Employment Decision Making. *Journal of Applied Psychology, 100*, 128-161.

Koenig, A., Eagly, A., Mitchell, A., & Ristikari, T. (2011). Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms. *Psychological Bulletin, 137*, 616-642.

Kvaløy , O., & Luzuriaga, M. (2014). Playing the trust game with other people's money. *Experimental Economics, 17*, 615-630.

Leibbrandt, A., & List, J. (2015). Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment. *Management Science, 61*(9), 2016-2024.

Levati, M. V., Miettinen, T., & Rai, B. (2011). Context and Interpretation in Laboratory Experiments: The Case of Reciprocity. *Journal of Economic Psychology, 32*(5), 846-856.

Moran, B. B. (1992). Gender Differences in Leadership. *Library Trends, 40*(3), 475-491.

Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete too Much? *Quarterly Journal of Economics, 122*(3), 1067-1101.

Nikiforakis, N., & Normann, H.-T. (2008). A Comparative Statics Analysis of Punishment in Public Goods Experiments. *Experimental Economics, 11*(4), 358-369.

Northouse , P. G. (2012). *Introduction to Leadership: Concepts and Practice.* Los Angeles: SAGE.

Nosenzo, D., Offerman, T., Sefton, M., & van der Veen, A. (2010). *Inducing Good Behavior: Bonuses verus Fines in Inspection Games*. CEDEX Working Paper 2010-21. Nottingham, UK.

Nosenzo, D., Offerman, T., Sefton, M., & van der Veen, A. (2016). Discretionary Sanctions and Rewards in the Repeated Inspection Game. *Management Science, 62*(2), 502-517

Rauhut, H. (2009). Higher Punishment, Less Control? Experimental Evidence on the Inspection Game. *Rationality and Society, 21*(3), 359-392.

Reuben, E., & Timko, K. (2018). On the effectiveness of elected male and female leaders and team coordination. *Journal of the Economic Science Association, 4*(2), 123-135.

Ridgeway, C. (2001). Gender, Status, and Leadership. *Journal of Social Issues, 57*(4), 637-655.

Riley-Bowles, H., Babcock, L., & Lai, L. (2007). Social incentives for gender diVerences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes, 103*, 84-103.

Rosener, J. B. (1990). Ways Women Lead. *Harvard Business Review, 68*(6), 119-125.

Samuelson, L. (2005). Economic Theory and Experimental Economics. *Journal of Economic Literature, 43*(1), 65-107.

Sandberg, S. (2013). *Lean In: Women, Work and the Will to Lead*. New York: Knopf Doubleday.

Sax, L. (2005). *Why Gender Matters: What Parent and Teachers Need to Know about the Emerging Science of Sex Differences*. New York: Doubleday.

Schwieren, C. (2012). The Gender Wage Gap in Experimental Labor Markets. *Economics Letters, 117*(3), 592-595.

Shahriar, A., Unda, L., & Alam, Q. (2020). Gender differences in the repayment of microcredit: The mediating role of trustworthiness. *Journal of Banking & Finance, 110*, 1-16.

Smith, A. (1984). *The Theory of Moral Sentiments.* Indianapolis, IN, United States: Liberty Fund Inc.

Sundén, A. E., & Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *American Economic Review, 88*(2), 207-211.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. *The American Economic Review, 80*(1), 234-248.

Vandegrift, D., & Yavas, A. (2009). Men, women, and competition: An experimental test of behavior. *Journal of Economic Behavior and Organization, 72*(1), 554-570.

Walters, A. E., Stuhlmacher, A. F., & Meyer, L. L. (1998). Gender and Negotiator Competitiveness: A Meta-analysis. *Organizational Behavior and Human Decision Processes, 76*(1), 1-29.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys, 19*, 479-511.

Wood, R., Corcoran, M., & Courant, P. (1993). Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyer's Salaries. *Journal of Labor Economics, 11*, 417-441.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*, 75-98.