MONASH University

MONASH BUSINESS SCHOOL

Department of Economics
ISSN number 1441-5429

The gender leadership gap in competitive and cooperative institutions

Discussion Paper no. 2024-10

Catherine C. Eckel, Lata Gangadharan, Philip J. Grossman, Miranda Lambert and Nina Xue

Abstract:

This study investigates the impact of the institutional setting on the gender leadership gap. Motivating our study is the belief that women shy away from competitive environments and tend to prefer cooperative environments. We design an experiment using a modified Centipede game to test whether leaders can foster cooperation under two incentive schemes: competitive ("winner takes all") versus cooperative (equal earnings distribution). The leader, whose gender is revealed, sends a message providing strategic advice to their group. We find that male and female leaders are similarly effective in enhancing efficiency and are anticipated to perform equivalently. However, in the competitive context, a gender gap emerges: Female leaders receive lower evaluations than male leaders for offering identical advice. Interestingly, this bias is not observed in the cooperative context, suggesting that the congruence of the environment with gender stereotypes has important implications for leadership evaluations. Randomly-selected female leaders are evaluated 50% higher in a cooperative, as compared to a competitive environment. Thus, achieving gender equality in leadership requires careful attention to the institutional design of organizations. Notably, men consistently demonstrate a higher propensity to lead, regardless of the surrounding context.

Keywords: gender, leadership, institutional environment, evaluation, experiment

JEL Classification: C92, J16, J71, M14

Catherine C. Eckel: Texas A&M University (email: ceckel@tamu.edu); Lata Gangadharan: Monash University (email: Lata.Gangadharan@monash.edu); Philip J. Grossman: Monash University (email: Philip.Grossman@monash.edu); Miranda Lambert: Texas A&M University (email: miranda.lambert@tamu.edu); Nina Xue: Monash University (email: Nina.Xue1@monash.edu).

The gender leadership gap in competitive and cooperative institutions*

Catherine C. Eckel †, Lata Gangadharan ‡, Philip J. Grossman §, Miranda Lambert ¶, Nina Xue ‖

20 May 2024

Abstract

This study investigates the impact of the institutional setting on the gender leadership gap. Motivating our study is the belief that women shy away from competitive environments and tend to prefer cooperative environments. We design an experiment using a modified Centipede game to test whether leaders can foster cooperation under two incentive schemes: competitive ("winner takes all") versus cooperative (equal earnings distribution). The leader, whose gender is revealed, sends a message providing strategic advice to their group. We find that male and female leaders are similarly effective in enhancing efficiency and are anticipated to perform equivalently. However, in the competitive context, a gender gap emerges: Female leaders receive lower evaluations than male leaders for offering identical advice. Interestingly, this bias is not observed in the cooperative context, suggesting that the congruence of the environment with gender stereotypes has important implications for leadership evaluations. Randomly-selected female leaders are evaluated 50% higher in a cooperative, as compared to a competitive environment. Thus, achieving gender equality in leadership requires careful attention to the institutional design of organizations. Notably, men consistently demonstrate a higher propensity to lead, regardless of the surrounding context.

JEL Classification: C92, J16, J71, M14
Keywords: gender, leadership, institutional environment, evaluation, experiment

*We are grateful to participants at the Economic Science Association (ESA) World Meeting 2022, University of Lyon, Panthéon-Sorbonne University, Australia, and New Zealand Workshop on Experimental Economics (ANZWEE) 2022, the Gender, Norms and Economics Workshop 2022, the Behavioral Research in Economics Workshop (BREW) 2022, CERGE-EI, Young Economists' Meeting 2023, the University of Arizona, the Economic Psychology Seminar, the 2024 ASSA Annual Meeting, the Australian Gender Economics Workshop 2024, Presidency College, Kolkata, New York University, Abu Dhabi, University of Economics, Bratislava, and Queensland University of Technology. This project received approval from the Monash University Human Research Ethics Committee (project number 26620). Funding was provided by a Monash University, Department of Economics research grant and the Australian Research Council (DP170101167).

†Texas A&M University, ceckel@tamu.edu
‡Monash University, Lata.Gangadharan@monash.edu
§Monash University, Philip.Grossman@monash.edu
¶Texas A&M University, miranda.lambert@tamu.edu
‖Monash University, Nina.Xue1@monash.edu

1 Introduction

The gender gap in leadership positions is well documented (Goldin, 2021; McKinsey & Company, 2022). This chasm is particularly stark in the private sector, in which only 6% of S&P 500 companies have female CEOs and women hold less than a third of all management roles (Catalyst, 2022). The gap is less pronounced in industries that are not-for-profit or have some nonprofit characteristics. For example, women hold 43% of CEO positions in the nonprofit sector (AAUW, 2022) and make up 19% of leaders in hospitals, where nonprofit and for-profit firms often coexist (Catalyst, 2022). There are several potential reasons for the lack of women in leadership (for a review of relevant literature, see Eckel et al., 2021). In this paper, we address the role of the institutional environment in explaining the gender leadership gap and its impact on the effectiveness and evaluation of leaders and self-selection into leadership roles.

The institutional environment, which refers to the rules, norms, understandings, beliefs, and taken-for-granted assumptions in an organization, is increasingly recognized for its role in gender disparities in organizations (Bohnet, 2016). Simple changes in elements of the institutional environment, such as rules about leader selection, can have major implications for the selection of female leaders (Erkal et al., 2022). Competitiveness is often seen as a key characteristic of a leader, and this expectation may also impact the gender leadership gap. Consistent with findings on gender differences in preferences for cooperation (e.g., Bilén et al., 2021) and competition (e.g., Niederle and Vesterlund, 2007), women have been found to prefer working in cooperative rather than more competitive environments (e.g., Kuhn and Villeval, 2015; Wozniak, 2016). While generally women tend to shy away from competition (e.g., Niederle and Vesterlund, 2007) and fail to match men's increased performance in competitive settings (e.g., Gneezy et al., 2003), they are more willing to compete when they can do so in a team (e.g., Healy and Pate, 2011; Dargnies, 2012), or can share their winnings with others (e.g., Cassar and Rigdon, 2021a,b). Less is known, however, about how these gender differences in preferences for the work environment manifest in positions of leadership.

An important element of the institutional environment, and the focus of this paper, is the incentive structure.[1] Kuhn (2009) shows that compensation based on individual performance leads to perceptions of a more individualist culture, which is linked to more competitive behavior (e.g., Leibbrandt et al., 2013), while an organization with team-based incentives is more likely to be perceived as having a collectivist culture (e.g., Cox et al., 1991). Competitive incentives, such as a large individual bonus for the top performer in a team, can

[1]Other aspects of organizations can also influence whether the institutional environment is more competitive or cooperative, for example: setting common goals for a team, whether collaboration is actively encouraged, and hiring practices.

motivate workers, but one employee winning necessarily means that others lose out or receive a lesser reward.[2] Alternatively, an organization may choose to motivate workers using a more cooperative incentive scheme, such as equally dividing the surplus that is generated by the team.[3] Our purpose is to examine the role of the organizational environment on leadership outcomes for women. We conjecture that in a relatively more competitive environment, women may find leadership roles less attractive and may be stereotyped as lacking the requisite characteristics of a good leader. Alternatively, congruent with the stereotype that women are more prosocial and other regarding, a more cooperative environment may encourage more women to seek leadership positions in which their leadership styles may be assessed more favorably.

To this end, we design a laboratory experiment to isolate the effect of gender from other characteristics that might affect leader effectiveness or evaluation. Disentangling causal relationships between gender and leadership outcomes can be difficult or impossible using observational data alone because of the complex interactions that underlie a leader's accomplishments and the assessment of leaders. Experimental methods are therefore useful for isolating the effect of the leader's gender on followers' decisions and the evaluation of male and female leaders without confounds.

We employ the Centipede game, introduced by Rosenthal (1981), as our workhorse because it captures key elements of an organization in which total productivity is maximized by repeated cooperation within a team; but individual workers are tempted to pursue their self-interest at the team's expense by free riding on the efforts of others or by taking credit for another's work. In this game, the risk that a teammate may defect incentivizes players to be the first to stop cooperating (for reviews see Camerer, 2003; Krockow et al., 2016). Participants first play the Centipede game in anonymous pairs without a leader. They are then matched in new groups of three with one participant, whose gender is revealed to the group, randomly selected to be the leader. Another advantage of the Centipede game is that payoffs can be easily varied to make the game more competitive or cooperative, allowing us to test our hypotheses about the effect of the environment on the gender leadership gap.

The function of the leader is to persuade others to voluntarily coordinate and choose actions that are best for the collective (Kruse, 2013). The quantity and type of leader communication have proven to be important factors in motivating teams to pursue a socially

[2] A related example is the performance review system developed by Enron which involved grading employees based on their relative performance and firing the bottom 15 percent every year, known internally as "rank-and-yank" (Duckworth, 2016). Tournament incentives can also backfire and lead to sabotage on the part of workers who are less likely to win (Carpenter et al., 2010).

[3] Charness and Grieco (2023) show that cooperative incentives are more effective than competitive incentives at fostering creativity within a group.

optimal outcome (e.g., Weber et al., 2001; Brandts et al., 2015; Dong et al., 2018). Indeed, leaders are more effective when they can make a free-form plea as opposed to a preset suggestion (e.g., Charness et al., 2023). In our context, the leader acts as an external advisor whose task is to send a free-form, nonbinding message persuading their followers to maximize total earnings for the group, an outcome that also maximizes the leader's own payoff. We expect that, in line with the literature, this type of leadership will produce a significantly higher likelihood of reaching a socially optimal outcome. The contribution of our paper is to show the effect of the gender of the leader and of the institutional setting on leader effectiveness.

Our experimental design varies both the institutional environment (competitive or cooperative) and the gender of the leader. We vary the environment by modifying the allocation of payoffs in a Centipede game that is otherwise identical. To model a setting in which project funding is uncertain and susceptible to external shocks, we implement a possible early ending to the game, randomly determined by the computer (Krockow et al., 2018).[4] The competitive environment *(Comp)* is competitive in payoffs, in the sense that the first player to end the game receives a substantially larger share of profits. In the cooperative environment *(Coop)*, either player can also choose to end the game by exiting, with the same payoff consequences. If both players cooperate until the final node, or the end node that is determined by the computer, then payoffs are the same for both players. We then measure leader effectiveness by observing followers' exit choices, which determine the total productivity of the group. An equal number of women and men are recruited to the session, and the assignment to groups and to leadership roles is randomly determined, which gives us the necessary exogenous variation in the gender of the leader without calling unnecessary attention to the role of gender in the study.

Our results show that leaders are effective at raising productivity, and that whether advice comes from male or female leaders does not matter for followers' choices. Despite equal effectiveness, however, female leaders receive more negative evaluations than male leaders, but only in the competitive environment. Female leaders in the competitive environment are evaluated 50% lower than their counterparts in the cooperative environment; evaluations of male leaders in the two environments are similar. Women are also more likely than men to receive monetary penalties from male followers in the competitive setting. We find no such gender difference in evaluations or penalties in the cooperative setting.

We examine several potential mechanisms underlying the results. First, we find that the gender gap in evaluations in the competitive environment cannot be explained by beliefs

[4]This also ensures that both players have an equal chance of ending up with a larger share at the terminal node.

about leader effectiveness, as elicited beliefs are accurate in predicting both the positive effect of a leader and the absence of a gender difference. Second, based on a sentiment analysis and an examination of the content of leaders' messages, we find that male and female leaders do not differ on average in the advice they give, but even after controlling for the content of leaders' messages, the bias against women persists. Our third mechanism explores the perceptions of followers and finds that female leaders receive an additional penalty in evaluations when their advice is perceived to be "bad" or "selfish;" that is, when gender stereotypes (i.e., women are more fair-minded) are thought to be violated. Our findings thus suggest that traditional gender stereotypes play a key role in the biased evaluation of female leaders. Finally, men consistently report a greater willingness to lead than women in both environments, with evidence from participants suggesting that gender stereotypes also play an important role in self-perceptions about leadership ability.

Our research makes several contributions. First, there is a body of recent work showing, in a variety of settings, no gender differences in the performance of leaders (Reuben and Timko, 2018; Grossman et al., 2019; Shurchkov and van Geen, 2019).[5] Heursen et al. (2023) further explore whether male and female leaders differ in their relative effectiveness using a more sensitive measure and still report a null result. We contribute to this literature by directly comparing leadership outcomes for men and women in a competitive and a cooperative setting, showing that these results hold in both environments.

Second, there is a literature that explores the effect of the environment on perceptions of leaders. Eagly et al. (1995) find that women are perceived as less effective leaders than men in more male-oriented industries such as the military, while women are more favored in industries related to education, government, and social services. Mengel et al. (2019) document a bias against women in university teaching evaluations, despite similar student performance under male and female instructors. This bias is larger in mathematical courses, which are typically considered to be more male dominated. This finding is also related to several recent papers exploring the role of gender-based stereotypes on evaluations, by both women themselves and related to confidence (e.g., Coffman, 2014; Bordalo et al., 2019; Exley and Kessler, 2022) and by others (e.g., Sarsons et al., 2021; Erkal et al., 2023).[6] Our contribution is to isolate the role of the competitive/cooperative element of the institution and show that women leaders receive negatively biased evaluations in the competitive environment. This might underlie women's preference to work in more cooperative environments.

[5]Shurchkov and van Geen (2019) also find that female managers are less likely to choose competitive incentives to motivate worker effort, but only when their gender is revealed.

[6]Stereotypes have the tendency to exaggerate true underlying differences based on a "kernel of truth" (e.g., Bordalo et al., 2019), for example, the gender difference in risk (e.g., Eckel and Grossman, 2002) and prosocial preferences (e.g., Cason et al., 2022).

There is a considerable literature suggesting that the gender leadership gap arises because men and women have different preferences. There is robust evidence that women tend to be more risk averse (e.g., Eckel and Grossman, 2008) and less competitive (e.g., Niederle and Vesterlund, 2007) than men, two qualities that are generally considered important for leadership.[7] Our finding that women report being less willing to lead in both environments complements the work on preferences and points to an additional explanation: women's self-perceptions.[8] Women perceive their abilities to be lower than those of their male counterparts, despite their equivalent performance in practice.

A further contribution is to expand the Centipede game in two ways. To the best of our knowledge, we are the first to examine the effect of leadership in raising team productivity in the context of the Centipede game. The game is well suited for studying leadership because increasing social efficiency is built into the game and depends on the ability of team members to cooperate and forgo individual gains. We also make a methodological contribution through an experimental design that exogenously varies the nature of the environment, while keeping all else constant, including total payoffs and the leader's incentives. Our design could be adapted to study other aspects of leadership across the two environments.

Finally, our study offers insights into both the supply and the demand side of leadership. On the supply side, our findings indicate that women exhibit lower willingness to *supply* leadership (i.e., they are less willing to take on leadership roles) than men. This occurs irrespective of the environment. On the demand side, the *demand* for female leaders appears to be less than the demand for male leaders. They face greater scrutiny and are evaluated and penalized more harshly than male leaders, but only in the competitive environment. The mechanism underpinning the results on both sides of the market stem from prevailing stereotypes. In the competitive environment, the stereotypes that lead to followers' biased perceptions about the advice given by female leaders is reinforced by the self-stereotypes that women hold about their abilities being lower, thus explaining the stark gender leadership gaps observed in such environments, such as competitive private sector fields (e.g., tech and banking). In the cooperative environment, gender differences appear only on the supply side of leadership, and this may explain the more modest gender gaps observed in the not-for-profit sector.

An important policy implication arising from our research is that the effectiveness of strategies aimed at reducing the gender leadership gap in organizations is contingent on the prevailing institutional environment. In situations in which the observed gender gap

[7] Mas and Pallais (2017) also find that women tend to prefer more flexible working arrangements.

[8] In our study, participants are informed that followers can choose to reward or penalize their leader, however leaders do not receive their followers' evaluations. This rules out the possibility of receiving negative feedback, which Chakraborty and Serra (2023) show could discourage women from leading.

primarily stems from supply side factors, policies that promote proactive participation and uptake of leadership roles by women may be particularly effective. However, in cases where the main barrier lies on the demand side, such policies may backfire. Encouraging women to be assertive and pursue leadership roles could potentially lead to a backlash against them, further discouraging them from considering future leadership positions. We discuss this further in the Conclusion. A promising way forward for organizations wishing to address the gender leadership gap is to, where possible, create more cooperative environments that could help to "even the playing field" for female leaders.

2 Experimental design

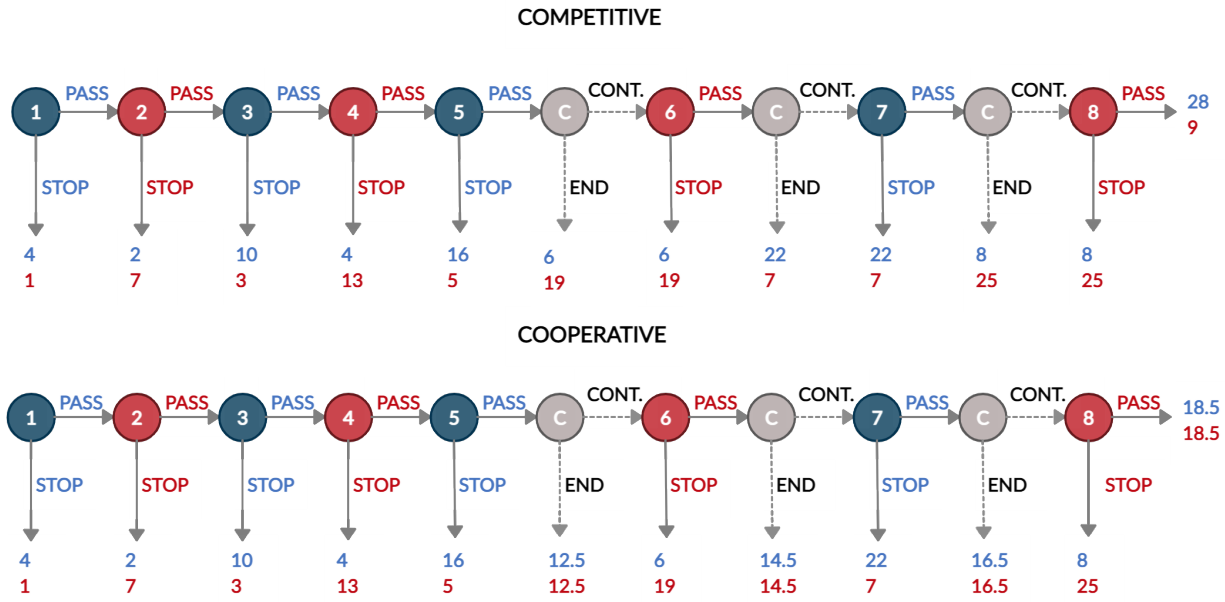
The Centipede game, introduced in Rosenthal (1981), is a strategic game in which players alternate turns deciding whether to take a larger portion of an increasing pot, or to pass the decision to the other player. Each player faces a dilemma: Taking the pot now yields immediate gains, but waiting can potentially lead to higher rewards. Figure 1 illustrates the game as we implemented it in our experiment. The game is played in pairs and each pair consists of one Blue player (who makes decisions in the four odd decision nodes) and one Red player (who makes decisions in the four even decision nodes). The Blue player moves first choosing between exiting ("Stop"), earning 4 for himself and 1 for the Red player, or passing to the Red player ("Pass"). The Red player then has a parallel choice, whether to exit, earning 7 for herself and 2 for the Blue player, or to pass back to the Blue player. This continues until one player exits or the terminal node is reached. Total payoffs increase with each round and are maximized if the game reaches the terminal node. However, the game is inherently competitive, in that the payoffs are unequal at each node and the decision maker has an incentive to exit first, beating out their counterpart.

The game is solvable by backward induction, with the prediction that players will stop at the first opportunity. However, contrary to the game-theoretic solution, players rarely stop in the early nodes (McKelvey and Palfrey, 1992). Proposed explanations include altruistic concerns for the opponent, backward induction being too cognitively demanding, and non-equilibrium beliefs about the opponent (e.g. Palacios-Huerta and Volij, 2009; Gamba and Regner, 2019; García-Pola et al., 2020b).

An important feature of the Centipede game is that if players choose to pass at each decision node, the total payoff continues to grow, and this growing productivity is built into the game itself. This means that once leaders are introduced, they have a potential role in improving group outcomes.[9] Another important requirement that is satisfied by the

[9]The data confirms that this is indeed the case in Section 3.1.

Figure 1: Decisions and payoffs in *Comp* and *Coop*



Notes: *Comp* and *Coop* are identical in case either player chooses to Stop. The two environments only differ in the distribution of payoffs at the end node and those determined by the computer.

Centipede game is that it can be easily adapted to be either competitive or cooperative in nature.

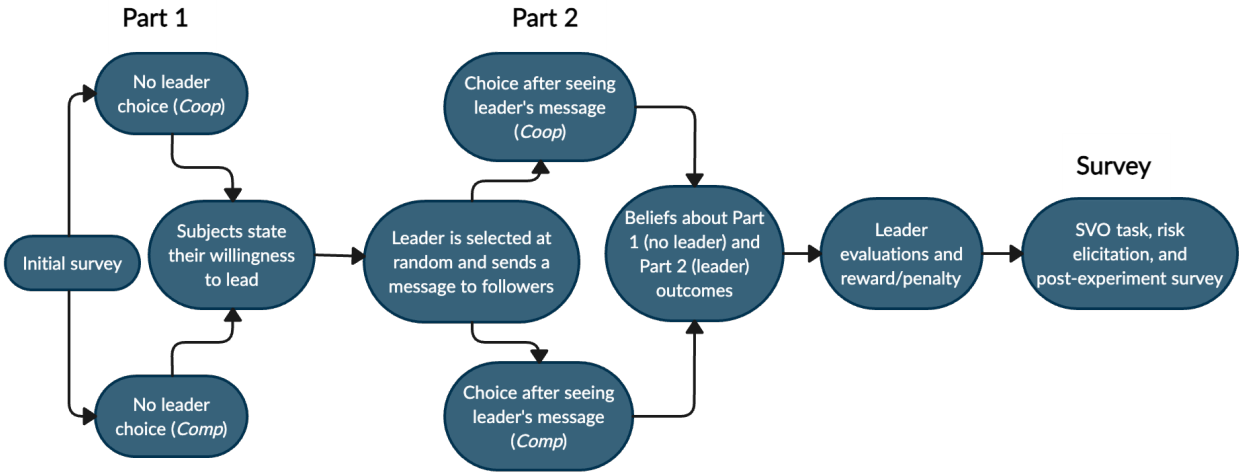
A key element of our implementation is the possibility that the game will end randomly. In many collaborative interactions, there is a degree of uncertainty about the length of time that the interaction will continue. This uncertainty is implemented here by introducing a random termination node. Following Krockow et al. (2018), we introduce a probability that the game will end at each node after round 5.[10] This ensures that both players have an equal chance of ending up with a larger share at the computer-determined terminal node in the competitive environment. This random ending also gives us an opportunity to alter the game to make it a more cooperative environment. We do this by manipulating the random-ending; total payoffs are the same, but instead are equalized across the two players. Thus, we have a between-subjects experiment with two main treatments: Competitive (*Comp*, upper panel of Figure 1), and Cooperative (*Coop*, lower panel of Figure 1). These are explained in more

[10]In contrast, an alternative use of random endings is to mimic the incentive effects of infinite play with discounting, but without having to play forever, This method, proposed by Roth and Murnighan (1978), links the number of expected repetitions of the stage game to the discount factor.

detail below.[11]

In an initial survey, subjects select a gender-specific alias, which is later used to reveal the leader's gender to their group without drawing undue attention to gender.[12] Gender (including the gender of followers) is otherwise not revealed or discussed. We also asked for subjects' age, field, and year of study to reduce the salience of gender in the survey. In Part 1, subjects play the Centipede game in anonymous pairs and subsequently reveal their willingness to be a leader in Part 2. In Part 2, subjects are rematched into new groups of three, consisting of one leader and two anonymous followers. Followers make a second choice in the Centipede game after seeing their leader's chosen alias and a message from their leader. Next, both followers and leaders report their beliefs about Part 1 and 2 outcomes. Followers also evaluate their leader and can choose to adjust their leader's earnings. Finally, the post-experiment survey elicits social preferences, risk preferences, demographic information, and explanations for decisions made in the game. The experimental timeline is summarized in Figure 2.

Figure 2: Timeline for *Comp* and *Coop*



[11]For a further discussion of why we chose the Centipede game, see Appendix A.

[12]See the full list of names in Appendix B. This name-based approach follows Chakraborty and Serra (2023) and avoids experimenter demand bias. If subjects select "non-binary", "gender-diverse", or "my gender is not listed", they can choose between two gender-neutral names.

2.1 Choices without a Leader (Part 1)

2.1.1 Competitive Environment

In Part 1, subjects are randomly matched in pairs and play a Centipede game using the strategy method.[13] Each player makes an exit choice regarding when to "Stop" (at their first, second, third, or fourth decision node), or to "Always Pass" (i.e., never Stop). The Blue player decides between stopping at exit node 1, 3, 5, 7, or to Always Pass, while the Red player decides between stopping at exit node 2, 4, 6, 8, or to Always Pass. After node 5, the computer may randomly end the game before either node 6, 7, 8, or never, with equal probability, and payoffs are equivalent to the game reaching the next node. Payoffs are determined by the earliest Stop decision, or the terminal node determined randomly by the computer, whichever comes first. This element of uncertainty is important as it means either player could end up with the larger share at the terminal node and neither player has an obvious advantage at the outset. For example, if both Blue and Red chose Always Pass, but the computer chose to end the game before node 7, then the game would end at node 7, Blue would receive \$22, and Red would receive \$7. If instead the computer randomly ends the game before node 6, Blue would receive \$6, and Red would receive \$19. The environment is competitive in payoffs as one player receives a significantly larger share of the joint profits. The Blue and Red players must therefore compete with one another to receive the larger share while at the same time trying to maximize the total profits. Total payoffs are potentially maximized if both players choose Always Pass (\$37), with Blue receiving \$28 and Red receiving \$9.

2.1.2 Cooperative Environment

The cooperative environment (Figure 1) in Part 1 is identical to *Comp*, should either player choose to stop before the final node. For example, if the computer does not end the game before node 6, Blue chooses Always Pass, and Red chooses to Stop at node 6, then Red will earn \$19 and Blue will earn \$6. We chose to retain the unequal payoffs following a Stop decision in *Coop* to model the risk that either player could stop cooperating with their teammate and take a larger share for themselves.

[13]The strategy method allows us to obtain decisions pertaining to all possible decision nodes, including those that are not reached in realized play. Further, this approach meant that feedback about the outcomes could be delayed until the end of the session, thus allowing the elicitation of beliefs about exit choices. García-Pola et al. (2020a) find that the exit choice tends to be earlier under "hot" (realized) play as compared to "cold" play (strategy method) when the game has constant or increasing total payoffs. Since we use the strategy method to elicit all exit choices, this design choice should have no impact on any treatment or gender comparisons.

The cooperative nature of this environment is reflected by the fact that players have the opportunity to receive an equal share of group earnings if they both cooperate and reach the most socially efficient outcome. For example, if the computer chooses not to end the game and both players choose to Always Pass, the total payoff is maximized ($37) and both players earn $18.50. Likewise, if both players choose Always Pass and the computer chooses to end the game before node 7, then both players would receive $14.50.

2.2 Choices with a leader (Part 2)

Before participants are informed of their role in Part 2, we measure willingness to lead (WTL) by asking each subject to indicate (on a scale from 1 to 10) how much they want to be the leader of their group. We randomly assign participants to leadership positions (and subjects are aware of this) to ensure that we have an equal sample of male and female leaders and to reduce selection bias.[14]

The leader's gender is revealed through their chosen alias. The other two group members (followers) remain in the same role (Blue or Red) and game as in Part 1 (either *Comp* or *Coop*).[15] The leader has a minimal advisory role and is instructed to send a message to their followers to explain potential strategies. The leader has no decision to make in the game. We explore the role of leaders in managing followers' beliefs, which has been shown to improve social efficiency in games that require coordination (e.g., Sahin et al., 2015; Gächter and Renner, 2018). The interests of leaders and followers coincide, as leaders receive the average earnings of Blue and Red in Part 2, plus an additional $3 to represent a "leadership premium." Thus, the leader is responsible for setting the goal while the followers are responsible for achieving it. Note that the leader's incentives are the same across treatments as total earnings are the same in *Comp* and *Coop*.

After seeing the leader's message and making an exit choice, followers are asked to evaluate the effectiveness of their leaders (on a scale from 1 to 7). Importantly, this evaluation is made before followers receive feedback about the outcomes of both Part 1 and Part 2. We elicit ratings of leaders to model commonly used tools such as 360-degree feedback, in which remuneration and promotion decisions are made based on unincentivized assessments by peers, managers, and subordinates. Followers also have the chance to adjust (either increase or decrease) the leader's earnings at a cost, with every $0.25 spent corresponding to a ±$1 change in the leader's payoff (up to ±$3). One follower in each group is chosen at random to have their choice implemented for the leader. We also elicit (incentivized) beliefs about

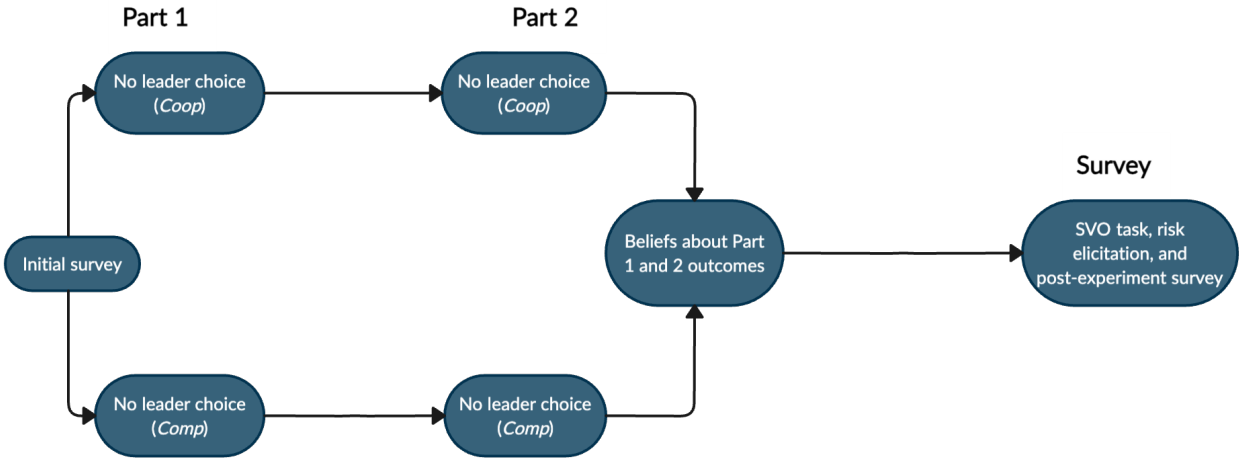
[14]For instance, those who self-select into leadership may be perceived to be more effective and be evaluated differently, thus making causal inferences about leadership and gender difficult.

[15]This allows us to directly compare choices with and without a leader at the individual level.

when the game will end in both Part 1 and Part 2.

To account for potential learning effects, we conduct two treatments (*Comp-Control* and *Coop-Control*), in which Part 1 is simply repeated and participants make a second decision in Part 2 without a leader. See Figure 3 for a summary of the experimental timeline. This allows us to isolate the effect of a leader on followers' decisions and rule out a difference in choices due to experience alone.

Figure 3: Timeline for *Comp-Control* and *Coop-Control*



2.3 Post-experiment survey

In the post-experiment survey, we elicit social preferences using the Social Value Orientation (SVO) measure (Murphy et al., 2011), in which participants make a series of allocations for themselves and a randomly chosen participant in the same session. The SVO measure ranges from Competitiveness (maximizing the difference between payoffs) to Altruism (maximizing the other's payoff), based on the preferred allocations. Second, we elicit individual risk preferences using the Eckel and Grossman (2002) measure, in which subjects are given a choice between six lotteries, each with a 50% chance of winning a prize. Finally, participants answer a demographic survey on their place of birth, ethnicity, education, and mother and father's education. In open-ended questions, we also ask participants to explain their: 1) decisions in the game; 2) reasons for stated WTL; 3) assessments of the leader (including leaders' self-assessments); and 4) reasons for rewarding/penalizing the leader. Feedback is only provided after the completion of the post-experiment survey.

2.4 Predictions

For payoff-maximizing agents, the subgame perfect equilibrium is that each player chooses to Stop at the first node in both the cooperative and competitive games, with or without a leader. In contrast, our behavioral hypothesis is that leaders are effective at encouraging their followers to make later exit choices. This is based on previous work showing that leaders have a powerful effect in many coordination games (Brandts et al., 2007; Sahin et al., 2015; Cooper and Weber, 2020). It is less clear whether the positive effect of leadership holds in both competitive and cooperative settings. Therefore, as a first step, we examine whether leaders are effective in raising productivity in both environments.

We subsequently consider three key research questions. First, are female leaders more effective in a cooperative environment than in a competitive one? Second, conditional on equal effectiveness, are female leaders are evaluated and rewarded on par with male leaders? Finally, are women more willing to become leaders when the environment is more cooperative in nature?

Consistent with gender norms around social roles (Eagly, 1987) in which women are expected to demonstrate more communal attributes than men, women are shown to prefer working in more cooperative environments (e.g., Kuhn and Villeval, 2015). Building on this, role congruity theory proposes that the perceived inconsistency between stereotypes about women and expectations about leaders would lead to less favorable attitudes towards female than male leaders (Ridgeway, 2001; Eagly and Karau, 2002). We expect *Coop* to create a more favorable environment for female leaders. The setting is more congruent with gender stereotypes about women being more caring and other regarding, while a more competitive environment is more consistent with gender stereotypes about men being more assertive and ambitious. We conjecture that the cooperative environment may improve the effectiveness of female leaders and reduce the gender gap in the assessment of leaders and willingness to become leaders.

2.5 Procedures

The experiment was programmed using oTree (Chen et al., 2016). Sessions were conducted at the Monash Laboratory for Experimental Economics at Monash University, using Sona to recruit subjects, and the Experimental Economics Laboratory at the University of Melbourne, using ORSEE (Greiner, 2015) for recruitment, between July - November 2021.[16] Due to stay-at-home orders in Melbourne, online sessions of approximately 18 subjects per session

[16]Our results do not differ across the Monash and Melbourne subject pools.

were conducted via Zoom with similar conditions to a laboratory environment.[17] We report results from a total of N=400 participants (N=300 in the main treatments and N=100 in the control treatments).[18] Subjects are informed that following the completion of the survey, either Part 1 or Part 2, plus either the SVO task or the risk task, will be chosen at random to be paid. On average, participants received $19.24 in *Comp* and $22.37 in *Coop* and the experiment lasted approximately one hour. Our experimental design, power analysis, and research questions were preregistered on AsPredicted.org (pre-registration #60725).

3 Results

3.1 Effect of leaders on exit nodes

In a preliminary analysis, we first check whether leaders have a positive impact on follower productivity by examining exit choices, which take a value between 1 and 5 for each player's decision nodes. Each player can choose to exit at one of the four decision nodes corresponding to their color (at every other node). Alternatively, they may choose to Always Pass (see Figure 1). For example, an exit choice of 3 for Blue (Red) is equivalent to the game ending at node 5 (6). Unless otherwise specified, we use two-tailed Mann-Whitney tests to compare differences in means. Without a leader, the average exit choice is 4.02 in *Comp* and 5.79 in *Coop* ($p < 0.01$). We find no gender difference in these exit choices: In *Comp*, women exit at exit choice 3.92 and men at 4.13 ($p = 0.37$); in *Coop*, women exit at 5.81 and men at 5.69 ($p = 0.68$). Exit choices also do not differ by the player's role with Blue players exiting at 3.07 and Red players exiting at 2.93 ($p = 0.51$) in *Comp*. Similarly, in *Coop*, Blue players exit at 3.91 while Red players exit at 3.69 ($p = 0.17$).

Figure 4 shows that leaders have a significantly positive effect on exit choices in both *Comp* (no leader: 4.02 vs. leader: 5.79, $p < 0.01$) and *Coop* (no leader: 5.76 vs. leader: 7.66, $p < 0.01$).[19] On average, the presence of a leader encourages subjects to make a later exit choice in both *Comp* and *Coop* of 24% and 19% respectively. This result is robust to demographic controls in OLS regressions (Table 1).[20] Following the leader's message, Blue players tend to exit later than Red players in *Comp* (3.84 vs. 3.54, $p = 0.04$) but not in *Coop*

[17]For example, sessions were anonymized, instructions were read out loud by the experimenter, and subjects could ask private questions to the experimenter via the chat function.

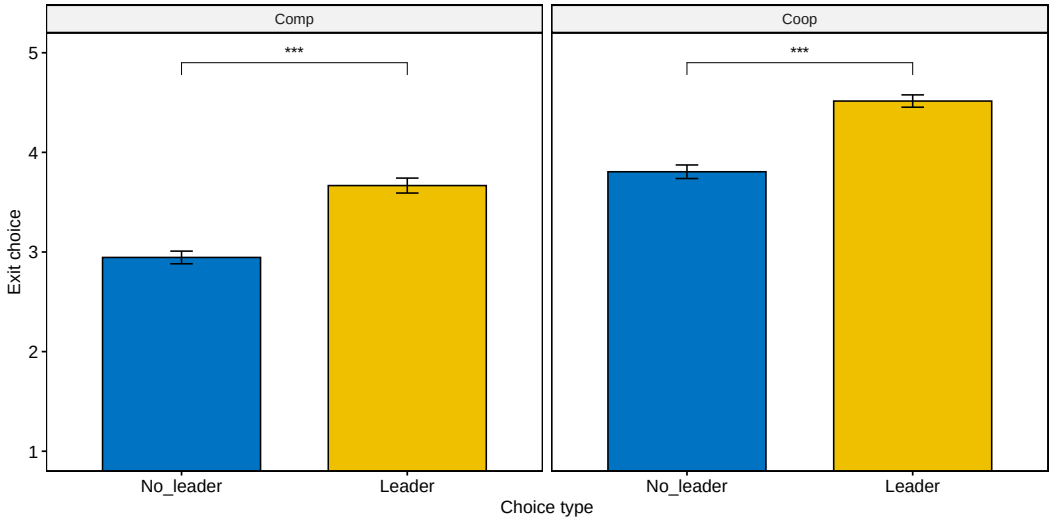
[18]We exclude from our analysis 5 participants who reported their gender as "non-binary" or "gender diverse" in *Comp* and 2 participants who reported their gender as "non-binary" in *Coop*.

[19]The results are consistent after excluding participants who were leaders in Part 2 for *Comp* (3.07 vs. 3.67, $p < 0.01$, Wilcoxon signed-rank test) and *Coop* (3.76 vs. 4.52, $p < 0.01$, Wilcoxon signed-rank test).

[20]All results from the OLS regression analyses hold when using an ordered probit regression. See Appendix C-E.

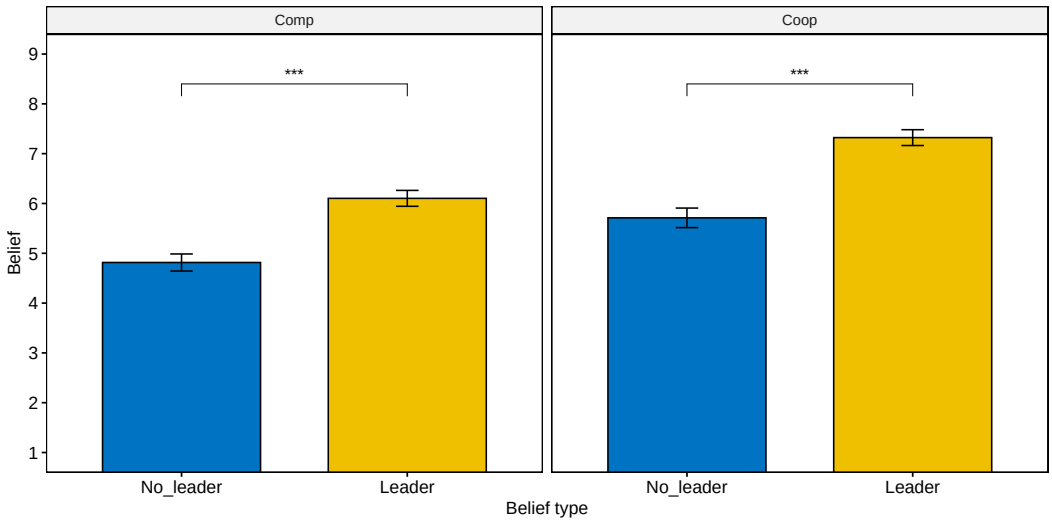
(4.56 vs. 4.46, $p = 0.15$). Leaders can help to coordinate followers' expectations regarding their opponent's choice and act as "belief managers" in the presence of uncertainty (e.g., Gächter and Renner, 2018). The difference in actual exit choices parallels beliefs about exit nodes. Without a leader, participants expect the game to end at node 4.84 in *Comp* and 5.83 in *Coop*. In the presence of a leader, these expectations increase significantly to 6.03 ($p < 0.01$) and 7.33 ($p < 0.01$), respectively (see Figure 5). Overall, choices in the game are consistent with participants' beliefs.

Figure 4: Exit choices with and without a leader



Note: Exit choices take a value between 1 and 5. Error bars represent standard errors.

Figure 5: Beliefs about exit nodes with and without a leader



Note: Beliefs about exit nodes take a value between 1 and 9. Error bars represent standard errors.

Table 1: Determinants of exit choices

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Leader	0.72***	0.67***	0.70***	0.70***
	(0.12)	(0.12)	(0.11)	(0.12)
Female	−0.06	−0.08	0.24	0.14
	(0.16)	(0.14)	(0.16)	(0.15)
Constant	2.98***	0.68	3.67***	4.01***
	(0.13)	(0.87)	(0.13)	(0.82)
Controls	No	Yes	No	Yes
R^2	0.10	0.25	0.11	0.23
Adj. R^2	0.10	0.21	0.10	0.19
Num. obs.	484	484	496	496

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the individual level in parentheses. The dependent variable is the exit choice (between 1 and 5). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Without a leader, choices do not differ significantly when participants are asked to make a second decision in the Centipede game in the control treatments (see Appendix C.2). We therefore conclude that it is the leader, rather than learning or experience, that drives the later exit choices in both *Comp* and *Coop*.

3.2 Leader Effectiveness: The role of the leader's gender on exit choices

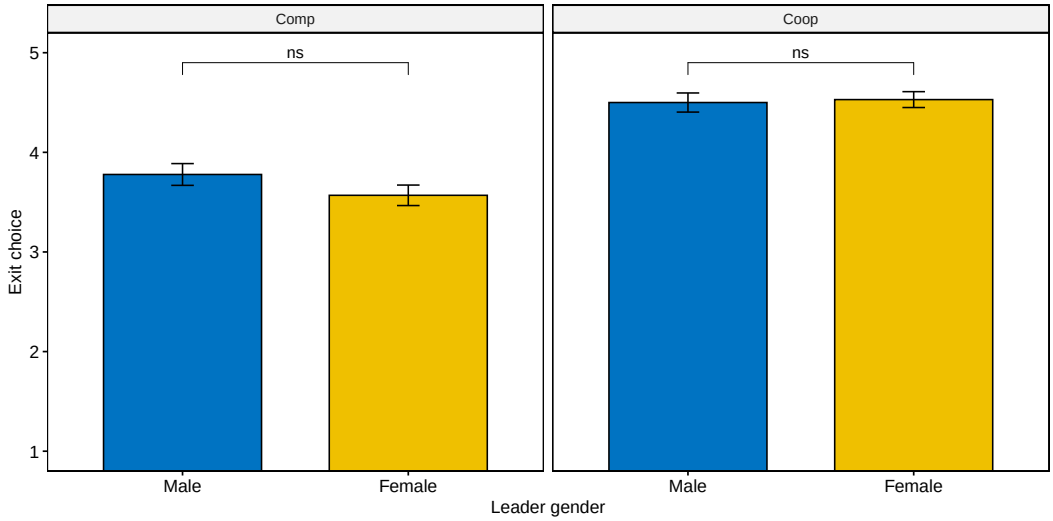
Next, we compare exit choices under male and female leaders. Similar to Section 3.1, we use the participant's exit choice as the dependent variable. On average, we find no significant gender difference in exit choices in *Comp* (M: 3.80 vs. F: 3.59, $p = 0.13$) or in *Coop* (M: 4.50 vs. F: 4.52, $p = 0.71$), as seen in Figure 6. The interaction between leader and follower gender is not significant in the regression analysis in Table 2 for *Comp* ($p = 0.39$, column 2) and for *Coop* ($p = 0.31$, column 4).[21] In *Coop*, female followers tend to exit later than male followers in the presence of a leader ($p < 0.01$, column 4), suggesting that women are more responsive to leadership in the more cooperative setting. This leads us to:

Result 1: Male and female leaders are equally effective in *Comp* and *Coop*.

[21] This result is robust to an alternative measure of leader effectiveness, using the difference in exit choices (see Appendix C.3).

Female followers are more responsive to leadership when the environment is more cooperative.

Figure 6: Exit choices by leader gender



Note: Exit choices take a value between 1 and 5 for each possible stopping decision of Blue and Red. Error bars represent standard errors.

Table 2: Determinants of exit choices by leader gender

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	0.13	0.08	0.08	0.17
	(0.30)	(0.36)	(0.31)	(0.30)
Female follower	0.14	0.01	0.59***	0.58***
	(0.37)	(0.37)	(0.20)	(0.22)
Female leader x Female follower	−0.64	−0.42	−0.16	−0.32
	(0.46)	(0.49)	(0.30)	(0.31)
Constant	3.70***	3.77***	4.18***	3.49***
	(0.24)	(1.29)	(0.24)	(0.99)
Controls	No	Yes	No	Yes
R^2	0.04	0.23	0.08	0.34
Adj. R^2	0.03	0.13	0.07	0.25
Num. obs.	192	192	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

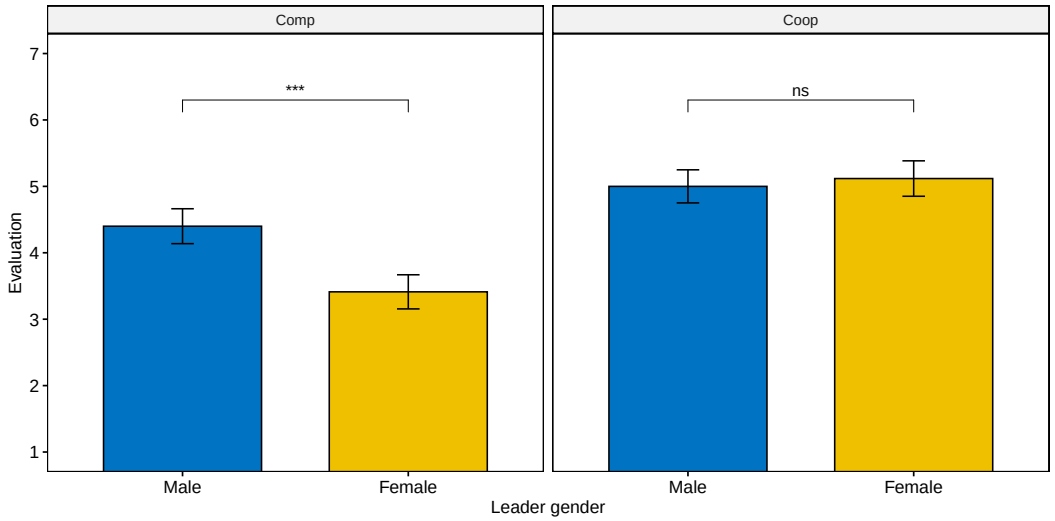
Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the exit choice (between 1 and 5). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

3.3 Leader evaluations and discretionary payments

3.3.1 Leader evaluations

Figure 7 presents the average evaluation scores (reported on a scale from 1 to 7, from "not effective at all" to "extremely effective") received by male and female leaders. A key insight is that, in *Comp*, the average evaluation received by female leaders is 22.5% lower than that received by male leaders (M: 4.40 vs. F: 3.41, $p = 0.04$). This gender gap disappears in the more cooperative environment (M: 5.00 vs. F: 5.12, $p = 0.83$). Strikingly, female leaders are evaluated 50% lower in the competitive setting ($p < 0.01$), whereas men are only evaluated 14% lower ($p = 0.12$). Similarly, when we examine the distribution of evaluations, female leaders are more likely to receive evaluations on the lower end of the scale in *Comp* while male leaders are more likely to receive higher evaluations ($p = 0.02$, Kolmogorov-Smirnov test), see Figure D.1 in Appendix D. In *Coop*, we find no significant difference in the distribution of evaluations ($p = 0.78$, Kolmogorov-Smirnov test).

Figure 7: Evaluations by leader gender



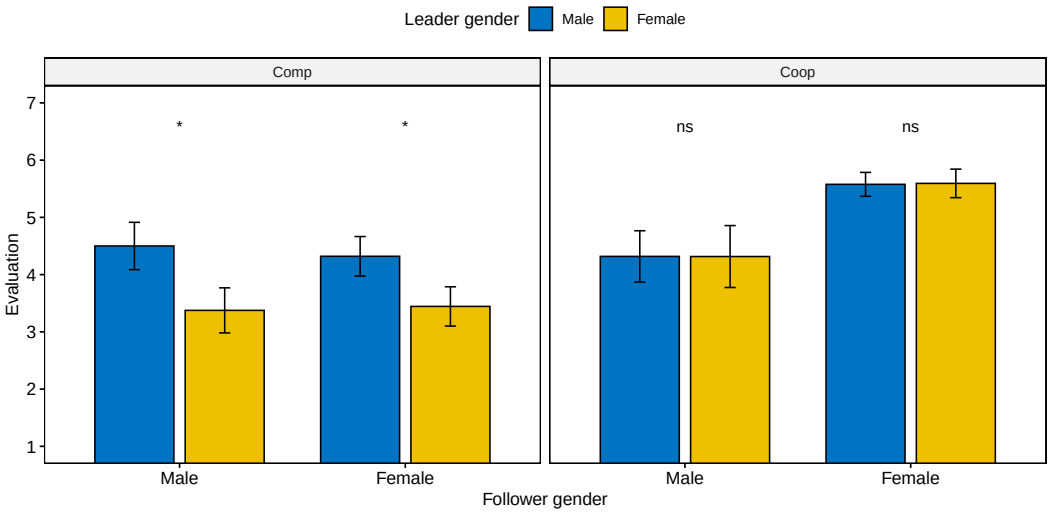
Notes: Evaluations take a value between 1 and 7. Error bars represent standard errors.

Figure 8 shows that the bias against female leaders comes from both male (M: 4.50 vs. F: 3.38, $p = 0.06$) and female (M: 4.32 vs. F: 3.44 , $p = 0.07$) followers, with no significant effect of follower gender (M: 3.89 vs. F: 3.87, p=0.96). The gender gap in evaluations is persistent across followers' colors: Red and Blue players do not differ in their evaluations of leaders.[22] The regression analysis in Table 3 confirms this result and these results hold when we control

[22]In *Comp*, male leaders receive a mean evaluation of 4.09 from Blue and 4.70 from Red ($p = 0.42$), and female leaders receive 3.40 from Blue and 3.42 from Red ($p = 0.81$). Similarly in *Coop*, male (5.04 vs. 4.96, $p = 0.64$) and female leaders (4.96 vs. 5.28, $p = 0.62$) receive similar evaluations from Blue and Red.

for followers' own choices and beliefs (see Tables D.1 and D.2 in Appendix D). Consistent with the previous finding that female followers are more responsive to leadership in *Coop*, female followers also give both male and female leaders significantly higher evaluations in *Coop* than male followers ($p < 0.01$, column 4).

Figure 8: Evaluations by leader and follower gender



Notes: Evaluations take a value between 1 and 7. Error bars represent standard errors.

3.3.2 Leader payment adjustments

Recall that followers can adjust the payment of leaders through a reward or penalty, at a cost of \$0.25 per \$1.00, up to a maximum of \$3.00. Twenty-seven percent of followers chose to incur this cost, with two-thirds of those positively adjusting and one-third negatively adjusting their leader's payment.[23] On average in *Comp*, male leaders receive a positive adjustment of \$0.19 while female leaders receive a negative adjustment of \$0.09, though this difference is not statistically significant ($p = 0.19$). When we examine average payment adjustments by follower gender, we find that this difference is driven by male followers who tend to positively adjust male leaders' payments by \$0.53 but negatively adjust female leaders' payments ny \$0.32. This gender gap of \$0.85 in payments for the leader is statistically significant ($p = 0.04$). We find no such difference for female followers in *Comp* (M: -\$0.05 vs. F: \$0.08, $p = 0.61$). In *Coop*, the average positive adjustment is higher for female leaders

[23] In *Comp*, 19% positively adjust payment and 11% negatively adjust payments. In *Coop*, 18% positively adjust payments and 6% negatively adjust payments. According to a χ^2 test, this difference is not significant ($p = 0.42$).

Table 3: Determinants of leader evaluations

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	-1.13^{*}	-1.28^{**}	-0.00	0.15
	(0.60)	(0.64)	(0.70)	(0.70)
Female follower	-0.18	-0.23	1.26^{**}	1.39^{***}
	(0.65)	(0.66)	(0.53)	(0.52)
Female leader x Female follower	0.25	0.22	0.02	-0.72
	(0.79)	(0.89)	(0.72)	(0.84)
Constant	4.50^{***}	7.95^{***}	4.32^{***}	2.62
	(0.50)	(2.29)	(0.53)	(1.75)
Controls	No	Yes	No	Yes
R^2	0.07	0.35	0.12	0.33
Adj. R^2	0.04	0.14	0.09	0.11
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

but not significantly so (M: \$0.00 vs. F: \$0.19, $p = 0.31$).[24]

In Table 4, we pool *Comp* and *Coop* and examine whether followers make a positive or negative payment adjustment (extensive margin, columns 1 and 2) and the average payment adjustment amount (intensive margin, columns 3 and 4). The negative interaction term shows that male followers are more likely to make a negative payment adjustment for female leaders ($p = 0.03$, column 2). Similarly, on the intensive margin, male followers give female leaders significantly more negative adjustments as compared to male leaders ($p = 0.05$, column 4).

Additionally, leader evaluations positively predict adjustments at both the extensive ($p < 0.01$, column 2) and intensive margins ($p = 0.01$, column 4). This provides support for the validity of leader evaluations as a measure of followers' assessments of their leaders and lends further credibility to our data on evaluations. We also observe a significantly positive relationship between followers' beliefs about when the game will end and the choice to adjust the leader's payment ($p = 0.06$, column 2) as well as the adjustment amount ($p = 0.04$, column 4). Our findings on the evaluation of leaders are summarized as follows:

[24]In *Coop*, male followers tend to positively adjust payments to both male and female leaders (M: \$0.15 vs. F: \$0.12, $p = 1.00$), while female followers negatively adjust the payments of male leaders on average (-\$0.15) but reward female leaders (\$0.23), but this difference is not significant ($p = 0.17$).

Table 4: Determinants of payment adjustments

	Payment adjustment		Adjustment amount	
	(1)	(2)	(3)	(4)
Leader evaluation	0.08***	0.06***	0.13***	0.10**
	(0.02)	(0.02)	(0.04)	(0.04)
Coop	−0.08	−0.07	−0.20	−0.18
	(0.07)	(0.07)	(0.13)	(0.14)
Female leader	0.15*	0.14*	0.26*	0.22
	(0.08)	(0.08)	(0.15)	(0.15)
Male follower	0.25**	0.23**	0.46***	0.47**
	(0.10)	(0.10)	(0.17)	(0.19)
Belief	0.02	0.04*	0.06	0.09**
	(0.02)	(0.02)	(0.04)	(0.05)
Female leader x Male follower	−0.32**	−0.30**	−0.57**	−0.55*
	(0.14)	(0.14)	(0.27)	(0.28)
Constant	−0.45***	−0.38	−1.00***	−0.80
	(0.15)	(0.42)	(0.31)	(0.82)
Controls	No	Yes	No	Yes
R^2	0.12	0.30	0.10	0.24
Adj. R^2	0.09	0.19	0.07	0.11
Num. obs.	195	195	195	195

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the probability of a payment adjustment which =1 if the adjustment is positive, =-1 if the adjustment is negative, and =0 if no adjustment is made (columns 1 and 2) and the adjustment amount is between -$3 and $3 (columns 3 and 4). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Result 2: Female leaders are evaluated less favorably than male leaders in *Comp*, but not in *Coop*. Evaluations of female leaders are 50% lower in the competitive setting while for evaluations of male leaders do not depend on the environment. Relatedly, male followers tend to, on average, reward male leaders but penalize female leaders in *Comp*.

3.4 Potential mechanisms underlying the gender gap in evaluations

In this section, we investigate potential explanations for the gender gap in leader evaluations in *Comp*, despite there being no gender difference in effectiveness. We first investigate whether beliefs about the effectiveness of male and female leaders can explain this gap

(Section 3.4.1). Second, we examine leaders' messages to see if gender differences exist in the message length and whether this has an impact on evaluations (Section 3.4.2). Third, we explore the content of leaders' messages (Section 3.4.3) and test whether male and female leaders differ in the types of advice given, and the impact of advice on evaluations. We find that none of these mechanisms explain the observed gender gap. We do find evidence, however, that followers' perceptions about leaders' advice differ by gender (Section 3.4.4). Female leaders in *Comp* are systematically penalized in evaluations regardless of the advice given. Our data suggests that women receive this additional penalty when their advice is perceived to be inconsistent with traditional gender stereotypes.

3.4.1 Do beliefs about leader effectiveness differ by leader gender?

Given that male and female leaders are equally effective, we investigate whether the gender gap in evaluations is driven by *beliefs* about leader effectiveness. Recall that to elicit beliefs regarding exit nodes, we ask participants when they believe the game will end. Overall, beliefs regarding exit nodes are reasonably accurate in predicting no significant difference between male and female leaders in *Comp* (M: 6.23 vs. F: 6.05, $p = 0.63$) and *Coop* (M: 7.42 vs. F: 7.24, $p = 0.55$).[25] Participants also correctly predict later exits in the cooperative setting for both male ($p < 0.01$) and female leaders ($p < 0.01$). Similarly, in the regression analysis (Table 5), we do not find any evidence that participants anticipate different exit decisions for male and female leaders in *Comp* ($p = 0.31$, column 2) or *Coop* ($p = 0.77$, column 4). These results are consistent under an alternative measure of beliefs about leader effectiveness by taking the difference between beliefs with and without a leader (see Table E.1 in Appendix E).

3.4.2 Does the length of leaders' messages differ by leader gender?

On average, the messages sent by leaders contained 22.24 words in *Comp* and 20.14 words in *Coop*. The number of words in the leader's message is positively correlated with evaluations in *Comp* ($p < 0.01$, column 2, Table 6) but does not predict evaluations in *Coop* ($p = 0.11$, column 4). We find no significant gender difference in the message length in *Comp* (M: 27.74 vs. F: 17.56, $p = 0.23$) and in *Coop* (M: 21.38 vs. F: 19.00, $p = 0.96$). We therefore find no evidence that the gender difference in leader evaluations in *Comp* is driven by the length of leaders' messages.

[25]This result holds when we only examine leaders' beliefs in *Comp* (M: 6.22 vs. F: 5.87, $p = 0.43$) and *Coop* (M: 7.33 vs. F: 7.33, $p = 0.40$).

Table 5: Determinants of followers' beliefs about exit nodes

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.35	−0.16	−0.09	0.59
	(0.66)	(0.64)	(0.71)	(0.80)
Female follower	0.55	0.54	0.45	0.59
	(0.51)	(0.58)	(0.46)	(0.50)
Female leader x Female follower	0.06	0.01	0.05	−0.39
	(0.74)	(0.87)	(0.79)	(0.91)
Constant	5.85***	8.38***	7.09***	7.81***
	(0.44)	(2.35)	(0.48)	(2.02)
Controls	No	Yes	No	Yes
R^2	0.03	0.19	0.01	0.31
Adj. R^2	0.00	−0.06	−0.02	0.09
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the belief about the exit node (between 1 and 9). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

3.4.3 Does the content of leaders' messages differ by leader gender?

We examine the content of leaders' messages in two ways. First, we investigate whether there are gender differences in the tone of leaders' messages. To test for this, we conduct a sentiment analysis, a natural language processing (NLP) technique which is used to determine whether the emotional tone of text data is positive, negative, or neutral.[26] Each leader's message is given a sentiment score, based on the number of words that are associated with emotions as well as the strength of the emotion, whereby a negative score denotes negative emotions, a positive score represents positive emotions and zero means the message is emotionally neutral. Messages sent by male leaders have a mean sentiment score of 0.17 while for female leaders, the mean sentiment score is 0.11, but this difference is not statistically significant ($p = 0.59$). We also find no significant gender difference when we examine messages in *Comp* and *Coop* separately.[27]

Second, we perform a more detailed analysis of the message content. We employed three coders from Texas A&M University to classify the content of leaders' messages and the explanations given by followers for their evaluations. The coders worked independently

[26]We use the package *sentimentr* in R to conduct the analysis (Rinker, 2019).

[27]In *Comp*, the mean sentiment score is 0.12 for male leaders and 0.02 for female leaders ($p = 0.28$) and in *Coop*, the mean sentiment score is 0.21 for male leaders and 0.20 for female leaders ($p = 0.72$).

Table 6: Effect of message length on leader evaluations

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
# Words	0.03***	0.03***	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Female leader	−0.55	−0.96*	0.33	−0.21
	(0.53)	(0.53)	(0.59)	(0.56)
# Words x Female leader	−0.01	0.00	−0.01	−0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Constant	3.63***	6.65***	4.62***	1.37
	(0.31)	(2.18)	(0.47)	(1.78)
R^2	0.16	0.43	0.02	0.36
Adj. R^2	0.13	0.23	−0.01	0.13
Num. obs.	100	96	100	99

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The control variables are: gender, risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

without knowing the research questions or subjects' decisions in the experiment. The coders were given a summary of the instructions (including comprehension questions) that were provided to participants to aid in their understanding of the game and the implications of leaders' messages (see Appendix J for the instructions), but were not informed about the treatments or leaders' aliases. A list of categories was provided to each coder for each response type, explained in more detail below. We used Cohen's Kappa (Cohen, 1960) to assess interrater reliability, or the degree to which coders are in agreement after accounting for chance agreement. Overall, we find substantial agreement (Kappa values between 0.60-0.80) in almost all categories (see Appendix F for the Cohen's Kappa value for each individual category).[28]

Leaders' messages were classified into four categories: "Profit maximizing" (74%), "Stop early" (18%), "Tailored" (4%), and "No advice" (4%).[29] Profit maximizing advice instructs followers to Always Pass, or to never Stop. Stop early advice encourages followers to choose Stop prior to the final node, despite this being suboptimal for the total group (and the leader's) payoff. Tailored advice consists of different advice for Blue and Red.[30] Messages

[28]Kappa values between 0.40-0.60 generally indicate "moderate" agreement; values between 0.60-0.80 indicate "substantial" agreement; and values between 0.80-1.00 suggest "almost perfect" agreement (Landis and Koch, 1977).

[29]In *Coop*, leaders' messages are classified into the same four categories: Profit maximizing (90%), Stop early (4%), Tailored (0%), and No advice (6%).

[30]An example of tailored advice is: "Blue: stop at round 5. Red always pass."

that fail to give any meaningful advice are classified as No advice.[31] Given the small proportion of Tailored and No advice, we combine all messages that are not Profit maximizing into one category and label this as "Other advice." As a robustness check, we indeed find that followers tend to exit significantly later after receiving Profit maximizing advice as compared to Other advice ($p < 0.01$, column 2), see Appendix G.

Male and female leaders do not differ in the types of advice given in *Coop*, with a vast majority of leaders giving Profit maximizing advice (M: 88% vs. F: 92%, $p = 0.66$, Fisher's exact test). A smaller proportion of leaders in *Comp* give Profit maximizing advice (M: 78% vs. F: 70%, $p = 0.75$, Fisher's exact test) overall and the gender difference is also not significant.

Examining leader evaluations conditional on the type of advice given (Figure 9), we find no gender difference in *Coop* for leaders who give Profit maximizing advice (M: 5.21 vs. F: 5.32, $p = 0.62$) and Other advice (M: 3.50 vs. F: 2.75, $p = 0.51$). However, in *Comp*, female leaders receive lower evaluations than their male counterparts when giving both Profit maximizing (M: 4.49 vs. F: 3.57, $p = 0.05$) and Other advice (M: 4.10 vs. F: 3.00, $p = 0.08$). Also consistent with Result 2, female leaders who give Profit maximizing advice in *Coop* are evaluated 49% higher than their counterparts in *Comp* ($p < 0.01$). On the other hand, male leaders who give Profit maximizing advice receive similar evaluations across the two environments ($p = 0.13$). These results are supported by the regression analysis in Table 7. In *Comp*, female leaders are penalized in evaluations even after controlling for the type of advice ($p = 0.05$, column 2). In *Coop*, leaders who give advice that is not Profit maximizing tend to receive lower evaluations ($p = 0.03$, column 4), while female followers tend to give higher ratings than male followers ($p = 0.02$).

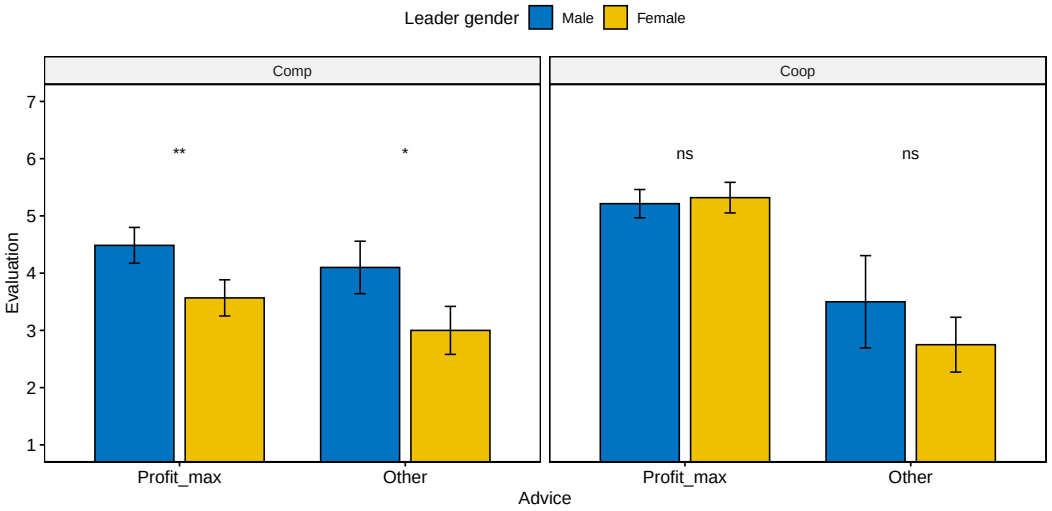
To summarize, we find no evidence that female leaders receive lower evaluations than male leaders due to differences in *how* they lead. The sentiment analysis does not uncover any gender differences in message tone. Further, conditional on giving the same type of advice (Profit maximizing or Other), female leaders still receive lower evaluations than male leaders in the competitive environment. While evaluations of male leaders do not depend on the environment, female leaders receive substantially lower evaluations in the competitive setting for the *same* advice.

3.4.4 Do followers' perceptions about leader's messages differ by leader gender?

In the post-experiment survey, followers were asked to explain their evaluation of their leader, which was then classified by the coders as rating the leader's advice to be "Good," "Bad,"

[31] An example of no advice is: "Do what you think is best."

Figure 9: Leader evaluations by advice and leader gender



Notes: The evaluation of the leader takes a value between 1 and 7. Error bars represent standard errors.

"Selfish," or "Other."[32] Table 8 examines the relationship between the evaluation and evaluation reason. As expected, relative to advice that is deemed to be Good, followers provide lower evaluations for leaders when the advice is judged to be Bad ($p < 0.01$, column 2), Selfish ($p < 0.01$) or Other ($p = 0.07$). However, it is striking that female leaders receive an *additional* gender-based penalty relative to male leaders, when their advice is deemed to be Bad advice or Selfish advice ($p = 0.04$ and $p = 0.02$, respectively, column 2), relative to male leaders. In particular, in the competitive setting, female leaders receive lower evaluations than male leaders for giving Bad (F: 2.42 vs. M: 3.29, $p = 0.06$) or Selfish advice (F: 2.22 vs. M: 4.00, $p = 0.02$). However, we do not observe this in *Coop* for Bad (F: 3.27 vs. M: 3.67, p=0.57) or Selfish advice (F: 5.00 vs. M: 4.00, $p = 0.76$).

Taken together, the results show that the gender gap in evaluations cannot be explained by differences in actual leader effectiveness, expectations about effectiveness, differences in the length of leaders' messages, differences in the tone of the message, nor by differences in the type of advice given by male and female leaders. Instead, our results suggest that gender stereotypes play a key role as the bias against female leaders in the competitive setting persists irrespective of the content of the leader's message. Female leaders tend to receive an additional penalty in evaluations when their advice is perceived to be Bad or Selfish. This

[32]The following question was asked of followers: "Please explain your evaluation of your group Leader." The evaluation is coded as Good advice (if the follower described the leader as having good, clear or persuasive advice), Bad advice (insufficient, bad, unclear or not persuasive advice), Selfish advice (disproportionately benefiting the leader), or Other advice. See Appendix J for instructions given to coders and sample responses. As a robustness check, Figure H.1 in Appendix H shows that Profit maximizing advice is more likely to be deemed Good as compared to Other advice.

Table 7: Determinants of leader evaluations by message type

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	-0.92^{**}	-0.95^{**}	-0.02	-0.32
	(0.47)	(0.48)	(0.35)	(0.39)
Other advice	-0.39	-0.11	-1.66^{*}	-1.60^{**}
	(0.55)	(0.53)	(0.87)	(0.75)
Female follower	-0.07	-0.13	1.12^{***}	0.85^{**}
	(0.38)	(0.39)	(0.34)	(0.35)
Female leader x Other advice	-0.18	-0.83	-0.45	-0.57
	(0.74)	(0.69)	(0.99)	(1.05)
Constant	4.52^{***}	8.03^{***}	4.60^{***}	2.29
	(0.41)	(2.19)	(0.39)	(1.78)
Controls	No	Yes	No	Yes
R^2	0.08	0.37	0.21	0.40
Adj. R^2	0.04	0.16	0.18	0.20
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The baseline advice is Profit maximizing advice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

suggests the incongruity between stereotypes about gender and leadership roles (Eagly and Karau, 2002) causes female leaders to be punished more harshly for acting in a way that is inconsistent with normative beliefs about gender.

3.5 Willingness to Lead

Next, we turn to the supply side of the gender leadership gap and ask if there is a gender difference in the willingness to lead and whether this is contingent on the institutional environment. We find that men are more willing to lead, regardless of the environment, as shown in Figure 10: *Comp* (M: 6.88 vs. F: 5.75, $p = 0.03$) and *Coop* (M: 7.82 vs. F: 6.06, $p < 0.01$). In contrast to our expectations, our results show that women do not report a higher willingness to lead even when the environment is cooperative (*Coop*: 6.06 vs. *Comp*: 5.75, $p = 0.57$).

We examine WTL for men and women separately in the regression analysis in Table 9. While there is some evidence that men are more willing to become leaders in the cooperative environment, this is not the case for women. When we pool both male and female participants

Table 8: Determinants of leader evaluations by evaluation reason

	(1)	(2)
Coop	0.45**	0.35
	(0.22)	(0.23)
Female follower	0.50**	0.43*
	(0.22)	(0.23)
Female leader	0.01	−0.04
	(0.24)	(0.26)
Bad advice	−2.14***	−1.92***
	(0.40)	(0.39)
Selfish advice	−1.76***	−1.62***
	(0.51)	(0.52)
Other advice	−2.29**	−2.27*
	(1.04)	(1.22)
Female leader x Bad advice	−0.75	−1.01**
	(0.49)	(0.50)
Female leader x Selfish advice	−1.32**	−1.31**
	(0.63)	(0.62)
Female leader x Other advice	0.78	0.93
	(1.21)	(1.37)
Constant	5.19***	5.81***
	(0.27)	(1.11)
Controls	No	Yes
R^2	0.51	0.55
Adj. R^2	0.49	0.47
Num. obs.	195	195

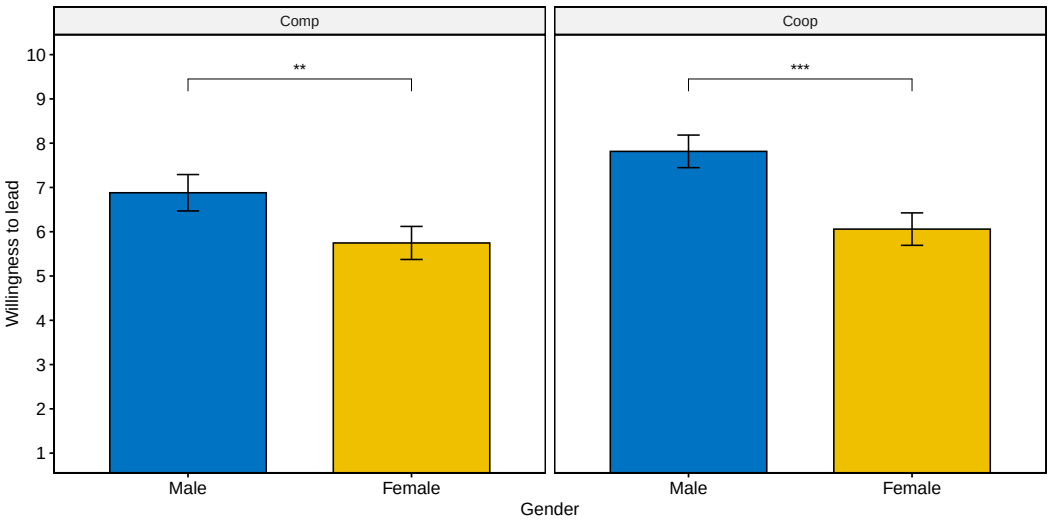
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The baseline treatment is *Comp* and the baseline evaluation is Good advice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

in the regression analysis, we find that willingness to lead is significantly lower for women ($p = 0.03$, column 6), but we do not find a significant interaction between the environment and gender ($p = 0.28$, column 6). We summarize our findings on willingness to lead as follows:

Result 3: Women, as compared to men, report a lower willingness to lead in both environments and are not more willing to lead in the cooperative environment. There is some evidence that men are more willing to lead in the cooperative environment.

Figure 10: Willingness to Lead (WTL) by gender



Note: Willingness to lead takes a value between 1 and 10. Error bars represent standard errors.

3.5.1 Why does reported willingness to lead differ by gender?

To further investigate the gender gap in WTL, we investigate participants' survey responses explaining their reported willingness to become leaders. Subjects' explanations for their stated WTL were classified by the coders into one of six categories: 1) Do not want to lead due to a lack of ability ("No ability"), 2) Do not want to lead due to a belief that payoffs would be lower ("No payoff"), 3) Indifference ("Indifferent"), 4) Want to lead due to a belief that payoffs would be higher ("Yes payoff"), 5) Want to lead due to ability ("Yes ability"), and 6) Other reasons ("Other").[33] Figure 11 summarizes the proportion of each category for men and women. Among women, 34% stated that they did not wish to be the leader due to a perceived lack of ability while only 14% of men gave the same response. Conversely, 35% of men stated that they wanted to be the leader because the leader's payoffs are higher and/or less risky while only 20% of women offered this explanation. According to a χ^2 test, the reasons given by male participants are significantly different from those given by female participants ($p < 0.01$).[34] Our results suggest that not only do stereotypes play a key role in the evaluation of female leader by others, *self*-stereotyping is also evident in individuals' perceptions of their own leadership abilities.

[33]Participants answered the following question: "You stated that your enthusiasm for becoming the Leader was [WTL] out of 10. Please explain your answer."

[34]The results are similar when we separately examine *Comp* and *Coop*. In *Comp*, 15% of men and 33% of women cite "No ability" and 46% of men and 24% of women cite "Yes payoff" ($p = 0.03$). In *Coop*, 12% of men and 36% of women cite "No ability" and 23% of men and 15% of women cite "Yes payoff" ($p < 0.01$).

Table 9: Determinants of Willingness to Lead

	Female		Male		All	
	(1)	(2)	(3)	(4)	(5)	(6)
Coop	0.31	0.04	0.93*	0.93*	0.93	0.99*
	(0.52)	(0.58)	(0.55)	(0.56)	(0.57)	(0.59)
Female					−1.13**	−1.25**
					(0.54)	(0.56)
Coop x Female					−0.62	−0.85
					(0.77)	(0.80)
Constant	5.75***	4.32	6.88***	16.60***	6.88***	16.94***
	(0.38)	(5.78)	(0.39)	(4.51)	(0.40)	(5.11)
Controls	No	Yes	No	Yes	No	Yes
R^2	0.00	0.09	0.02	0.27	0.06	0.13
Adj. R^2	−0.00	−0.06	0.01	0.11	0.05	0.04
Num. obs.	163	163	132	132	295	295

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors in parentheses. The dependent variable is the reported willingness to lead (between 1 and 10). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

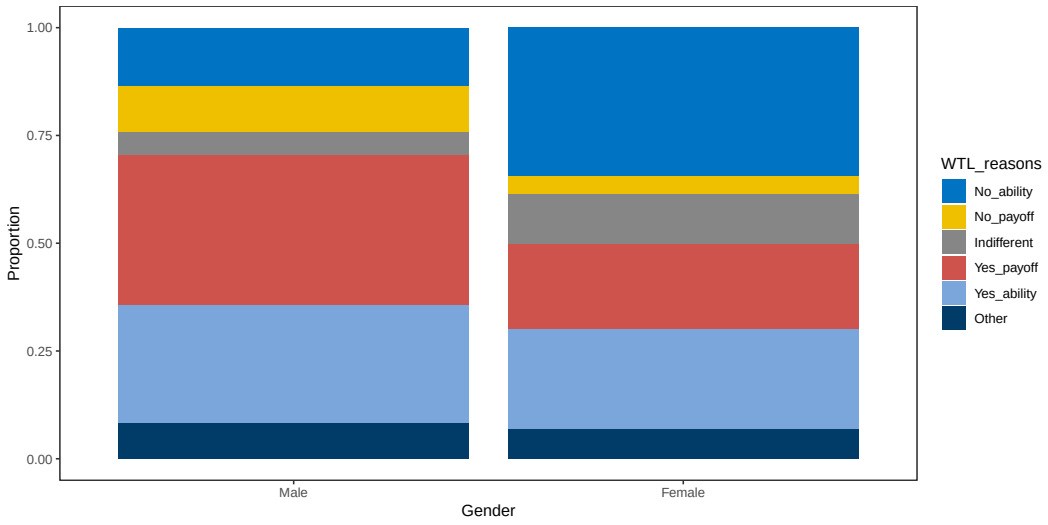
3.5.2 Does leaders' willingness to lead play a role in leadership outcomes?

In the context of our experiment, we made the conscious decision to use a random mechanism to select leaders in order to achieve a gender balance of leaders and to isolate the causal effect of gender in leadership outcomes. In reality, leaders are chosen with some input from the candidates themselves. We use reported WTL as a proxy to identify individuals who might be more likely to seek out and be selected for leadership roles and explore whether WTL is correlated with leader effectiveness and evaluations. In other words, are individuals with a stronger desire to lead necessarily better leaders, and does this desire affect how they are perceived as leaders?

We first explore whether the chosen leader's reported WTL is correlated with their effectiveness. One possibility is that participants who express more interest in leading are more motivated or more suited to leadership and, as a result, are more persuasive in their messages. We do not find evidence of this in either environment (see Table 10) as the leader's WTL does not have a significant effect on the exit choices of followers ($p = 0.40$ and $p = 0.63$, columns 2 and 4).

Next, we investigate whether any correlations exist between WTL and leader evaluations. While there are many possible explanations for why reported WTL may differ across participants, the general desire for leadership may be reflected in leaders' messages. This,

Figure 11: Willingness to Lead reasons



in turn, could affect followers' assessments of their leaders, for example, if they perceive differences in confidence.

Overall, there is no significant correlation between WTL and leader evaluations for female leaders in either environment (see Table I.2 in Appendix I). However, when we examine this correlation for male leaders only (Table I.3), we find that male leaders who are more willing to lead are evaluated better in *Comp* while the sign reverses in *Coop*, with higher WTL more likely to be penalized in evaluations. One potential explanation is that men who are more willing to lead are more likely to send messages that are consistent with traditional leader stereotypes (e.g., that leaders are more self-confident and authoritative). These qualities are congruent with the more competitive environment and may result in better evaluations; however, the same qualities could create a discord in the more cooperative environment and could lead to less favorable evaluations.

Overall, the greater reluctance of women to become leaders seems to be driven by a perceived lack of ability, despite the fact that men and women are equally effective as leaders and the leader's earnings (i.e., the average of the two followers' earnings plus a small bonus) are less risky than that of followers by construction.[35] Our exploratory analysis of the relationship between willingness to lead and leadership outcomes shows that WTL is not correlated with leader effectiveness. However, our results suggest that for male leaders, a greater WTL is associated with higher evaluations in the competitive setting and lower evaluations in the cooperative environment. Again, these findings are indicative of the

[35]Though leaders do not earn significantly more than followers on average, leader earnings are also not lower than follower earnings. In *Comp*, leaders earn \$20.25 while followers earn \$18.74 on average ($p = 0.18$). In *Coop*, leaders earn an average of \$22.81 while followers earn an average of \$22.16 ($p = 0.25$).

Table 10: Effectiveness by leader's WTL by treatment

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
WTL of leader	0.01	0.05	0.01	−0.02
	(0.03)	(0.05)	(0.03)	(0.03)
Female leader		0.36		−0.33
		(0.46)		(0.33)
Female follower		−0.25		0.43***
		(0.23)		(0.15)
WTL of leader x Female leader		−0.08		0.04
		(0.07)		(0.05)
Constant	3.61***	3.56***	4.42***	3.40***
	(0.21)	(1.34)	(0.18)	(1.01)
Controls	No	Yes	No	Yes
R^2	0.00	0.24	0.00	0.34
Adj. R^2	−0.00	0.13	−0.00	0.24
Num. obs.	192	192	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the exit choice (between 1 and 5). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

important interaction between gender stereotypes and the institutional environment.

4 Conclusion

Many explanations have been proposed to explain the gender leadership gap. One understudied explanation is the role of the institutional environment. We design an experiment using a modified Centipede game to test whether leaders can foster higher productivity under a competitive versus a cooperative environment. We then examine how the environment interacts with leader gender and gender stereotypes. We hypothesize that a cooperative environment may help close the gender leadership gap.

We show that leaders are effective at encouraging later exit choices by their followers, regardless of the environment. This is driven in part by the shifting of followers' beliefs about exit choices. We find no significant difference in the effectiveness of male and female leaders, in both the competitive and cooperative settings. A difference does exist, however, in the evaluation of leaders. Female leaders are evaluated as being significantly worse than male leaders, but only in the competitive environment. In line with our predictions, this penalty against women disappears in the cooperative environment as evaluations of female leaders

increase by 50% while evaluations of male leaders remain unchanged from the competitive setting.

We show that this gender gap in evaluations in the competitive setting is not driven by differences in beliefs about effectiveness, differences in the length of leaders' messages, or by differences in the content of leaders' advice. Instead, we find evidence that stereotypes play a key role in followers' perceptions of their leaders. Female leaders are especially penalized when their advice is perceived to be inconsistent with stereotypes about women. We find that women consistently express a reluctance to lead as compared to their male counterparts, suggesting that stereotypes also matter for leaders' perceptions about their own effectiveness. We thus provide robust evidence on both the demand and supply side of leadership, identify gender differences on both sides and the mechanisms that explain these findings.

Our findings highlight the importance of the institutional environment in fostering misperceptions about female leaders, who despite delivering similar outcomes as male leaders, are assessed as being less effective in the competitive environment. Such biases against female leaders have important practical implications. These include decisions on promotions and pay raises, which often depend on evaluations by others. Organizations wishing to reduce the gender leadership gap should carefully assess whether the organizational culture might be contributing to unfavorable biases against female leaders and consider policies which might help mitigate these biases. Approaches such as the "Lean in" (Sandberg, 2013) suite of policies, often touted as effective, may primarily address the supply-side factors contributing to the gender leadership gap. However, in organizations in which the demand side presents a significant barrier, or when both demand and supply sides have a role to play, implementing such policies could inadvertently lead to unintended consequences. This may include subjecting female leaders to unjustly heightened levels of negative assessments and potential backlash, ultimately discouraging not only the future involvement of these women in leadership positions but also other women in the organization.

Admittedly, not all organizations can easily change their institutional environments. For such organizations, our results emphasize the need to be cognizant of potential biases in the evaluation process. While the incentive structure is just one facet of culture, a promising avenue for future research is to examine the role of other contributing factors such as the process of generating ideas, the organizational structure, and conflict resolution processes. While a major advantage of the random selection of leaders in our experiment is the ability to isolate the role of gender in leadership outcomes, an interesting question for future research is whether the interaction between the environment and leadership outcomes depends on the leader selection mechanism and whether other mechanisms such as selection by merit, voting by followers, or by willingness to lead, could lead to different outcomes.

References

AAUW (2022). Barriers and bias: The status of women in leadership. https://www.ncgs.org/wp-content/uploads/2017/11/Barriersand-Bias-The-Status-of-Women-in-Leadership.pdf.

Bilén, D., Dreber, A., and Johannesson, M. (2021). Are women more generous than men? a meta-analysis. *Journal of the Economic Science Association*, 7(1):1–18.

Bohnet, I. (2016). *What works*. Harvard university press.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–73.

Brandts, J., Cooper, D. J., and Fatas, E. (2007). Leadership and overcoming coordination failure with asymmetric costs. *Experimental Economics*, 10:269–284.

Brandts, J., Cooper, D. J., and Weber, R. A. (2015). Legitimacy, communication, and leadership in the turnaround game. *Management Science*, 61(11):2627–2645.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton university press.

Carpenter, J., Matthews, P. H., and Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review*, 100(1):504–17.

Cason, T. N., Gangadharan, L., and Grossman, P. J. (2022). Gender, beliefs, and coordination with externalities. *Journal of Public Economics*, 214:104744.

Cassar, A. and Rigdon, M. L. (2021a). Option to cooperate increases women's competitiveness and closes the gender gap. *Evolution and Human Behavior*, 42(6):556–572.

Cassar, A. and Rigdon, M. L. (2021b). Prosocial option increases women's entry into competition. *Proceedings of the National Academy of Sciences*, 118(45):e2111943118.

Catalyst (2022). Women in management: Quick take. https://www.catalyst.org/research/women-in-management/.

Chakraborty, P. and Serra, D. (2023). Gender and Leadership in Organisations: the Threat of Backlash. *The Economic Journal*, 134(660):1401–1430.

Charness, G., Feri, F., Meléndez-Jiménez, M. A., and Sutter, M. (2023). An Experimental Study on the Effects of Communication, Credibility, and Clustering in Network Games. *The Review of Economics and Statistics*, 105(6):1530–1543.

Charness, G. and Grieco, D. (2023). Creativity and corporate culture. *The Economic Journal*, 133(653):1846–1870.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree — An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Cooper, D. J. and Weber, R. A. (2020). Recent advances in experimental coordination games. *Handbook of experimental game theory*, pages 149–183.

Cox, T. H., Lobel, S. A., and McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal*, 34(4):827–847.

Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, 58(11):1982–2000.

Dong, K., Montero, M., and Possajennikov, A. (2018). Communication, leadership and coordination failure. *Theory and Decision*, 84(4):557–584.

Duckworth, A. (2016). *Grit: The Power of Passion and Perseverance*. Ebury.

Eagly, A. H. (1987). Reporting sex differences. *American Psychologist*, 42(7):756–757.

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

Eagly, A. H., Karau, S. J., and Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, 117(1):125.

Eckel, C., Gangadharan, L., Grossman, P. J., and Xue, N. (2021). The gender leadership gap: Insights from experiments. In *A Research Agenda for Experimental Economics*. Edward Elgar Publishing.

Eckel, C. C. and 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. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1:1061–1073.

Erkal, N., Gangadharan, L., and Koh, B. H. (2023). Do women receive less blame than men? Attribution of outcomes in a prosocial setting. *Journal of Economic Behavior and Organization*, 210:441–452.

Erkal, N., Gangadharan, L., and Xiao, E. (2022). Leadership selection: Can changing the default break the glass ceiling? *The Leadership Quarterly*, 33(2):101563.

Exley, C. L. and Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381.

Gächter, S. and Renner, E. (2018). Leaders as role models and 'belief managers' in social dilemmas. *Journal of Economic Behavior and Organization*, 154:321–334.

Gamba, A. and Regner, T. (2019). Preferences-dependent learning in the centipede game: The persistence of mistrust. *European Economic Review*, 120:103316.

García-Pola, B., Iriberri, N., and Kovářík, J. (2020a). Hot versus cold behavior in centipede games. *Journal of the Economic Science Association*, 6(2):226–238.

García-Pola, B., Iriberri, N., and Kovářík, J. (2020b). Non-equilibrium play in centipede games. *Games and Economic Behavior*, 120:391–433.

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

Goldin, C. (2021). *Career and family: Women's Century-Long Journey Toward Equity*. Princeton University Press.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Grossman, P. J., Eckel, C., Komai, M., and Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior and Organization*, 161:197–215.

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

Heursen, L., Ranehill, E., and Weber, R. A. (2023). Are women less effective leaders than men? Evidence from experiments using coordination games. Working paper.

Krockow, E. M., Colman, A. M., and Pulford, B. D. (2016). Cooperation in repeated interactions: A systematic review of centipede game experiments, 1992–2016. *European Review of Social Psychology*, 27(1):231–282.

Krockow, E. M., Colman, A. M., and Pulford, B. D. (2018). Dynamic probability of reinforcement for cooperation: Random game termination in the centipede game. *Journal of the Experimental Analysis of Behavior*, 109(2):349–364.

Kruse, K. (2013). What is leadership. *Forbes magazine*, 3:1–3.

Kuhn, K. M. (2009). Compensation as a signal of organizational culture: The effects of advertising individual or collective incentives. *The International Journal of Human Resource Management*, 20(7):1634–1648.

Kuhn, P. and Villeval, M. C. (2015). Are women more attracted to co-operation than men? *The Economic Journal*, 125(582):115–140.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Leibbrandt, A., Gneezy, U., and List, J. A. (2013). Rise and fall of competitiveness in individualistic and collectivistic societies. *Proceedings of the National Academy of Sciences*, 110(23):9305–9308.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59.

McKelvey, R. D. and Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica: Journal of the Econometric Society*, 60(4):803–836.

McKinsey & Company (2022). Women in the workplace. https://womenintheworkplace.com/.

Mengel, F., Sauermann, J., and Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2):535–566.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

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

Palacios-Huerta, I. and Volij, O. (2009). Field centipedes. *American Economic Review*, 99(4):1619–35.

Reuben, E. and 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. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, 57(4):637–655.

Rinker, T. W. (2019). sentimentr: Calculate text polarity sentiment.

Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory*, 25(1):92–100.

Roth, A. E. and Murnighan, J. K. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17(2):189–198.

Sahin, S. G., Eckel, C., and Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, 1(1):100–113.

Sandberg, S. (2013). *Lean In: Women, Work, and the Will to Lead*. New York: Knopf.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, 129(1):101–147.

Shurchkov, O. and van Geen, A. V. (2019). Why female decision-makers shy away from promoting competition. *Kyklos*, 72(2):297–331.

Weber, R., Camerer, C., Rottenstreich, Y., and Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science*, 12(5):582–598.

Wozniak, D. (2016). Gender differences for public good contributions by competitive versus group work environments. Available at SSRN 2824902.

A The Centipede game

Why did we select the Centipede game instead of the many possible cooperative and competitive games that populate the experimental repertoire? For example, experimentalists are accustomed to adapting the public goods game to mimic a remarkable variety of field situations – why not here? While it is not common for experimental papers to contain a detailed discussion of alternative designs, we decided it would be informative to document our thought process behind our design choices.

Our experimental model requires a game that can be varied to have a competitive version or a more cooperative version, but that otherwise looks similar. Most games used to study cooperation are lacking in one or more of the characteristics we required. We wanted a game in which there are potential gains to cooperation, mimicking the effect of successful cooperation in the field, and creating scope for the introduction of leadership to increase productivity and payoffs. We needed a game that could be competitive, in the sense that an action could increase one player's payoff at the expense of another's, but where the players were symmetric. And we needed the cooperative and competitive versions of the games to be as similar as possible.

We started with cooperative games and tried to make them competitive but discovered that games with gains to cooperation are difficult to make competitive. For example, we considered the public goods workhorse, which can easily be made into a pure coordination game by changing the off-diagonal payoffs (see, for example, Sahin et al., 2015), and has symmetric players. But it is not straightforward to devise a competitive parallel of the public goods game or coordination game.

We considered the trust game, but in this game, players are not symmetric because the trustor and the trustee make different decisions under different levels of risk. Therefore, a better strategy seemed to be to start with a game that was competitive, and then make it more cooperative. The widely-used tournament of Niederle and Vesterlund (2007) is very hard to make cooperative, though Cassar and Rigdon (2021a,b) take a step in that direction. They eliminate the gender gap in competition by allowing subjects to split the prize. However, it is hard to alter the game so that there are gains to cooperation.

We turned to the Centipede game, which is inherently competitive, and designed a way for it to be more cooperative. The game is ideal as there is scope for the leader to increase productivity, but at the same time, it can be easily adapted to be either competitive or cooperative. In the competitive version, players are symmetric and can compete based on payoffs. While in the cooperative version, there are gains to cooperation as the choice to cooperate does not merely redistribute a fixed pie. The changes are subtle, but have an

impact on behavior in expected ways, as seen by the experimental results.

B List of aliases

Table B.1: Female, male, non-binary and gender-diverse names

Female	Male	Non-binary	Gender-diverse
Isabella	William	Alex	Lillian
Chloe	Jack	Sam	John
Charlotte	Lachlan	Taylor	Ellie
Olivia	Joshua	Nicky	Andrew
Mia	Thomas	Robin	Stephanie
Emily	Oliver	Sasha	Sebastian
Sophie	Noah	Lee	Stacey
Ruby	Ethan	Blake	Hamish
Amelia	James	Kelly	Abby
Ella	Lucas	Andy	George
Lily	Sam		Alicia
Grace	Ben		Charles
Jessica	Daniel		Alexis
Hannah	Liam		Edward
Emma	Alexander		Jade
Sarah	Ryan		Joel
Lucy	Jacob		Phoebe
Holly	Isaac		Christopher
Eva	Matthew		
Laura	Luke		
Hayley	Jake		
Elizabeth	Nicholas		
Molly	Harry		
Samantha	Oscar		
Caitlin	Dylan		
Claire	Michael		
Anna	Connor		
Gabriella	Joseph		
Stella	Adam		
Eliza	Angus		
Madeline	Henry		
Alice	Nathan		
Angela	Caleb		
Amy	Anthony		
Natalie	Patrick		
Jenny	Scott		

C Exit choices

C.1 Exit choices with and without a leader

According to a Kolmogorov-Smirnov test, the distribution of exit choices is significantly different with a leader than without a leader in both *Comp* ($p < 0.01$) and *Coop* ($p < 0.01$), see Figure C.1.

Figure C.1: Distribution of exit choices with and without a leader

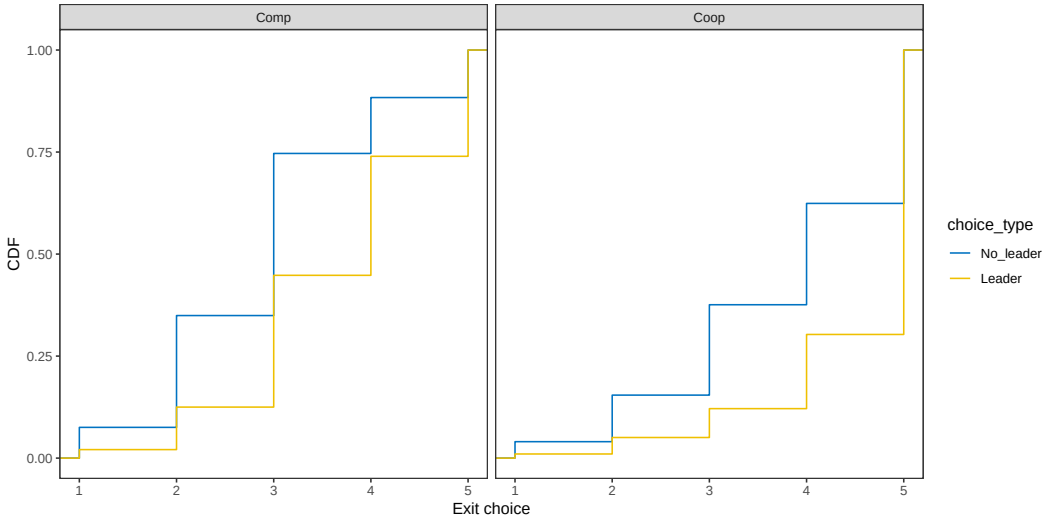


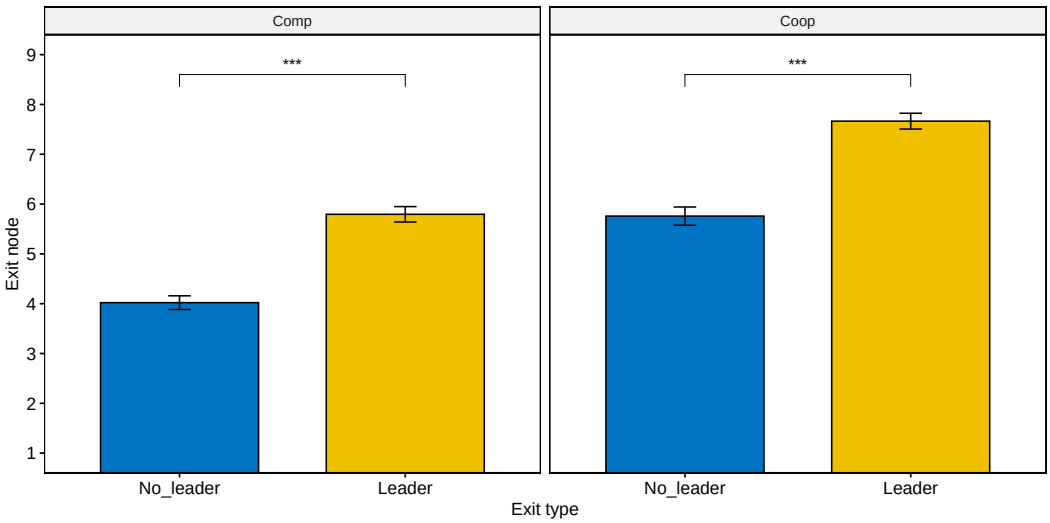
Figure C.2 shows that leaders have a significantly positive effect on exit choices in both *Comp* (no leader: 4.02 vs. leader: 5.79, $p < 0.01$) and *Coop* (no leader: 5.76 vs. leader: 7.66, $p < 0.01$). On average, the presence of a leader encourages subjects to select a later exit node in both *Comp* and *Coop* of 44% and 33% respectively.

C.2 Control treatment

In *Comp-Control* and *Coop-Control*, we control for the possibility of experience or a learning effect that is unrelated to the effect of a leader. We find no evidence of this when participants are asked to make a second choice before receiving any feedback in *Comp-Control* ($p = 0.24$, Column 2, Table C.3) or in *Coop-Control* ($p = 0.18$, Column 4).

This absence of a learning effect is also reflected in beliefs (Table C.5) in *Comp-Control* ($p = 0.70$, Column 2) and *Coop-Control* ($p = 0.16$, Column 4). Women appear to be more optimistic in their beliefs in the more cooperative treatment ($p = 0.04$), despite not choosing differently from men ($p = 0.11$).

Figure C.2: Exit nodes with and without a leader



Note: Exit nodes take a value between 1 and 9. Error bars represent standard errors.

C.3 Leader effectiveness using a difference measure

We construct a variable which takes the difference between exit choices with a leader and without a leader, and find that on average leaders increase followers' exit choices by 0.54 in *Comp* (one-tailed t-test, $p < 0.01$) and 0.76 in *Coop* (one-tailed t-test, $p < 0.01$). However, we do not find a significant difference between male and female leaders in *Comp* (0.41 vs. 0.65, $p = 0.20$), or in *Coop* (0.73 vs. 0.79, $p = 0.61$). Table C.6 supports Result 2 that male and female leaders do not differ in effectiveness in *Comp* ($p = 0.14$, column 2) and in *Coop* ($p = 0.83$, column 4).

Table C.1: Determinants of exit choices (ordered probit)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Leader	0.70^{***}	0.72^{***}	0.81^{***}	0.88^{***}
	(0.10)	(0.10)	(0.11)	(0.11)
Female follower	-0.04	-0.05	0.28^{***}	0.15
	(0.10)	(0.10)	(0.10)	(0.11)
Controls	No	Yes	No	Yes
AIC	1390.89	1356.79	1222.29	1203.19
BIC	1415.98	1452.98	1247.53	1295.74
Log Likelihood	-689.45	-655.40	-605.14	-579.60
Deviance	1378.89	1310.79	1210.29	1159.19
Num. obs.	484	484	496	496

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the exit choice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table C.2: Determinants of exit choices by leader gender (ordered probit)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	0.14	0.09	0.02	0.03
	(0.23)	(0.27)	(0.25)	(0.29)
Female follower	0.15	0.01	0.81^{***}	0.88^{***}
	(0.23)	(0.25)	(0.26)	(0.29)
Female leader x Female follower	-0.66^{**}	-0.44	-0.16	-0.45
	(0.31)	(0.36)	(0.36)	(0.41)
Controls	No	Yes	No	Yes
AIC	540.12	535.18	364.11	340.16
BIC	562.92	623.13	387.13	402.64
Log Likelihood	-263.06	-240.59	-175.05	-151.08
Deviance	526.12	481.18	350.11	302.16
Num. obs.	192	192	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the exit choice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table C.3: Determinants of exit choices in Control treatments

	Comp-Control		*Coop-Control*	
	(1)	(2)	(3)	(4)
Second choice	−0.16	−0.16	0.27	0.27
	(0.13)	(0.14)	(0.19)	(0.20)
Female	0.14	−0.02	0.53	0.60
	(0.29)	(0.43)	(0.34)	(0.37)
Constant	3.08***	3.77***	3.08***	2.26
	(0.26)	(1.22)	(0.28)	(2.12)
Controls	No	Yes	No	Yes
R^2	0.01	0.19	0.05	0.38
Adj. R^2	−0.00	0.10	0.04	0.30
Num. obs.	196	196	196	196

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the individual level in parentheses. The dependent variable is the exit choice (between 1 and 5). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table C.4: Determinants of exit choices in Control treatments (ordered probit)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Second choice	−0.16	−0.17	0.21	0.26
	(0.15)	(0.15)	(0.15)	(0.16)
Female	0.13	0.14	0.38**	0.36**
	(0.15)	(0.17)	(0.15)	(0.18)
Controls	No	Yes	No	Yes
AIC	581.70	581.19	601.19	557.27
BIC	601.37	620.53	620.86	622.83
Log Likelihood	−284.85	−278.60	−294.60	−258.63
Deviance	569.70	557.19	589.19	517.27
Num. obs.	196	196	196	196

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the exit choice. The control variables are: risk preferences, age, education, ethnicity, mother's education, and father's education.

Table C.5: Determinants of beliefs about exit choices

	Comp-Control		*Coop-Control*	
	(1)	(2)	(3)	(4)
Second choice	−0.16	−0.16	0.65	0.65
	(0.43)	(0.42)	(0.52)	(0.47)
Female	0.05	0.31	1.23**	1.35**
	(0.45)	(0.56)	(0.52)	(0.64)
Constant	5.03***	4.34**	4.35***	−3.73
	(0.43)	(2.15)	(0.45)	(3.02)
Controls	No	Yes	No	Yes
R^2	0.00	0.19	0.07	0.40
Adj. R^2	−0.02	−0.00	0.05	0.24
Num. obs.	98	98	98	98

Notes: Ordinary least squares regression with standard errors clustered at the individual level in parentheses. The dependent variable is the belief about the exit node (between 1 and 9). The baseline belief is the exit node for the first choice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table C.6: Determinants of the difference in choices

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	0.88**	0.63	−0.12	−0.10
	(0.37)	(0.42)	(0.38)	(0.45)
Female follower	0.40	0.15	0.34	0.44
	(0.37)	(0.41)	(0.36)	(0.40)
Female leader x Female follower	−1.04**	−0.79	0.24	0.22
	(0.50)	(0.58)	(0.50)	(0.57)
Constant	0.20	2.33	0.55**	0.64
	(0.27)	(1.49)	(0.26)	(2.06)
Controls	No	Yes	No	Yes
R^2	0.06	0.27	0.04	0.25
Adj. R^2	0.03	0.04	0.01	0.01
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors in parentheses. The dependent variable is the difference in choice with a leader minus the choice without a leader. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

D Evaluations and rewards

Figure D.1: Distribution of evaluations by leader gender

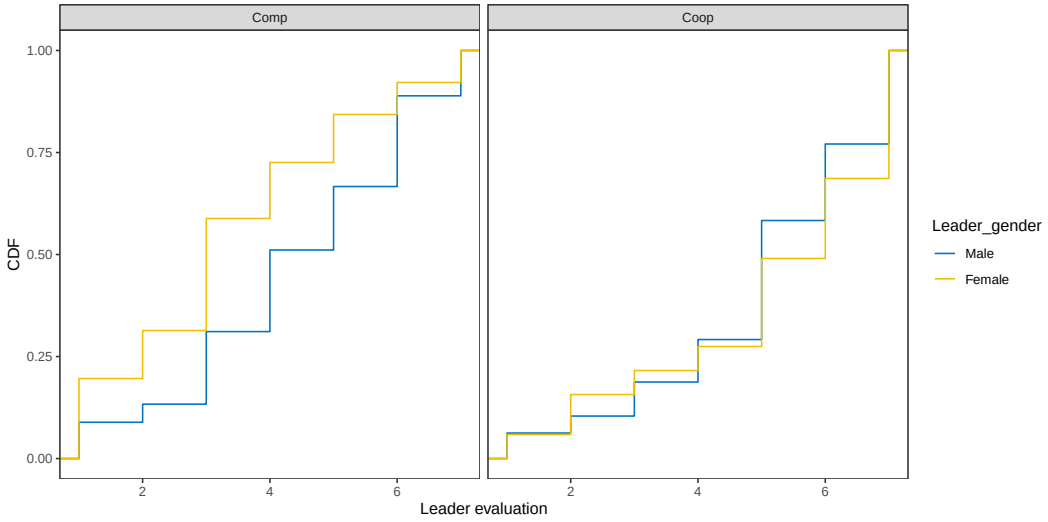
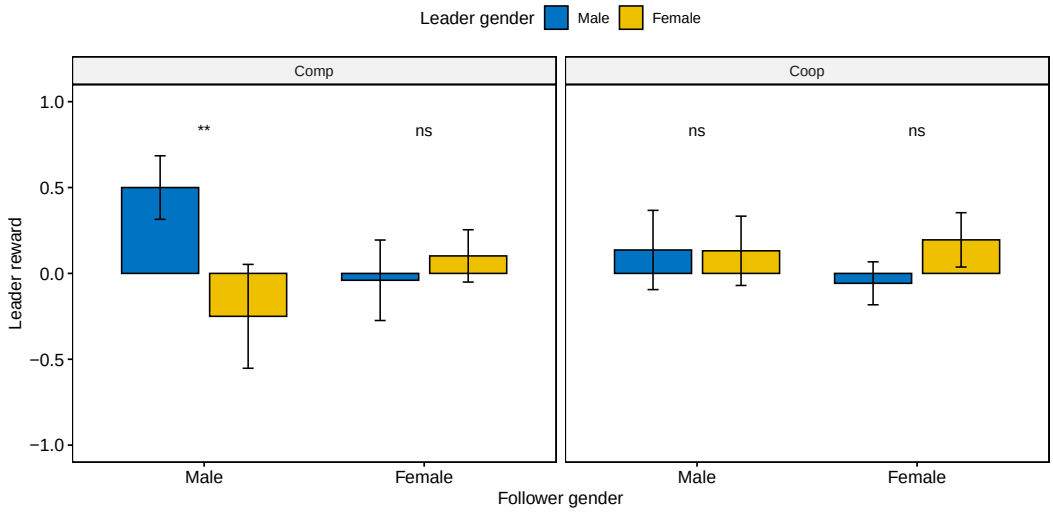


Figure D.2: Rewards by leader and follower gender



Notes: The leader's reward takes a value between -$3 and $3. Error bars represent standard errors.

Table D.1: Determinants of leader evaluations (controlling for choice)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	-1.21^{**}	-1.32^{**}	-0.06	0.04
	(0.55)	(0.54)	(0.58)	(0.58)
Female follower	-0.27	-0.23	0.82	1.01^{**}
	(0.61)	(0.56)	(0.50)	(0.51)
Choice	0.64^{***}	0.55^{***}	0.74^{***}	0.66^{***}
	(0.17)	(0.16)	(0.22)	(0.24)
Female leader x Female follower	0.66	0.45	0.14	-0.51
	(0.76)	(0.76)	(0.63)	(0.73)
Constant	2.11^{***}	5.89^{***}	1.23	0.31
	(0.77)	(2.17)	(0.98)	(2.03)
Controls	No	Yes	No	Yes
R^2	0.20	0.42	0.24	0.40
Adj. R^2	0.18	0.34	0.22	0.31
Num. obs.	192	192	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table D.2: Determinants of leader evaluations (controlling for beliefs)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Belief	0.30***	0.33***	0.28***	0.31***
	(0.11)	(0.09)	(0.09)	(0.11)
Female leader	−1.02*	−1.23**	0.02	−0.03
	(0.59)	(0.60)	(0.63)	(0.60)
Female follower	−0.34	−0.40	1.13**	1.21**
	(0.62)	(0.63)	(0.51)	(0.53)
Female leader x Female follower	0.23	0.22	0.00	−0.60
	(0.75)	(0.83)	(0.68)	(0.77)
Constant	2.75***	5.24**	2.35***	−0.24
	(0.76)	(2.27)	(0.89)	(2.31)
Controls	No	Yes	No	Yes
R^2	0.16	0.44	0.21	0.41
Adj. R^2	0.12	0.24	0.18	0.21
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table D.3: Determinants of leader evaluations (ordered probit)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.61*	−0.87**	0.08	0.27
	(0.32)	(0.37)	(0.33)	(0.39)
Female follower	−0.08	−0.15	0.64**	0.84**
	(0.31)	(0.35)	(0.31)	(0.36)
Female leader x Female follower	0.12	0.12	−0.01	−0.57
	(0.42)	(0.50)	(0.43)	(0.49)
Controls	No	Yes	No	Yes
AIC	375.72	381.78	359.45	374.67
BIC	398.80	456.15	382.81	452.52
Log Likelihood	−178.86	−161.89	−170.72	−157.33
Deviance	357.72	323.78	341.45	314.67
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the evaluation of the leader. The control variables are: risk preferences, age, education, ethnicity, mother's education, and father's education.

Table D.4: Determinants of rewards (probit)

	(1)	(2)
Leader evaluation	0.40^{***}	0.46^{***}
	(0.09)	(0.13)
Coop	-0.60^{**}	-0.67^{*}
	(0.30)	(0.40)
Female leader	0.90^{**}	1.33^{**}
	(0.40)	(0.59)
Male follower	1.08^{***}	1.54^{**}
	(0.42)	(0.62)
Belief	0.02	0.10
	(0.07)	(0.11)
Female leader x Male follower	-0.96^{*}	-1.28^{*}
	(0.55)	(0.75)
Constant	-3.60^{***}	-9.08
	(0.73)	(376.76)
Controls	No	Yes
AIC	131.99	140.48
BIC	153.65	227.10
Log Likelihood	-59.00	-42.24
Deviance	117.99	84.48
Num. obs.	163	163

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the probability of a positive reward. The baseline Treatment is *Comp*. The control variables are: risk preferences, age, education, ethnicity, mother's education, and father's education.

E Beliefs

Table E.1: Determinants of differences in followers' beliefs under a leader

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	0.15	0.32	−0.42	0.08
	(0.65)	(0.76)	(0.75)	(0.90)
Female follower	−0.35	−0.33	0.14	0.35
	(0.64)	(0.73)	(0.69)	(0.80)
Female leader x Female follower	−0.19	−0.31	0.93	0.58
	(0.88)	(1.05)	(0.98)	(1.13)
Constant	1.35***	6.94**	1.32**	1.18
	(0.48)	(2.72)	(0.51)	(2.85)
Controls	No	Yes	No	Yes
R^2	0.01	0.18	0.03	0.22
Adj. R^2	−0.02	−0.08	−0.00	−0.04
Num. obs.	96	96	99	99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the difference in beliefs about the exit node (between 1 and 9) under a leader. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

Table E.2: Determinants of followers' beliefs about exit nodes (ordered probit)

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.17	−0.11	−0.12	−0.06
	(0.18)	(0.19)	(0.19)	(0.21)
Female follower	0.27	0.27	0.06	−0.09
	(0.18)	(0.19)	(0.20)	(0.21)
Controls	No	Yes	No	Yes
AIC	587.63	616.38	508.21	515.15
BIC	617.46	714.84	538.25	611.28
Log Likelihood	−283.81	−275.19	−244.11	−225.58
Deviance	567.63	550.38	488.21	451.15
Num. obs.	146	146	149	149

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordered probit regression with standard errors clustered at the individual level in parentheses. The dependent variable is the belief about the exit node. The control variables are: risk preferences, age, education, ethnicity, mother's education, and father's education.

F Cohen's Kappa values

Table F.1: Cohen's Kappa - Leaders' messages

Advice categories	Cohen's Kappa
Profit maximizing	0.96
Stop early	0.85
Tailored	0.78
No advice	0.83

Table F.2: Cohen's Kappa - Reasons for evaluations

Evaluation categories	Cohen's Kappa
Good	0.83
Bad	0.75
Selfish	0.62
Other	0.11

Table F.3: Cohen's Kappa - Reasons for reported willingness to lead

WTL categories	Cohen's Kappa
No ability	0.79
No payoff	0.61
Indifferent	0.70
Yes payoff	0.80
Yes ability	0.79
Other	0.52

G Effectiveness by leader advice

Table G.1 shows that groups in which leaders gave any advice other than Profit Maximizing advice are more likely to exit earlier ($p < 0.01$, column 4).

Table G.1: Effectiveness of leaders by advice

	(1)	(2)
Other advice	-1.12^{***}	-1.20^{***}
	(0.23)	(0.24)
Female leader		-0.11
		(0.15)
Female follower		0.08
		(0.14)
Constant	4.29^{***}	3.88^{***}
	(0.08)	(0.81)
Controls	No	Yes
R^2	0.16	0.26
Adj. R^2	0.16	0.21
Num. obs.	390	390

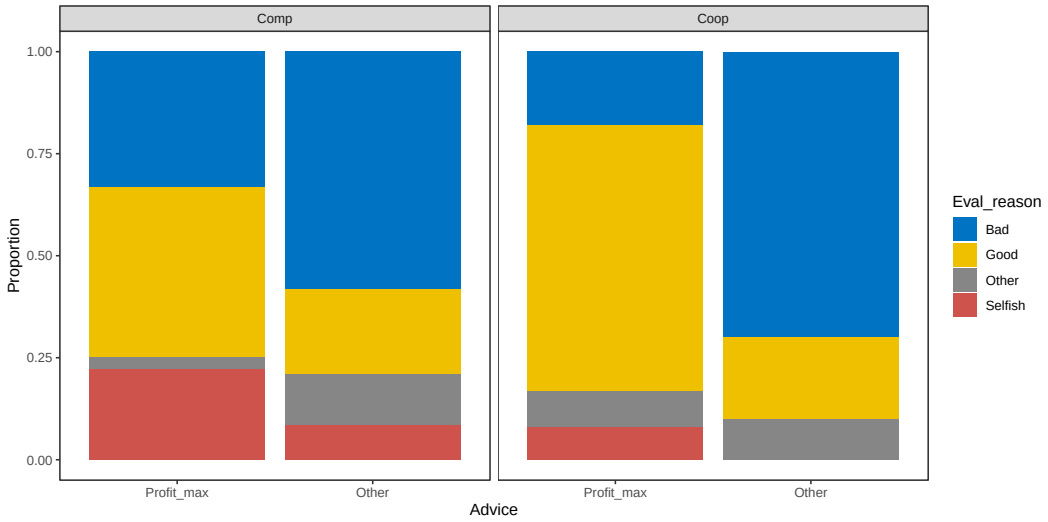
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the exit choice (between 1 and 5). The baseline advice is Profit maximizing advice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

H Reasons behind leader evaluations

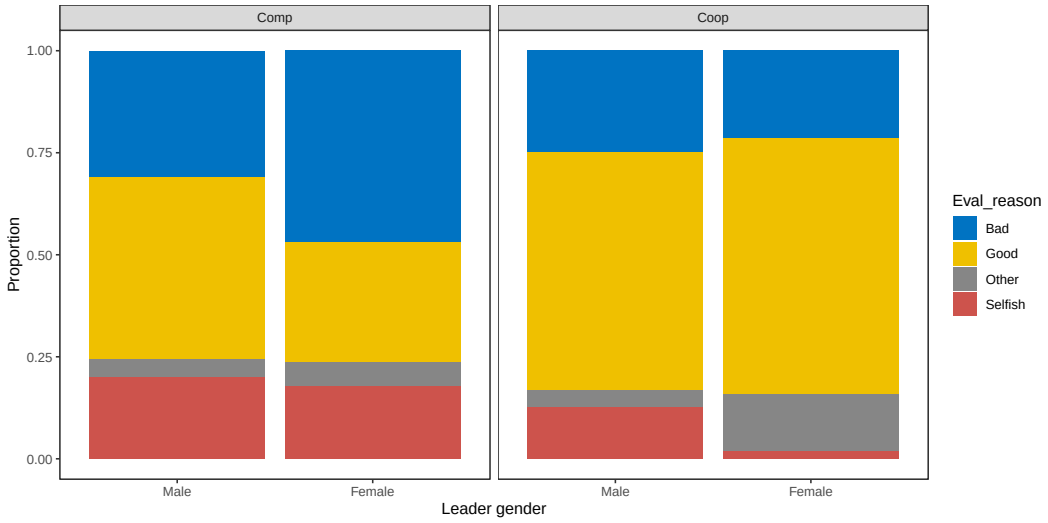
In *Comp*, Profit maximizing advice is more likely to be deemed Good (28% vs. 15%) and Selfish (14% vs. 5%) advice relative to Other advice ($p = 0.03$, Fisher's exact test, Figure H.1). Similarly, in *Coop*, Profit maximizing advice is more likely to be deemed Good (43% vs. 13%, $p < 0.01$, Fisher's exact test).

Figure H.1: Reasons for leader evaluations by advice



A majority of followers rate the leader's advice as Good in *Coop* (Figure H.2) and we find a marginal difference between the reasons given for male and female leaders ($p = 0.10$, Fisher's exact test), with messages by male leaders more likely to be deemed Selfish (M: 13% vs. F: 2%) than those by female leaders (whose advice is more likely to be classified in the Other category, M: 4% vs. F: 13%). In *Comp*, a larger proportion of female leaders are rated as giving Bad advice compared to male leaders (M: 30% vs. F: 46%) while women are less likely to be rated as giving Good advice (M: 43% vs. F: 31%), though these differences are not statistically significant ($p = 0.43$, Fisher's exact test).

Figure H.2: Reasons for leader evaluations by treatment and leader gender



I Willingness to lead

Table I.2: Leader evaluations by WTL by leader gender

	Female leader		Male leader	
	(1)	(2)	(3)	(4)
WTL of leader	0.03	0.02	0.06	0.20**
	(0.07)	(0.11)	(0.06)	(0.08)
Coop		1.22*		3.03***
		(0.73)		(0.81)
Female follower		0.75*		0.03
		(0.38)		(0.44)
WTL of leader x *Coop*		0.03		−0.35***
		(0.11)		(0.12)
Constant	4.09***	2.79	4.27***	4.04*
	(0.45)	(2.63)	(0.49)	(2.36)
Controls	No	Yes	No	Yes
R^2	0.00	0.36	0.01	0.39
Adj. R^2	−0.01	0.17	0.00	0.17
Num. obs.	102	102	93	93

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7).

Table I.3: Leader evaluations for male leaders by WTL by treatment

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
WTL of leader	0.17**	0.33***	−0.11	−0.32***
	(0.07)	(0.13)	(0.08)	(0.09)
Female follower		0.11		0.12
		(1.34)		(1.02)
WTL of leader x Female follower		−0.21		0.16
		(0.18)		(0.12)
Constant	3.37***	8.10**	5.87***	7.84***
	(0.51)	(3.60)	(0.53)	(1.91)
Controls	No	Yes	No	Yes
R^2	0.11	0.63	0.04	0.78
Adj. R^2	0.09	0.27	0.02	0.60
Num. obs.	45	45	48	48

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the evaluation of the leader (between 1 and 7).

J Instructions

J.1 Experimental instructions

Choose a name

During the experiment, no participant can identify any other participant by their decisions or earnings. To maintain anonymity throughout the experiment, please select an alias for yourself.

- ○ Isabella
- ○ Chloe

Next

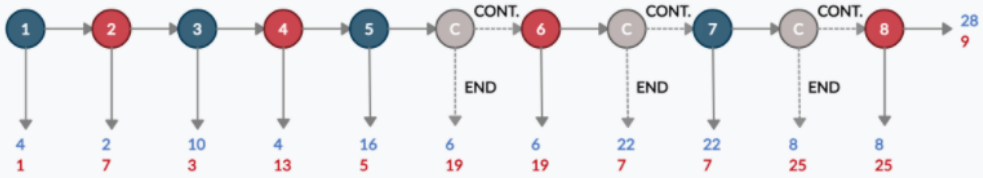
Competitive treatment

1. Your Decisions | 2. The Computer's Role | 3. Examples | 4. Comprehension

Decision making in Part 1A

The Computer's Role

After **round 5**, the Computer will randomly choose between **END** and **CONTINUE**. The Computer will make this choice before each of the remaining rounds. That is, the Computer may end funding for the project **before round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and STOP is chosen.



- If the Computer chooses END **before** either participant chooses STOP, then **the Computer's choice is implemented** and the earnings are shown below the Computer's choice.
- If the Computer chooses END **after** either participant chooses STOP, then **the earliest STOP is implemented** and the earnings are shown below the relevant participant's decision.
- If the Computer chooses **ALWAYS CONTINUE** and both participants choose **ALWAYS PASS**, then the earnings are shown to the right of **round 8**.

You will be informed of the outcome of the project at the end of the experiment.

Part 1A - Summary

You will be randomly assigned to be a **Blue** or **Red** participant. In Part 1A, you will be paired with another participant and as a pair, you will participate in a project which consists of **8 rounds** of decisions, made alternately by Red and Blue.

All you have to do is choose **when to STOP, or to ALWAYS PASS**. This means you are choosing between five options:

1. **STOP** at your **first** decision round (round **1** or **2**)
2. **STOP** at your **second** decision round (round **3** or **4**)
3. **STOP** at your **third** decision round (round **5** or **6**)
4. **STOP** at your **fourth** decision round (round **7** or **8**)
5. **ALWAYS PASS**

After **round 5**, the Computer may randomly choose to **END** funding for the project **before either round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and STOP is chosen. The project will end if either participant chooses STOP, or if the Computer chooses to END funding for the project.

You will be informed of the outcome of the project at the end of the experiment. If Part 1A is selected to be paid, you will receive your earnings from the project.

No leader choices

Part 1A

You have been randomly assigned to be a **Blue** participant and have been matched with a Red participant. Please make your decision for the project.

I choose to:

- ○ STOP at round 1
- ○ STOP at round 3
- ○ STOP at round 5
- ○ STOP at round 7
- ○ ALWAYS PASS

Next

1. The Roles | 2. The Leader | 3. Summary | 4. Comprehension

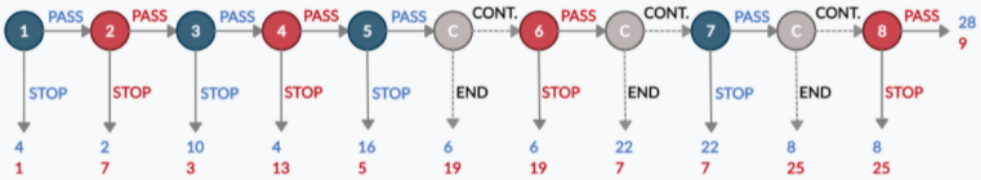
Part 1B - Overview

Summary

You will be assigned to new groups of three: one **Blue** participant, one **Red** participant and one **Leader**. In your new groups you will participate in a project which consists of 8 rounds of decisions, made alternately by Red and Blue.

If you are a Blue or Red participant, all you have to do is choose **when to STOP, or to ALWAYS PASS**. This means that you can choose between five options:

1. **STOP** at your **first** decision round (round **1** or **2**),
2. **STOP** at your **second** decision round (round **3** or **4**),
3. **STOP** at your **third** decision round (round **5** or **6**),
4. **STOP** at your **fourth** decision round (round **7** or **8**), or
5. **ALWAYS PASS** (i.e., never STOP).



After **round 5**, the Computer may randomly choose to **END** funding for the project **before either round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and STOP is chosen. The project will end if either participant chooses STOP, or if the Computer chooses to END funding for the project.

Before Blue and Red make their decisions, the **Leader** will send a message to group members to explain potential strategies. The Leader will not have any decisions to make in the project and will receive the average of the earnings of the Blue participant and the Red participant in the group, plus $3.

You will be informed of the outcome of the project at the end of the experiment. If Part 1B is selected to be paid, you will receive your earnings from the project.

Part 1B

Please indicate on a scale from 1 to 10, how much you want to be the Leader, where 1 means that you do not want to become the Leader at all and 10 means you very much want to become the Leader.

I want to be the Leader of my group:

On the next screen, you will find out whether you have been randomly assigned to be the **Leader**, the **Blue** participant, or the **Red** participant in your group.

Next

Leader

Part 1B - Your role

In Part 1B, you have been assigned to be the **Leader** of your group.

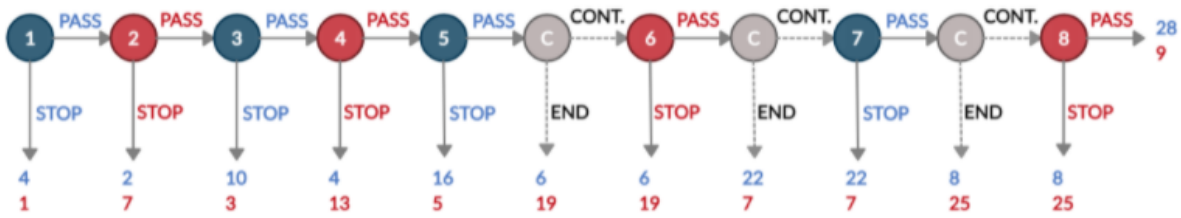
Your earnings will be maximized if your group members choose to **PASS** at every node. Please enter a message for your group.

Please follow the following rules when entering a message:

- You are not permitted to reveal personal information such as your name, age, ethnicity, field of study or other information that could identify you to your group.
- You are also not permitted to use offensive language or be disrespectful.

If you break any of these rules, you will be excluded from the study and you will not receive any earnings.

The figure representing the decisions of Blue and Red is repeated below.



Please enter your message.

Next

Leader beliefs

Part 1B

Please answer the following questions. You may receive an additional payment, depending on your response.

In **Part 1A**, at which round do you believe the project ended for your group (either by a participant or by the Computer)? You will receive an additional $1 if your guess is correct.

In **Part 1B**, at which round do you believe the project ended for your group (either by a participant or by the Computer)? You will receive an additional $1 if your guess is correct.

Please explain your responses above.

Next

Followers' choices

Part 1B - Decision

You have been randomly assigned to be a **Red** participant and have been matched with a Blue participant. Please make your decision for the project.

The Leader, **Chloe**, sent the following message to your group:

"please pass"

I choose to:

- ○ STOP at round 2
- ○ STOP at round 4
- ○ STOP at round 6
- ○ STOP at round 8
- ○ ALWAYS PASS

Next

Follower beliefs

Part 1B

Please answer the following questions. You may receive an additional payment, depending on your response.

In **Part 1A**, at which round do you believe the project ended for your group (either by a participant or by the Computer)? You will receive an additional $1 if your guess is correct.

In **Part 1B**, at which round do you believe the project ended for your group (either by a participant or by the Computer)? You will receive an additional $1 if your guess is correct.

How effective do you believe your group Leader, **Chloe**, was on a scale from 1 to 7 (1 = not effective at all, and 7 = extremely effective)?

Please explain your evaluation of your group Leader.

Next

Leader bonus

Part 1B

You and your group member now have the chance to increase or decrease the earnings of your group's Leader, **Chloe**. Based on your evaluation of the Leader, you can choose between one of seven options. Increasing or decreasing the Leader's earnings is costly, as shown in the table below.

Cost to you	Leader's earnings
$0.75	-$3.00
$0.50	-$2.00
$0.25	-$1.00
$0.00	$0.00
$0.25	+$1.00
$0.50	+$2.00
$0.75	+$3.00

For example, if you choose to **increase** the Leader's earnings by $2.00 (i.e., +$2.00), then you will have to pay $0.50. If you choose to **decrease** the Leader's earnings by $1.00 (i.e., -$1.00), then you will have to pay $0.25. If you do not wish to change the Leader's earnings, you can choose $0.00 at no cost.

Either your decision or your group member's decision will be chosen at random to be implemented for your Leader.

Recall, the Leader's earnings are the average of the earnings of the Blue and Red participant, plus $3. You will be informed of any changes to your Leader's earnings at the end of the experiment.

I choose to change the Leader's earnings by:

Leader survey

Survey

In Part 1A, how did you decide if or when to stop the project?

You stated that your enthusiasm for becoming the leader was 4 out of 10. Please explain your answer.

Please explain the reasoning behind the message you sent to your group.

How effective do you think you were as a Leader?

Any other feedback, suggestions or comments? Did the situation in the experiment remind you of anything?

Next

Follower survey

Survey

In Part 1A, how did you decide if or when to stop the project?

You stated that your enthusiasm for becoming the Leader was 0 out of 10. Please explain your answer.

In Part 1B, how did you decide if or when to stop the project?

Why did you choose to increase, decrease or not change the Leader's earnings?

What do you think was the gender of your leader?

Any other feedback, suggestions or comments? Did the situation in the experiment remind you of anything? Were the instructions clear?

Next

Cooperative treatment

Part 1A - Summary

You will be randomly assigned to be a **Blue** or **Red** participant. In Part 1A, you will be paired with another participant and as a pair, you will participate in a project which consists of **8 rounds** of decisions, made alternately by Red and Blue.

All you have to do is choose **when to STOP, or to ALWAYS PASS**. This means you are choosing between five options:

1. **STOP** at your **first** decision round (round **1** or **2**)
2. **STOP** at your **second** decision round (round **3** or **4**)
3. **STOP** at your **third** decision round (round **5** or **6**)
4. **STOP** at your **fourth** decision round (round **7** or **8**)
5. **ALWAYS PASS**

After **round 5**, the Computer may randomly choose to **END** funding for the project **before either round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and earnings are **shared equally** between Blue and Red. The project will end if either participant chooses STOP, or if the Computer chooses to END funding for the project.

You will be informed of the outcome of the project at the end of the experiment. If Part 1A is selected to be paid, you will receive your earnings from the project.

1. The Roles | 2. The Leader | 3. Summary | 4. Comprehension

Part 1B - Overview

Summary

You will be assigned to new groups of three: one **Blue** participant, one **Red** participant and one **Leader**. In your new groups you will participate in a project which consists of 8 rounds of decisions, made alternately by Red and Blue.

If you are a Blue or Red participant, all you have to do is choose **when to STOP, or to ALWAYS PASS**. This means that you can choose between five options:

1. **STOP** at your **first** decision round (round **1** or **2**),
2. **STOP** at your **second** decision round (round **3** or **4**),
3. **STOP** at your **third** decision round (round **5** or **6**),
4. **STOP** at your **fourth** decision round (round **7** or **8**), or
5. **ALWAYS PASS** (i.e., never STOP).

After **round 5**, the Computer may randomly choose to **END** funding for the project **before either round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and earnings are **shared equally** between Blue and Red. The project will end if either participant chooses STOP, or if the Computer chooses to END funding for the project.

Before Blue and Red make their decisions, the **Leader** will send a message to group members to explain potential strategies. The Leader will not have any decisions to make in the project and will receive the average of the earnings of the Blue participant and the Red participant in the group, plus $3.

You will be informed of the outcome of the project at the end of the experiment. If Part 1B is selected to be paid, you will receive your earnings from the project.

J.2 Instructions for coders

Classification Guidelines

Background: Participants are assigned into groups of three and participate in a project with a total of 8 rounds. Within each group, there is: one Blue participant, one Red participant, and one Leader. Blue and Red each have one decision to make and choose when to Stop, or to Always Pass. This means Blue and Red can each choose between five options:

1. Stop at their first decision round,
2. Stop at their second decision round,
3. Stop at their third decision round,
4. Stop at their fourth decision round, or
5. Always Pass (i.e., never Stop).

The leader sends a single message to the Blue and Red participant in their group to explain potential strategies. The Leader does not have any decisions to make in the project and receives the average earnings of Blue and Red plus a small bonus.

Before round 5, the project may end if either Blue or Red chooses Stop in rounds 1, 2, 3, or 4. **After** round 5, the project may end if either Blue or Red chooses Stop, or if the Computer chooses to end the project, whichever comes first. The Computer randomly decides whether to end the project after round 5, 6, 7, or never, with equal chances. Therefore, the project will end if either participant chooses Stop, or if the Computer chooses the end the project, whichever comes first.

Comprehension questions:

1) The Computer may randomly choose to end the project **after** which round?
    a. Round 4
    b. Round 5
    c. Round 6
    d. Round 7

2) If Blue chooses to Stop at round 7, Red chooses to Always Pass and the Computer chooses to end the project after round 5, when will the project end?
    a. After round 5
    b. After round 6
    c. After round 7
    d. Never

3) The Leader does not make any decisions in the project but can send a message to their group members.
    a. True
    b. False

4) How are the Leader's earnings determined?
    a. The Leader's earnings are the same as the earnings of Blue and Red, plus a small bonus
    b. The Leader's earnings are the sum of the earnings of Blue and Red
    c. The Leader's earnings are the average of the earnings of Blue and Red
    d. The Leader's earnings are the average of the earnings of Blue and Red, plus a small bonus

Advice Categories (L_message): This is the advice message sent from the leader to the 2 players (Blue and Red) before they must decide which of 8 rounds to stop at

- stop_early=1 if the leader advises both players to stop anywhere before round 8 or before the end.
- max_group=1 if the leader says to "go all the way" "stop after round 8", "never stop", or "always pass" etc.
- tailored=1 if the leader gives separate (or tailored) advice to each player, telling Blue to stop at round X and Red to stop at round Y
- no_advice=1 if the leader does not give any specific advice or tells the players to make their own decision

Leader evaluations (Belief_explanation): This variable asks participants to explain the evaluation they gave the leader. The 2 players (Blue and Red) were asked to evaluate the effectiveness of their leader on a scale from 1 to 7 (with 1=not effective at all and 7=very effective).

- eval_goodadvice=1 if the subject says they chose their rating based on the leader having good, clear or persuasive advice.
- eval_badadvice=1 if they think the leader gave insufficient advice, bad advice, unclear advice or advice that was not persuasive
- eval_selfishadvice=1 if the subject indicates that the leader's advice only or disproportionately benefitted the leader or was trying to maximize the leader's own earnings.
- Eval_otheradvice=1 if the evaluation does not fit in any other category

Willingness to Lead Reasoning (WTL_reason): This variable allows participants to explain why they were willing to lead, not willing to lead, or indifferent between being a Leader and Red/Blue. Participants were asked to indicate on a scale from 1 to 10, how much they want to be the Leader (1=they do not want to become Leader at all and 10=they very much want to become the Leader).

- wtl_noability=1 if the subject expresses that they did not want to lead because they doubt their leadership ability, their understanding of the game, their ability to persuade, or dislike leading.
- wtl_nopayoff=1 if the subject has chosen not to lead because they believe it is in their best financial interest to not be a Leader or because they prefer to make a decision as Blue or Red
- wtl_indiff=1 if the subject is indifferent between being the Leader or Blue or Red
- wtl_yespayoff=1 if the subject wanted to lead because they believe it is in their best financial interest to lead, or prefer the role of Leader
- wtl_yeslead=1 if the subject wants to lead because they think they are a good Leader, have high ability, believe they have good ideas that will make them a successful Leader or enjoy leading
- wtl_other=1 for any other reason or if it is left blank