MONASH University

MONASH BUSINESS SCHOOL

Department of Economics
ISSN number 1441-5429

Closing the Gender Leadership Gap: Competitive versus Cooperative Institutions

Discussion Paper no. 2025-17

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

Abstract:

Motivated by the stereotype that women are more cooperative and less competitive, we investigate how the institutional environment impacts the gender leadership gap. An experiment tests leaders' impact on earnings under competitive ("winner take all") versus cooperative (equal earnings distribution) incentive schemes. All leaders enhance efficiency similarly, but a gender gap emerges in the competitive context where women receive lower evaluations for identical advice. This bias disappears in the cooperative context where female leaders are evaluated 50% higher, suggesting that congruence between the environment and gender stereotypes has important policy implications. Men are more willing to lead, regardless of context.

Keywords: gender, leadership, institutio

JEL Classification: C92 D91 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: WU Vienna University of Economics and Business, (email: Nina.Xue@wu.ac.at).

Closing the Gender Leadership Gap: Competitive versus Cooperative Institutions*

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

25 September 2025

Abstract

Motivated by the stereotype that women are more cooperative and less competitive, we investigate how the institutional environment impacts the gender leadership gap. An experiment tests leaders' impact on earnings under competitive ("winner take all") versus cooperative (equal earnings distribution) incentive schemes. All leaders enhance efficiency similarly, but a gender gap emerges in the competitive context where women receive lower evaluations for identical advice. This bias disappears in the cooperative context where female leaders are evaluated 50% higher, suggesting that congruence between the environment and gender stereotypes has important policy implications. Men are more willing to lead, regardless of context.

JEL Classification: C92, D91, J16, J71, M14
Keywords: gender, leadership, institutional environment, performance evaluation, lab experiment, behavioral economics

*We are grateful to participants at several conferences and seminars for helpful comments and to Daniel Stephenson for his valuable assistance. 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
‖WU Vienna University of Economics and Business, Nina.Xue@wu.ac.at

1 Introduction

Women remain underrepresented in leadership roles despite closing, and in many cases surpassing, educational attainment gaps. In 2024, women held just 34% of managerial positions, 32.9% of board seats, and 34.1% of parliamentary seats across OECD countries, even as they outnumbered men in tertiary education (The Economist, 2025). This raises an important question: If women are academically outperforming men, why do they continue to be underrepresented in leadership positions? The persistent gender leadership gap is well documented (Bertrand et al., 2010; Goldin, 2021; McKinsey & Company, 2024), and a growing body of research focuses on differences in preferences and biased evaluations as contributing elements (see Eckel et al. (2021) for a review). In this paper, we examine an underexplored factor: The institutional environment itself may systematically shape leadership outcomes. Through a three-pronged approach, we investigate how institutional settings influence leader effectiveness, leader evaluation, and self-selection into leadership roles, revealing that the gender leadership gap may be as much about organizational design as it is about individual differences.

The institutional environment refers to the rules, norms, beliefs, and taken-for-granted assumptions in an organization, and is increasingly recognized for its role in gender disparities in organizations (Bohnet, 2016; Goldin and Katz, 2016). An important element of institutions, and the focus of this paper, is the incentive structure for workers. Competitive incentives, such as a large individual bonus for the top performer in a team, can motivate workers, but one employee's win necessarily means that others lose out.[1] 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.[2] Women have been shown to prefer working in cooperative, as opposed to competitive environments (Kuhn and Villeval, 2015; Wozniak, 2016), however, little is known about how these environments affect leadership.

In competitive incentive environments, effective leadership can exacerbate earnings inequalities among workers. Because this unequal outcome conflicts with gender stereotypes that portray women as more prosocial and other-regarding, successful female leaders may

[1] 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 potentially backfire and lead to sabotage on the part of workers who are less likely to win (Carpenter et al., 2010).

[2] Kuhn (2009) shows that compensation based on individual performance leads to perceptions of a more individualist culture, which is linked to more competitive behavior (Leibbrandt et al., 2013), while an organization with team-based incentives is more likely to be perceived as having a collectivist culture (Cox et al., 1991). Charness and Grieco (2023) show that cooperative incentives are more effective than competitive incentives at fostering creativity within a group.

be perceived as less effective than their male counterparts. Anticipating such biased evaluations, women may find leadership roles in competitive settings less attractive. In contrast, cooperative and egalitarian incentive structures distribute the rewards of leadership more equally, aligning more closely with stereotypical expectations of female behavior. In such environments, women's leadership styles may be evaluated more favorably, and women may be more willing to pursue leadership roles.

To test this conjecture, we introduce an experimental paradigm using variations of the centipede game, first introduced by Rosenthal (1981). The game captures key elements of an organization in which total productivity is maximized by repeated gains to cooperation within a team, but in the original version of the game, individual workers are tempted to pursue their self-interest at the team's expense; that is, final payoffs are competitive. We modify the game, altering the payoffs to the workers in order to create a cooperative variation. The leader, whose gender is revealed, is tasked with sending a free-form, nonbinding message to persuade followers to voluntarily coordinate and choose actions that are best for the collective, one of the key functions of leadership (Kruse, 2013).

A clear advantage of the centipede game is that it allows subtle changes to the payoff structure to shift the environment between more competitive and more cooperative, while keeping team payoffs constant. This makes it well-suited to modeling real organizations, which often involve both cooperation and competition. As we describe in Section 2, alternative paradigms were considered but were less suitable. For instance, many cooperative games (e.g., public goods, coordination) are difficult to make competitive, while others (e.g., trust games) lack role symmetry. Competitive games, on the other hand, such as the tournament used by Niederle and Vesterlund (2007) and adapted by Cassar and Rigdon (2021a,b), do not easily generate real gains from cooperation. The centipede game, in contrast, is inherently competitive, with gains to cooperation and scope for a leader to improve group payoffs, but, with only a subtle change to payoffs, could be made more cooperative.

We also considered using observational data or field settings. But disentangling causal relationships between gender and leadership outcomes was impossible using observational data alone because of the complex interactions that underlie a leader's accomplishments and assessments. Similarly, field experiments do not allow the subtle variation in institutions that our research question requires. In contrast, laboratory experiments allow us to isolate the effect of a leader's gender on followers' decisions and evaluations without introducing confounding factors.

In our laboratory experiment, we exogenously vary both the gender of the leader and the institutional environment—via the incentive structure—to be relatively more competitive or cooperative. Consider, for example, project funding that is uncertain and susceptible to

external shocks. To model such a setting, we implement a possible early ending to the game, randomly determined by the computer (Krockow et al., 2018). These possible ending nodes provide us with a place where the payoffs can be exogenously manipulated. The competitive environment *(Comp)* is competitive in payoffs, echoing early literature on contests (Weigelt et al., 1989), as the first player to end the game receives a substantially larger share of profits. In the cooperative environment *(Coop)*, either player can still choose to end the game, with the same payoff consequences. However, if both players cooperate until the final node, or the end node that is determined by the computer, then total payoffs are shared equally.[3]

Our results demonstrate the power of the institutional environment to impact the gender leadership gap. Leadership is successful in both settings, and with both male and female leaders: Followers' exit choices occur later on average, increasing total team payoffs compared to the no-leadership case. Despite equal effectiveness, female leaders receive more negative evaluations than male leaders, but only in the competitive institution; evaluations do not differ between male and female leaders in the cooperative institution. Female leaders in the competitive environment are evaluated 50% lower than their counterparts in the cooperative setting; evaluations of male leaders across the two environments are similar, but women are evaluated higher in the cooperative than competitive setting. 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. By making the institution more cooperative in worker payoffs, we alter the perception of leader effectiveness; women are no longer penalized in evaluations or in monetary rewards.

We examine several potential mechanisms underlying the results. First, we find that the gender gap in evaluations in the competitive institution cannot be explained by (incentivized) beliefs about leader effectiveness (i.e., expectations about followers' exit choices in the game), as beliefs are accurate in predicting both the positive effect of a leader and the lack of a gender difference.[4] Since both evaluations and beliefs offer measures of leader effectiveness, this discrepancy in which evaluations (that are more subjective in nature) show a gender gap but beliefs (about objective performance) do not, highlights the importance of *how* leaders are assessed. Second, based on sentiment analysis and analysis of the content of leaders' messages, we find that male and female leaders do not differ in the advice they give, but that even after controlling for the content of leaders' messages, the bias against women persists. Our third mechanism explores the perceptions of followers, which explains the lower evaluations of women relative to men in competitive settings, an effect not observed in coop-

[3]Indeed, based on the index of competitiveness and cooperativeness developed by Demuynck et al. (2022), our *Comp* environment is more competitive than our *Coop* environment.

[4]Thus, in our context, we find no evidence of inaccurate statistical discrimination, as documented in Bohren et al. (2023).

erative environments. Specifically, female leaders receive an *additional* gender-based penalty in evaluations when their advice is perceived to be "bad" or "selfish;" that is, when gender stereotypes (i.e., women being 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, as compared to women, in both environments, with evidence suggesting that gender stereotypes also play an important role in *self*-perceptions about leadership ability.[5]

Our research makes five key contributions. First, we contribute to the growing literature showing no gender differences in leadership performance (e.g., Reuben and Timko, 2018; Grossman et al., 2019; Heursen et al., 2023). We demonstrate that this result holds across institutional environments that differ only in their payoff structure—competitive versus cooperative— while holding all other factors constant. This suggests that the pronounced gender gap in leadership observed in the private sector, compared to the more modest gaps in the not-for-profit sector, cannot be attributed to differences in leaders' performance.[6]

Second, we build on evidence that gender gaps in leadership reflect differences in preferences for job attributes (Konrad et al., 2000; Blau and Kahn, 2017; Mas and Pallais, 2017) and for cooperation over competition (Gneezy et al., 2003; Kuhn, 2009; Healy and Pate, 2011; Dargnies, 2012; Cassar and Rigdon, 2021a). We show that women are less willing to lead regardless of the environment, suggesting that self-perceptions, rather than preferences alone, may limit leadership entry. Despite equal performance, women perceive themselves as less capable than men.

Third, we contribute to research on biased evaluations of women in leadership (Eagly et al., 1995; Mengel et al., 2019; Sarsons et al., 2021; Erkal et al., 2023). We show that the institutional context matters: Women are evaluated more negatively than men only in competitive environments, where their leadership success clashes with prevailing gender stereotypes. This could also underlie women's preference to work in more cooperative environments.

Fourth, we introduce a novel application of the centipede game to study leadership. We show that leadership in this setting can improve social efficiency by coordinating cooperation, and we offer a new experimental design that manipulates the environment (competitive vs. cooperative) without altering total payoffs or equilibrium strategies. This design can be used

[5]Born et al. (2022) consider how team gender composition affects willingness to lead and find that women are more willing to lead in female-majority teams than male-majority teams.

[6]The gender leadership gap 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). This chasm is less pronounced in not-for-profit industries, e.g., women hold 43% of CEO positions in the nonprofit sector (AAUW, 2022).

to study leadership dynamics more broadly.

Finally, we provide an integrated analysis of both the supply and demand sides of the gender leadership gap. Women are less willing to lead (supply), and are penalized more harshly by followers (demand), but only in competitive settings. These effects appear to stem from both external stereotypes and women's internalized beliefs. In cooperative environments, demand-side biases disappear, suggesting that institutional design could help to reduce gender disparities in leadership.

Our findings carry important policy implications. Interventions to increase women's leadership participation must be tailored to the institutional context. Supply-side interventions may be effective in egalitarian settings, but in competitive environments, such strategies may provoke backlash unless demand-side biases are addressed. Organizational reforms that reduce evaluation bias or shift toward more cooperative environments may help level the playing field for female leaders.

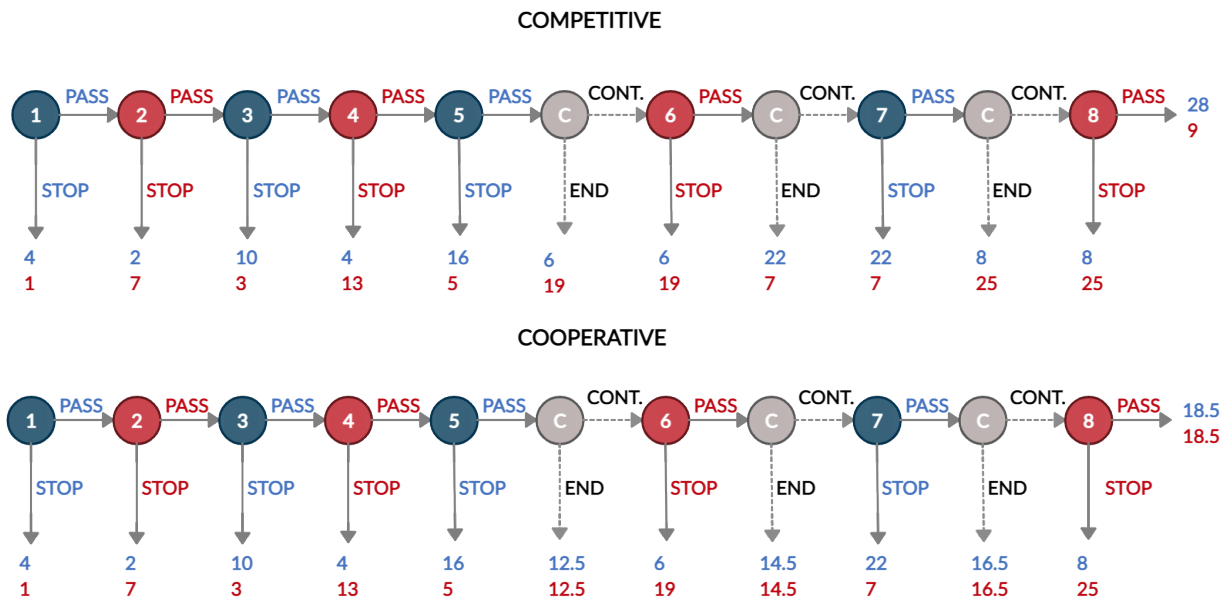
2 Experimental Design

We introduce an experimental paradigm for studying leadership, in which we exogenously vary both the gender of the leader and the institutional environment, within the context of team cooperation. The centipede game, introduced by 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, thereby increasing the pot further. Each player faces a dilemma: Taking the pot earlier yields immediate gains, but passing can potentially lead to higher rewards. Figure 1 illustrates the game as implemented in our experiment. The game is played in randomly assigned 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 decision to pass and are maximized if the game reaches the terminal node. In practice, we implement the game using a strategy method similar to García-Pola et al. (2020a) and explained in further detail below.[7] The game's incentive structure is inherently competitive, in that the payoffs

[7]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 means 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

are unequal at each node and the decision maker has an incentive to exit first, beating out their opponent.[8]

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 terminal node, determined by the computer.

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). One major advantage of using the centipede game is that it allows us to study the role of the leader in shifting the beliefs (and subsequent behavior) of followers off the equilibrium path toward cooperation, an important function of leadership (Gächter and Renner, 2018), and indeed we report evidence of this in Section 3.1.[9]

"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.

[8]See e.g., Weigelt et al. (1989) for early contest literature in which players compete to earn a larger payoff in a tournament.

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

A second benefit of using the centipede game is the gains to cooperation, which mimic successful cooperation in the field. 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, there is potential scope for the leader to improve productivity by encouraging players to pass.

Another important advantage is that the game can be easily adapted to be relatively more competitive or cooperative using only very subtle changes to the payoff structure. Such an environment that is not at the extremes (i.e., purely competitive or purely cooperative) is more representative of real organizations that often contain elements of both competition and cooperation. We considered many alternative games in the experimental repertoire but found that most games with gains to cooperation are difficult to make competitive (e.g., the public goods game or coordination game), or are not symmetric in players' roles (e.g., the trust game). We also considered starting with a competitive game. For instance, Cassar and Rigdon (2021a,b) take the widely used tournament of Niederle and Vesterlund (2007) and allow subjects to split the prize.[10] However, it is not straightforward to alter the game such that there are gains to cooperation. We therefore found the centipede game to be ideal as the game is inherently competitive, with gains to cooperation and scope for a leader to improve group payoffs, but can be made more cooperative with only a subtle change to payoffs. Despite this adjustment in payoffs being small, the payoff structure is more competitive in the competitive version than in the cooperative version of the game based on the index of competitiveness and cooperativeness in normal-form games by Demuynck et al. (2022). The index is based on the decomposition of a game into a competitive benchmark and a cooperative benchmark.[11]

A key innovation in our modification of the game 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: There is an equal probability that the game will end after node 5, 6, 7, or 8, a technique frequently used in experimental economics to prevent endgame effects.[12] While removing endgame effects is also useful for us, our reason

[10]While Niederle and Vesterlund (2007) examine the decision to enter into competition, we compare behavior in two environments, one of which is more competitive than the other in payoffs.

[11]More details are available upon request.

[12]Random endings are used in experiments 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. Krockow et al. (2018) are the first to test endgame effects in repeated games with alternating play, like the centipede game. They design their experiment explicitly to test the effect of different properties of random ending times on the length of game play.

for incorporating the random endgame is that it gives us a way to manipulate the incentive structure. The random ending ensures that both players have an equal chance of ending up with a larger share at the computer-determined terminal node, even in the competitive environment. But more importantly, this random ending also gives us an opportunity to alter the game to make the incentive structure more cooperative. In our novel cooperative environment, total payoffs are the same as in the competitive environment, 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 detail below.

In an initial survey, participants select a gender-specific alias, which is later used to reveal the leader's gender to their group without drawing undue attention to gender.[13] Gender (including the gender of followers) is otherwise not revealed or discussed during the experiment.[14] We also asked for participants' age, field, and year of study to reduce the salience of gender in the survey. In Part 1, all participants are randomized into either the *Comp* or *Coop* treatment and play the centipede game in anonymous pairs. In Part 2, participants are asked their willingness to be a leader in the game, knowing that their answers will not determine who is selected to be leaders. Participants are rematched into new groups of three, consisting of one leader and two anonymous followers and then play the same treatment of the game again. 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 (incentivized) beliefs about when the game would end in Parts 1 and 2. Followers also provide subjective evaluations of their leader and can choose to adjust their leader's earnings (all participants are aware of this possibility before making any decisions in Part 2). 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.

2.1 Choices without a Leader (Part 1)

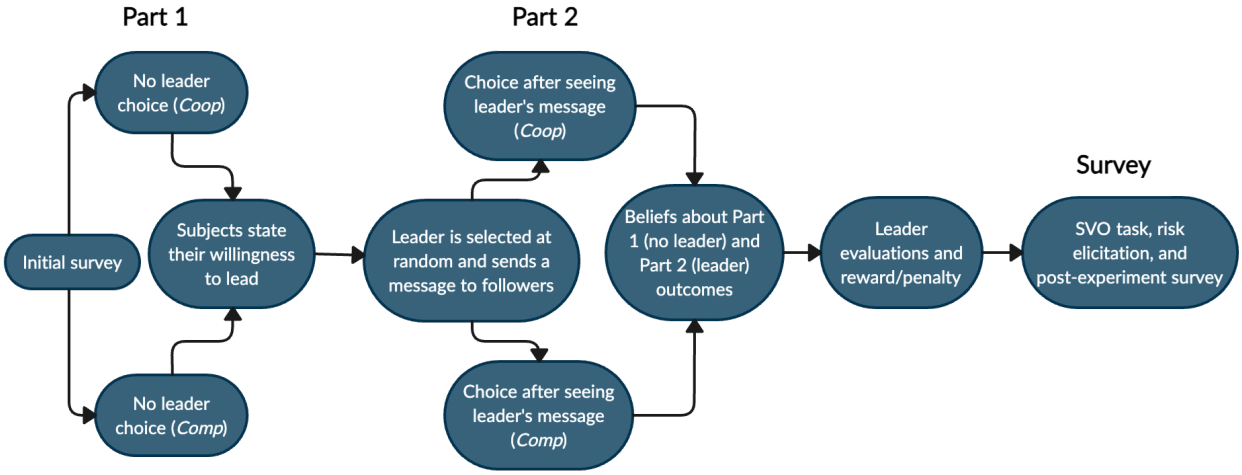
2.1.1 Competitive Environment

In Part 1, participants are randomly matched in pairs and play a centipede game using the strategy method. Each player makes an exit choice regarding when to "Stop" (at their

[13]See the full list of names in Appendix Table B.1. This name-based approach follows Chakraborty and Serra (2024) to avoid experimenter demand bias. If subjects select "non-binary," "gender-diverse," or "my gender is not listed," they can choose between two gender-neutral names.

[14]When asked what the situation in the experiment reminded them of, only four out of three hundred participants mentioned anything involving gender. Two participants commented on or suggested not using gendered names. Two participants said the leader's gender is not relevant for them or could not be assessed.

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



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 (AUD), 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 are in competition 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.

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 roles in Part 2, we measure willingness to lead (WTL) by asking each participant to indicate (on a scale from 1 to 10) how much they want to be the leader of their group. Note that we elicit WTL before participants make any decisions in Part 2, which precludes the possibility that leadership ambitions are affected by outcomes in the game. We randomly assign participants to leadership positions (and participants are aware of this) to ensure that we have an equal sample of male and female leaders and to reduce selection bias.[15,16]

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 environment as in Part 1 (either *Comp* or *Coop*).[17] The leader has no decision to make in the game itself and only has a minimal advisory role within their group. Leaders are instructed to send a message to their followers to explain potential strategies and were only given two restrictions: The leader could not give away identifiable information or use offensive language. We intentionally chose a free response format as it holds more external validity and is less likely to suffer from experimenter demand effects. Through this, we explore the role of leaders in managing followers' beliefs, which has been shown to improve social efficiency in games that require coordination (Sahin et al., 2015; Gächter and Renner, 2018). The interests of the leader and the group as a whole

[15]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.

[16]Based on their responses in the post-experiment survey, we find no evidence that participants failed to take the question seriously or responded untruthfully. When asked to explain their reported willingness to lead, participants gave thoughtful answers about their reasoning, e.g., citing confidence in their leadership abilities, analytical skills, etc. This is consistent with recent research showing the value of subjective beliefs Exley and Kessler (2022) and the comparability of incentivized and nonincentivized measures of beliefs Gangadharan et al. (2024).

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

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* for comparability. This allows us to isolate the impact of the incentive structure for workers on leadership outcomes without having to disentangle differences in leader incentives. Given that leaders are usually assessed based on different metrics, it is not uncommon for leaders to face a different set of incentives than workers in real-world organizations.

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 managers, peers, 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 $\pm$\$1 change in the leader's payoff (up to $\pm$\$3). One follower in each group is chosen at random to have their choice implemented for the leader. We also elicit incentivized beliefs about when the game will end in both Part 1 and Part 2.

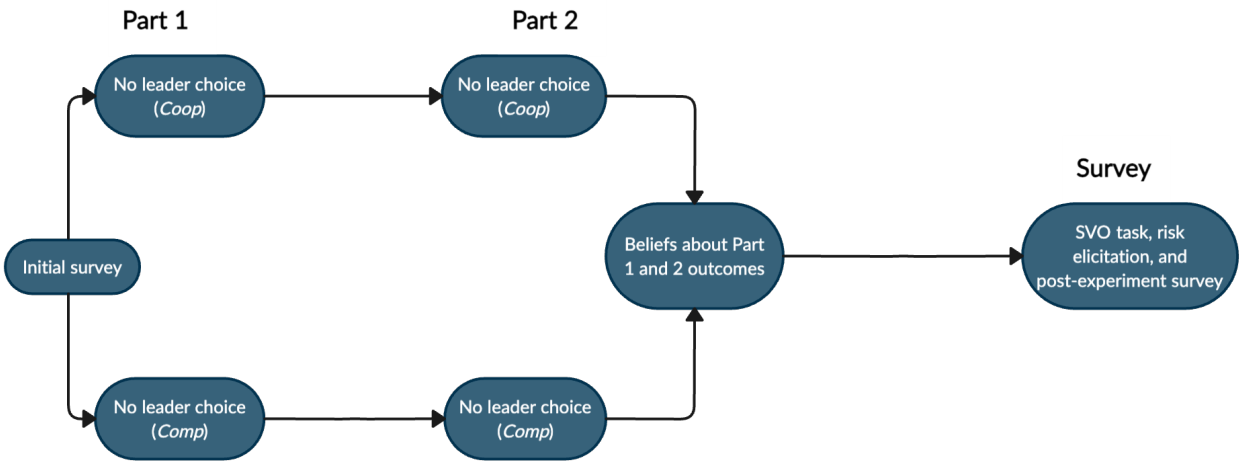
To account for potential learning or experience effects, we conduct two control treatments (*Comp-Control* and *Coop-Control*), in which Part 1 is simply repeated and participants make a second decision in Part 2 without a leader. See Figure 3 for a summary of the experimental timeline for the control treatments. This allows us to isolate the effect of a leader on followers' decisions and rule out a difference in choices due to experience alone.[18]

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 participants 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

[18]Our aim with the Control treatment was to test the effect of having a leader on follower decisions in different environments, rather than to isolate specific mechanisms explaining why leaders have an effect, which has been studied by others (see e.g., Van Knippenberg and Hogg (2003)).

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



father's education. In open-ended questions, we also ask participants to explain their: 1) decisions in the game; 2) reasons for their 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 and 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 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 (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 toward female than male leaders (Ridgeway, 2001; Eagly and Karau, 2002). We expect *Coop* to create a more favorable environment for female leaders. This 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 in 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 participants, and the Experimental Economics Laboratory at the University of Melbourne, using ORSEE (Greiner, 2015) for recruitment, between July - November 2021.[19] Due to stay-at-home orders in Melbourne, online sessions of approximately 18 participants per session were conducted via Zoom with similar conditions to a laboratory environment.[20] We report results from a total of N=400 participants (N=300 in the main treatments and N=100 in the control treatments).[21] The sample size was determined based on a power analysis conducted after we ran a small pilot study with 24 participants per treatment. We chose a sample size of 150 participants per treatment to be able to detect an effect size that is as small as 75% (i.e., Cohen's d = 0.37) of that observed in the pilot, with 90% power and $\alpha = 0.05$. Participants 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 #74160).

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

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

[21] We exclude from our analysis five participants who reported their genders as "non-binary" or "gender diverse" in *Comp* and two participants who reported their genders as "non-binary" in *Coop*.

3 Results

3.1 Effect of Leaders on Exit Choices

In an initial 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 (i.e., at every other node). For example, an exit choice of 3 for Blue (Red) is equivalent to the game ending at node 5 (6). Alternatively, they may choose to Always Pass (see Figure 1). 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: 2.95 vs. leader: 3.67, $p < 0.01$) and *Coop* (no leader: 3.81 vs. leader: 4.52, $p < 0.01$).[22,23] On average, the presence of a leader encourages participants 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).[24] 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* (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 (Gächter and Renner, 2018). The difference in actual exit choices parallels beliefs about exit nodes (between 1 and 9). 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 E.1 and Table E.1).[25] Overall, choices in the game are consistent with participants' beliefs.

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)

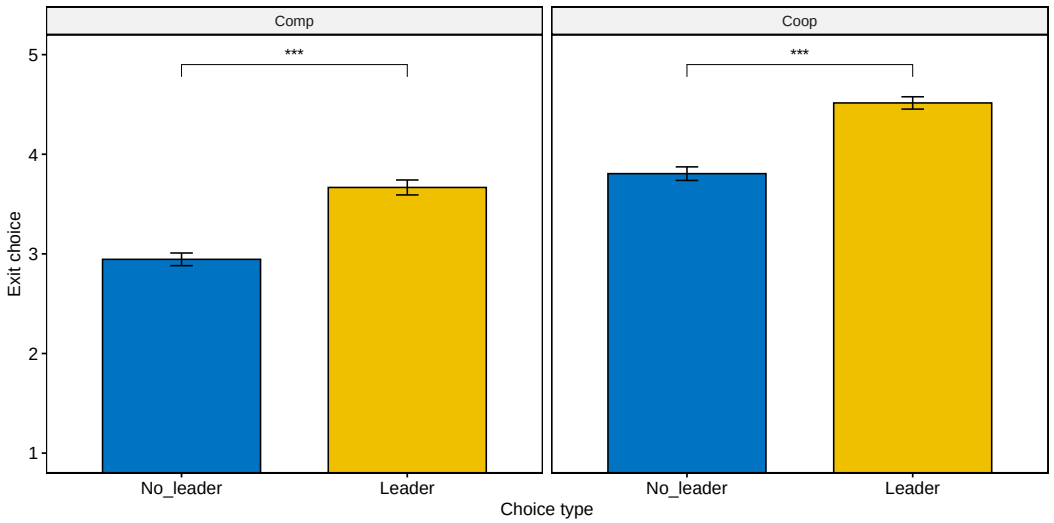
[22] See Appendix Figure C.1 for distribution of exit choices by round and Figure **??** for a comparison of mean exit choices with and without a leader.

[23] 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).

[24] All results from the OLS regression analyses hold when using an ordered probit regression. See Appendix C-E. See Appendix Table C.1 for the combined impact of having a leader and their gender.

[25] Appendix Table E.2 shows that differences in beliefs in rounds with and without a leader do not vary by leader or follower gender.

Figure 4: Exit choices with and without a leader



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

for both the competitive ($p = 0.24$) and cooperative environments ($p = 0.18$). We therefore conclude that it is the leader, rather than learning or experience, that drives later exit choices. Comparing across environments, followers tend to exit later and earn more in the cooperative setting than in the competitive one, both with and without a leader (all $p < 0.01$), suggesting that cooperative incentives can be more effective at increasing efficiency when there are gains to cooperation.

Next, we compare exit choices under male and female leaders. 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 5.[26] These results are supported by the regression analysis in Table 2.[27] This leads us to our first result:

Result 1: Male and female leaders are equally effective in both the competitive and cooperative environments.

3.2 Leader Evaluations and Discretionary Payments

3.2.1 Leader Evaluations

Figure 6 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

[26]See Figure C.3 for the distribution of exit choices by leader gender.

[27]We find no evidence to suggest a significant interaction between leader and follower gender (see Table C.2). This result is robust to an alternative measure of leader effectiveness, using the difference in exit choices (see Table C.4).

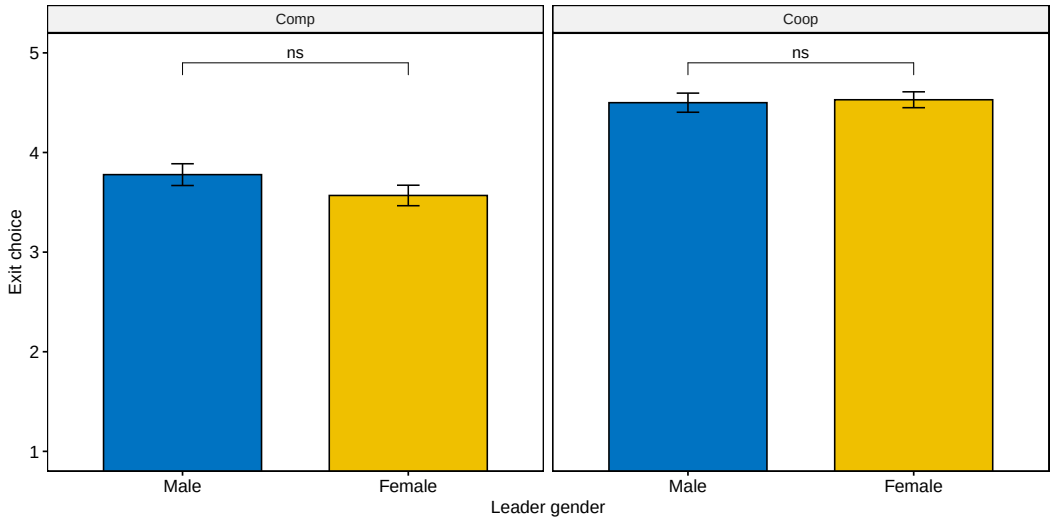
Table 1: Determinants of exit choices

	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.

Figure 5: 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.

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, $p < 0.01$). This gender gap disappears in the more cooperative environment (M: 5.00 vs. F: 5.12, $p = 0.51$). Strikingly, female leaders are evaluated 50% higher in the cooperative setting ($p < 0.01$), whereas men are only evaluated 14% higher ($p = 0.12$), though this difference for men is not significant. Similarly, when

Table 2: Determinants of exit choices by leader gender

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.21	−0.14	−0.01	−0.03
	(0.23)	(0.26)	(0.18)	(0.16)
Female follower	−0.20	−0.22	0.51***	0.41***
	(0.23)	(0.22)	(0.15)	(0.15)
Constant	3.89***	3.76***	4.23***	3.43***
	(0.20)	(1.31)	(0.20)	(1.01)
Controls	No	Yes	No	Yes
R^2	0.02	0.22	0.08	0.33
Adj. R^2	0.01	0.12	0.07	0.24
Num. obs.	192	192	198	198

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

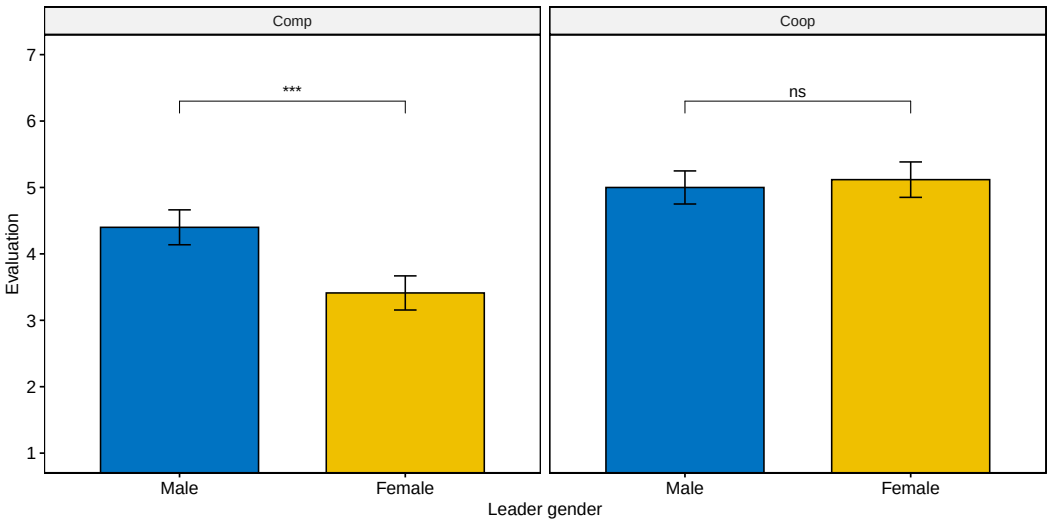
Notes: Ordinary least squares regression with standard errors clustered at the group level in parentheses. The dependent variable is the exit choice (between 1 and 5). The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

we examine the distribution of evaluations (Figure D.1), 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). In *Coop*, we find no significant difference in the distribution of evaluations ($p = 0.78$, Kolmogorov-Smirnov test).

Figure 7 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.[28] The regression analysis in Table 3 confirms this result and these results hold when we control for followers' own choices and beliefs (see Tables D.1 and D.2). Consistent with a general preference for working in cooperative environments, evaluations by female followers are significantly higher than evaluations by male followers in *Coop* ($p < 0.01$, column 4) and we do not observe a significant interaction between the gender of leaders and followers (see Table D.3).

[28]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.

Figure 6: Evaluations by leader gender



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

3.2.2 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.[29] On average in *Comp*, male leaders receive a positive adjustment of \$0.19 while female leaders receive a negative adjustment of \$0.09 ($p = 0.19$). 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 by \$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 but not significantly so (M: \$0.00 vs. F: \$0.19, $p = 0.31$).[30]

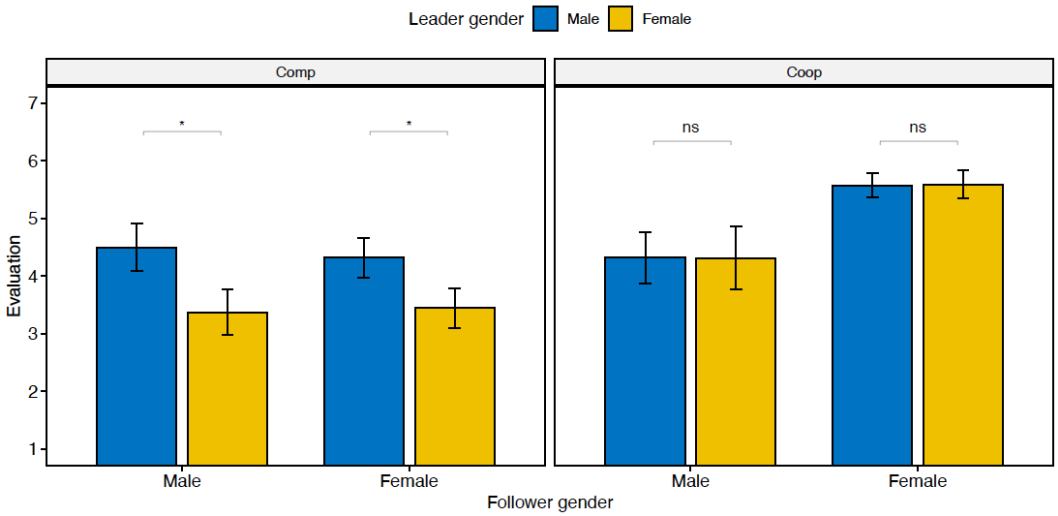
In Table 4, we pool *Comp* and *Coop* and examine whether followers make a positive or negative payment adjustment (columns 1-3) and the average payment adjustment amount (columns 4-6).[31] The negative interaction term shows that male followers are more likely

[29]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 between environments is not significant ($p = 0.42$).

[30]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 ($p = 0.17$).

[31]We show this separately for *Comp* and *Coop* in Figure D.2 as well as Table D.5.

Figure 7: Evaluations by leader and follower gender



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

to make a negative payment adjustment for female leaders ($p = 0.03$, column 3). Similarly, male followers give female leaders significantly more negative adjustments as compared to male leaders ($p = 0.05$, column 6).

Additionally, leader evaluations positively predict both payment adjustments ($p < 0.01$, column 2) and the adjustment amount ($p = 0.01$, column 5). 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 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:

Result 2: Female leaders are evaluated less favorably than male leaders in *Comp*, but not in *Coop*. Evaluations of female leaders are 50% higher in the cooperative setting while the 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*.

3.3 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 actual effectiveness. We first investigate

Table 3: Determinants of leader evaluations

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	-0.99^{***}	-1.16^{***}	0.01	-0.29
	(0.38)	(0.42)	(0.36)	(0.39)
Female follower	-0.05	-0.11	1.27^{***}	1.03^{***}
	(0.38)	(0.39)	(0.36)	(0.36)
Constant	4.43^{***}	7.96^{***}	4.31^{***}	2.50
	(0.37)	(2.27)	(0.42)	(1.78)
Controls	No	Yes	No	Yes
R^2	0.07	0.35	0.12	0.32
Adj. R^2	0.05	0.15	0.10	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.

whether beliefs about the effectiveness of male and female leaders can explain this gap (Section 3.3.1). Second, we examine leaders' messages to see if gender differences exist in the message length and whether this has an impact on evaluations (Section 3.3.2). Third, we explore the content of leaders' messages (Section 3.3.3) 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 3.3.4). 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.

3.3.1 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 incentivize participants to report when they expect 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, $p = 0.55$).[32] Participants also correctly predict later exits in the

[32]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 4: Determinants of payment adjustments

	Payment adjustment			Adjustment amount		
	(1)	(2)	(3)	(4)	(5)	(6)
Leader evaluation	0.08***	0.06***	0.06***	0.13***	0.10***	0.10**
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)
Coop	−0.08	−0.06	−0.07	−0.18	−0.17	−0.18
	(0.07)	(0.07)	(0.07)	(0.14)	(0.15)	(0.14)
Female leader	0.01	0.01	0.14*	0.01	−0.02	0.22
	(0.07)	(0.07)	(0.08)	(0.13)	(0.14)	(0.15)
Male follower	0.08	0.07	0.23**	0.17	0.18	0.47**
	(0.07)	(0.07)	(0.10)	(0.14)	(0.13)	(0.19)
Belief	0.02	0.04*	0.04*	0.06	0.09**	0.09**
	(0.02)	(0.02)	(0.02)	(0.04)	(0.05)	(0.05)
Female leader x Male follower			−0.30**			−0.55*
			(0.14)			(0.28)
Constant	−0.39***	−0.21	−0.38	−0.89***	−0.48	−0.80
	(0.15)	(0.42)	(0.42)	(0.31)	(0.84)	(0.82)
Controls	No	Yes	Yes	No	Yes	Yes
R^2	0.09	0.29	0.30	0.08	0.22	0.24
Adj. R^2	0.07	0.18	0.19	0.06	0.10	0.11
Num. obs.	195	195	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.

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* ($p = 0.75$, column 2) or in *Coop* ($p = 0.45$, column 4). These results are consistent under an alternative measure of beliefs about leader effectiveness, which is measured as the difference between beliefs about when the game would end with and without a leader (see Table E.2 in Appendix E), with no evidence of an interaction between leader and follower gender.

Note that by asking participants when the game will end, we are indirectly asking for an assessment of the effectiveness of the leader in encouraging a later exit choice. While the discrepancy between evaluations and beliefs is puzzling at first glance, these two measures differ in two main ways. First, in contrast to the evaluation question, the leader's alias

(and hence the leader's gender) is not explicitly mentioned when we elicit beliefs. Another key difference is that the belief question is tied to an objective measure of outcomes and is incentive-compatible, while the evaluation of the leader, though commonly used, is based on more subjective criteria in the measure of leadership quality. The fact that we observe a marked penalty for women in evaluations but not in beliefs suggests that *how* leaders are evaluated is important and gender differences may be especially pronounced when using more subjective measures. The observed gender bias is consistent with theories from psychology that greater ambiguity in performance criteria may result in more biases against women (Heilman and Haynes, 2008; Maas and Torres-Gonzalez, 2011; Heilman et al., 2024). Our findings also help to shed light on why women tend to be underrepresented in jobs that rely on more subjective measures of performance (Jirjahn and Stephan, 2004), why more structured evaluation templates could reduce the gender gap in start-up funding (Miller et al., 2023) and why women might receive lower ratings of "potential" despite higher performance (Benson et al., 2024). We further contribute to this literature by showing that the environment plays a crucial role in the evaluation process.

Table 5: Determinants of followers' beliefs about exit nodes

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.32	−0.15	−0.06	0.35
	(0.42)	(0.46)	(0.38)	(0.46)
Female follower	0.58	0.55	0.47	0.38
	(0.37)	(0.37)	(0.39)	(0.46)
Constant	5.83***	8.25***	7.08***	9.05***
	(0.36)	(2.51)	(0.42)	(2.26)
Controls	No	Yes	No	Yes
R^2	0.03	0.19	0.01	0.31
Adj. R^2	0.01	−0.05	−0.01	0.10
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.

3.3.2 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 both *Comp* ($p < 0.01$, column 2, Table 6) and *Coop* ($p = 0.04$, 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, $p = 0.96$). Notably, even after controlling for the length of the message, we still observe a penalty in evaluations for women in *Comp* ($p < 0.01$, column 2). We therefore find no evidence that the difference we observe in leader evaluations for men and women in *Comp* is driven by the length of leaders' messages.

Table 6: Effect of message length on leader evaluations

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
# Words	0.03^{***}	0.03^{***}	0.01	0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)
Female leader	-0.67^{*}	-0.95^{***}	0.09	-0.30
	(0.35)	(0.36)	(0.41)	(0.40)
Constant	3.68^{***}	6.65^{***}	4.72^{***}	1.35
	(0.29)	(2.17)	(0.42)	(1.76)
Controls	No	Yes	No	Yes
R^2	0.16	0.43	0.02	0.35
Adj. R^2	0.14	0.25	-0.00	0.15
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.

3.3.3 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.[33] 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

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

messages in *Comp* and *Coop* separately.[34]

Second, we perform a more detailed analysis of the message content by employing 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 without knowing the research questions or participants' 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).[35]

Leaders' messages were classified into four categories: "Profit maximizing" (74%), "Stop early" (18%), "Tailored" (4%), and "No advice" (4%).[36] 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.[37] Messages that fail to give any meaningful advice are classified as No advice.[38] 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 Table G.1.

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

[34]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 ($p = 0.72$). Messages from male leaders do not differ in sentiment across environments ($p = 0.21$), though they do for female leaders ($p < 0.01$).

[35]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).

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

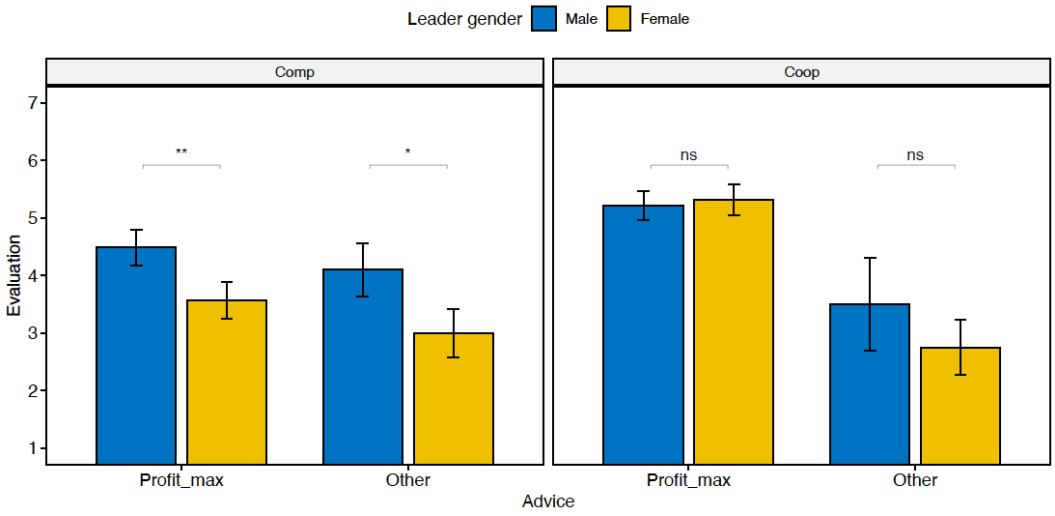
[37]An example of tailored advice is: "Blue: stop at round 5. Red always pass."

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

significant.

Examining leader evaluations conditional on the type of advice given (Figure 8), 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 ($p < 0.01$, column 2). In *Coop*, leaders who give Profit maximizing advice tend to receive higher evaluations ($p < 0.01$, column 4), while female followers tend to give higher ratings than male followers ($p = 0.01$).

Figure 8: 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.

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.

Table 7: Determinants of leader evaluations by message type

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	-0.96^{**}	-1.16^{***}	-0.07	-0.39
	(0.38)	(0.42)	(0.32)	(0.34)
Profit max advice	0.49	0.58	1.84^{***}	1.84^{***}
	(0.36)	(0.38)	(0.56)	(0.54)
Female follower	-0.07	-0.13	1.13^{***}	0.87^{**}
	(0.38)	(0.39)	(0.33)	(0.34)
Constant	4.05^{***}	7.68^{***}	2.77^{***}	0.32
	(0.45)	(2.23)	(0.58)	(1.94)
Controls	No	Yes	No	Yes
R^2	0.08	0.36	0.21	0.40
Adj. R^2	0.05	0.16	0.19	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 baseline advice is Other advice. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

3.3.4 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," "Selfish," or "Other."[39] 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 , Selfish or Other. 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.[40] In particular, in the competitive setting, female leaders receive lower evaluations than male leaders for giving Bad (F: 2.42

[39]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.

[40]In Table H.1 we confirm that this result is driven by the competitive environment, after controlling for interactions between leader and follower gender.

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, $p = 0.76$).

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.

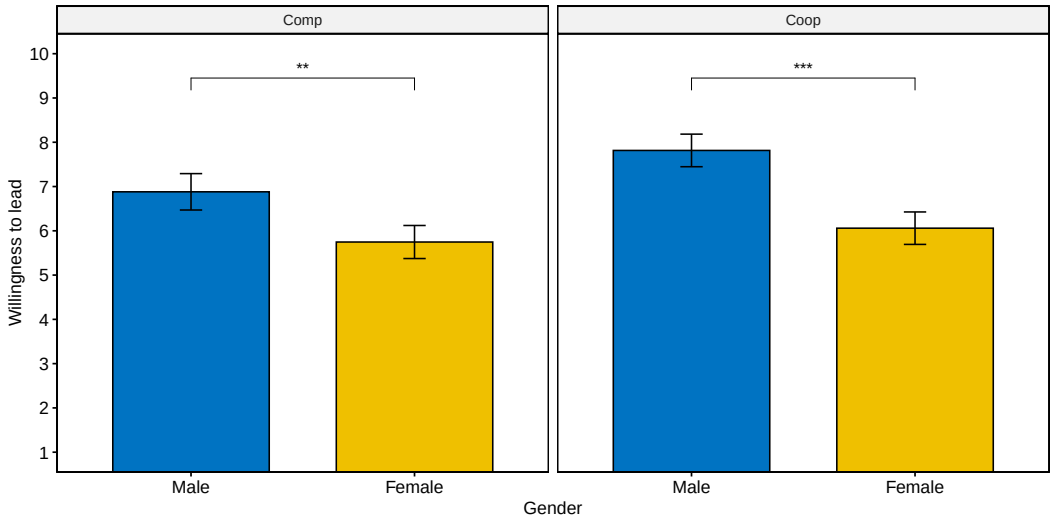
Taken together, the results show that the gender gap in evaluations cannot be explained by differences in actual leader effectiveness, expectations about effectiveness, differences in the length of leaders' messages, differences in the tone of the message, nor by differences in the type of advice given 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. Because women are stereotypically more egalitarian, their inability to avoid the unequal payoffs inherent in the competitive version of the game may also lead to harsher evaluations overall in that setting, especially when their advice is also selfish. This 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.

3.4 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 9: *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, $p = 0.57$).

Figure 9: Willingness to Lead (WTL) by gender



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

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 in the regression analysis, 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:

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.

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.

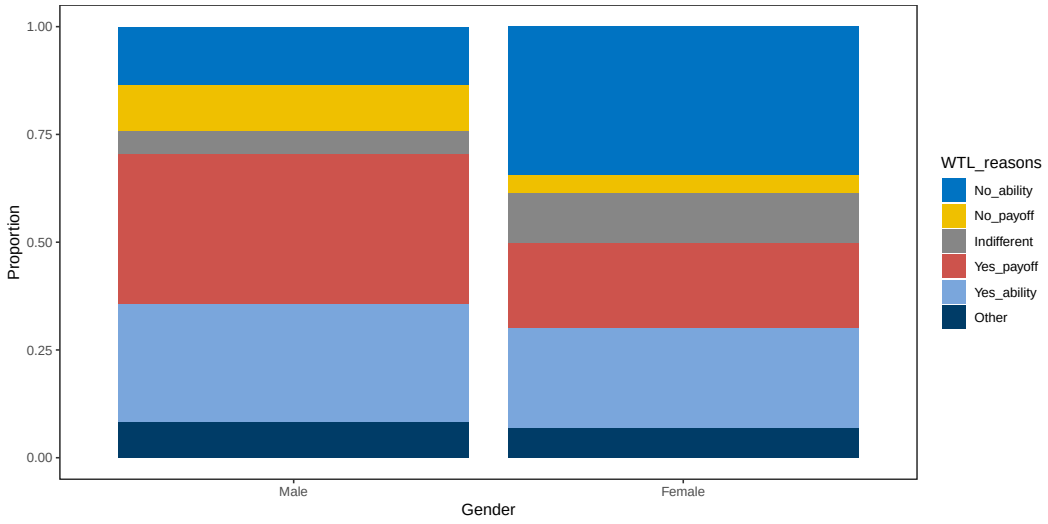
3.4.1 Why Does Reported Willingness to Lead Differ by Gender?

To further investigate the gender gap in WTL, we investigate participants' survey responses explaining their reported willingness to become leaders. Participants' 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").[41] Figure 10 summarizes the proportion of each category for

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

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 ($p < 0.01$).[42] 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.

Figure 10: Willingness to Lead reasons



3.4.2 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

[42]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" ($p < 0.01$).

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 ($p = 0.71$ and $p = 0.66$, columns 2 and 4).

Table 10: Effectiveness by leader's WTL by treatment

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
WTL of leader	0.01	0.01	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)
Female leader		−0.14		−0.01
		(0.27)		(0.17)
Female follower		−0.21		0.42***
		(0.23)		(0.15)
Constant	3.61***	3.72***	4.42***	3.27***
	(0.21)	(1.34)	(0.18)	(1.03)
Controls	No	Yes	No	Yes
R^2	0.00	0.23	0.00	0.33
Adj. R^2	−0.00	0.12	−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.

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

We find no significant correlation between WTL and leader evaluations for female leaders in either environment (see Table I.1). However, when we examine this correlation for male leaders only (Table I.2), 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.[43] 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 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 experimental paradigm using a modified centipede game to study leadership 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 and investigate three aspects: effectiveness, evaluation and self-selection into leadership.

We show that leaders are effective at encouraging later exit choices in 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 stark 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 to suggest that stereotypes play a key role in followers' perceptions of their leaders. Female leaders are espe-

[43] 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 ($p = 0.25$).

cially penalized when their advice is perceived to be inconsistent with traditional stereotypes about women. Our findings show that women consistently express a reluctance to lead as compared to their male counterparts, suggesting that stereotypes also matter for perceptions about their own effectiveness as leaders. We thus provide robust evidence on both the demand and supply side of leadership, whereby 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 female leaders are reinforced by the negative self-stereotypes that women hold about their own abilities, thus explaining the stark gender leadership gaps observed in 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 industries such as the not-for-profit sector.

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 subjective 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 the leadership desires of other women in the organization.

While we find in our set-up that the cooperative setting is more effective at encouraging cooperation and enhancing efficiency, 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. To mitigate these biases, highly competitive organizations might consider adopting more objective rather than subjective measures to assess leadership effectiveness. For example, such measures could include evaluating leaders based on their teams' ability to meet sales targets, complete projects with high client satisfaction, or other quantifiable key performance indicators.

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. 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.

Benson, A., Li, D., and Shue, K. (2024). Potential and the gender promotions gap. *Available at SSRN*.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bohnet, I. (2016). *What works*. Harvard University Press, Cambridge MA.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, pages 1–45.

Born, A., Ranehill, E., and Sandberg, A. (2022). Gender and willingness to lead: Does the gender composition of teams matter? *Review of Economics and Statistics*, 104(2):259–275.

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

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.

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. (2024). Gender and leadership in organisations: The threat of backlash. *The Economic Journal*, 134(660):1401–1430.

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.

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. In Capra, C. M., Croson, R. T., Rigdon, M. L., and Rosenblat, T., editors, *Handbook of Experimental Game Theory*, pages 149–183. Edward Elgar Publishing.

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.

Demuynck, T., Seel, C., and Tran, G. (2022). An index of competitiveness and cooperativeness for normal-form games. *American Economic Journal: Microeconomics*, 14(2):215–239.

Duckworth, A. (2016). *Grit: The Power of Passion and Perseverance*. Scribner, New York, NY.

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.

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.

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.

Gangadharan, L., Grossman, P. J., and Xue, N. (2024). Belief elicitation under competing motivations: Does it matter how you ask? *European Economic Review*, 169:104830.

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, Princeton, NJ.

Goldin, C. and Katz, L. F. (2016). A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, 34(3):705–746.

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.

Heilman, M. E., Caleo, S., and Manzi, F. (2024). Women at work: Pathways from gender stereotypes to gender bias and discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, 11(1):165–192.

Heilman, M. E. and Haynes, M. C. (2008). Subjectivity in the appraisal process: A facilitator of gender bias in work settings. *Beyond Common Sense: Psychological Science in the Courtroom*, pages 127–155.

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

Jirjahn, U. and Stephan, G. (2004). Gender, piece rates and wages: Evidence from matched employer–employee data. *Cambridge Journal of Economics*, 28(5):683–704.

Konrad, A. M., Ritchie Jr, J. E., Lieb, P., and Corrigall, E. (2000). Sex differences and similarities in job attribute preferences: A meta-analysis. *Psychological Bulletin*, 126(4):593.

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.

Maas, V. S. and Torres-Gonzalez, R. (2011). Subjective performance evaluation and gender discrimination. *Journal of Business Ethics*, 101:667–681.

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 (2024). 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.

Miller, A., Lall, S. A., Goldstein, M., and Montalvao, J. (2023). Asking better questions: The effect of changing investment organizations' evaluation practices on gender disparities in funding innovation. Policy Research Working Paper 10625, World Bank Group.

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.

The Economist (2025). The best places to be a working woman. https://www.economist.com/graphic-detail/2025/03/05/the-best-places-to-be-a-working-woman-in-2025.

Van Knippenberg, D. and Hogg, M. A. (2003). A social identity model of leadership effectiveness in organizations. *Research in Organizational Behavior*, 25:243–295.

Weigelt, K., Dukerich, J., and Schotter, A. (1989). Reactions to discrimination in an incentive pay compensation scheme: A game-theoretic approach. *Organizational Behavior and Human Decision Processes*, 44(1):26–44.

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

Online Appendices:

Design choices additional analysis, and instructions

A The Centipede Game

Readers sometimes ask why we adopted the centipede game instead of one 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.

Our experimental model requires a game that can be varied to produce versions that are relatively more or more cooperative, but that otherwise remain similar. Most games used to study cooperation are lacking in one or more of the characteristics required by our research question. We required 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 positively impact 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 first considered 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 also 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 participants to split the prize. Moreover, 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 more 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 impact behavior, as seen by the experimental results.

B List of aliases

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

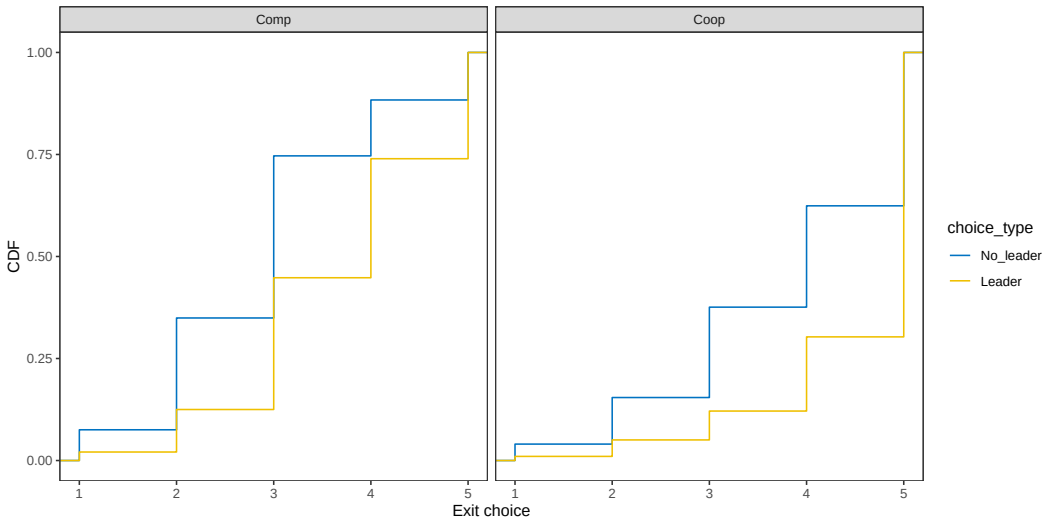
Female	Male	Non-binary	Gender-diverse
Isabella	William	Alex	Lillian
Chloe	Jack	Sam	John
Charlotte	Lachlan	Taylor	Ellie
Olivia	Joshua	Nicky	Andrew
Mia	Thomas	Robin	Stephanie
Emily	Oliver	Sasha	Sebastian
Sophie	Noah	Lee	Stacey
Ruby	Ethan	Blake	Hamish
Amelia	James	Kelly	Abby
Ella	Lucas	Andy	George
Lily	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.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 participants to select a later exit node in both *Comp* and *Coop* of 44% and 33% respectively.

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

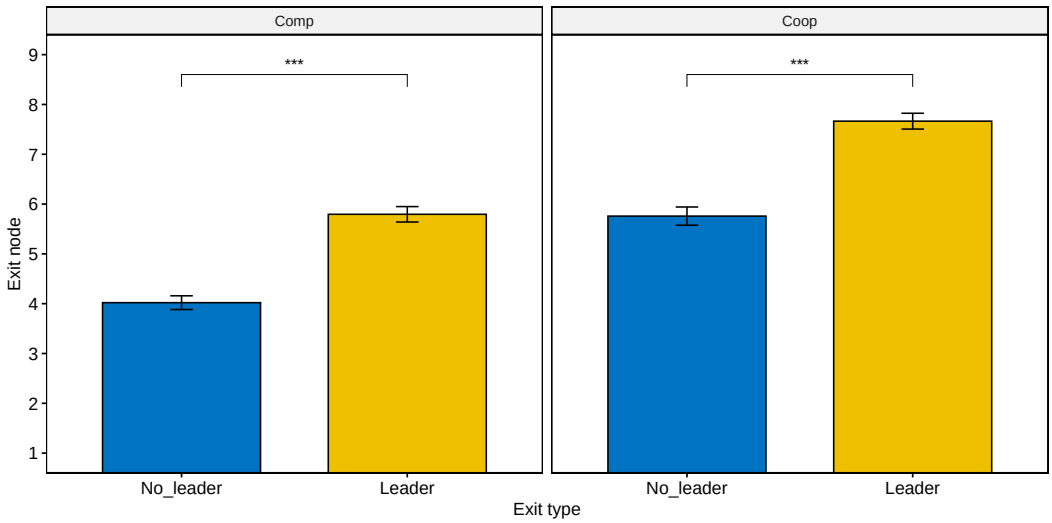


C.2 Control treatment

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 E.1) 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.4 supports Result 2 that male and female leaders do not differ in effectiveness in *Comp* ($p = 0.48$, column 2) and in *Coop* ($p = 0.91$, column 4).

Figure C.3: Distribution of exit choices by leader gender

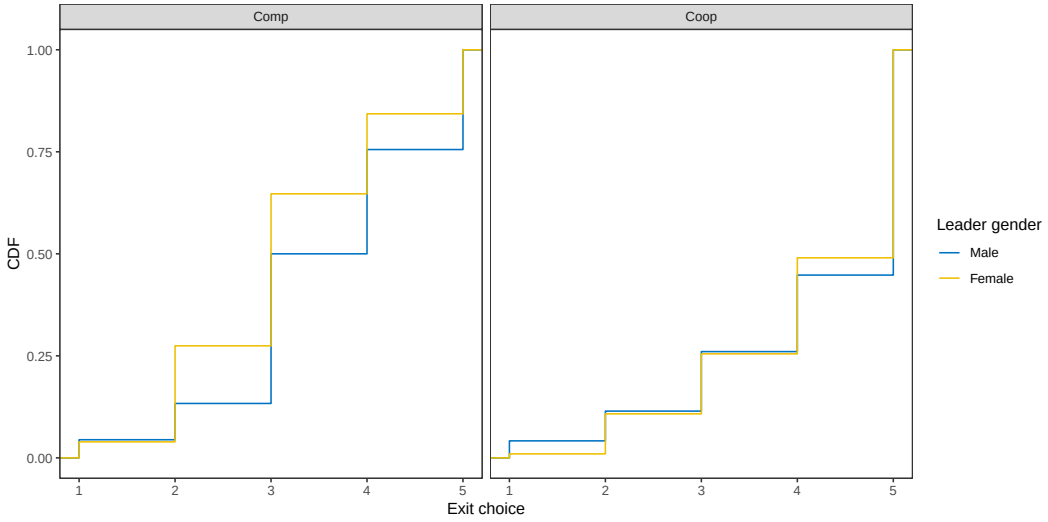


Table C.1: Determinants of exit choices

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Leader	0.85***	0.76***	0.45***	0.44***
	(0.18)	(0.19)	(0.16)	(0.17)
Female	0.04	−0.01	0.06	−0.04
	(0.18)	(0.17)	(0.19)	(0.19)
Leader x Female	−0.23	−0.17	0.44**	0.46**
	(0.24)	(0.25)	(0.22)	(0.22)
Constant	2.93***	0.68	3.77***	4.17***
	(0.15)	(0.87)	(0.14)	(0.81)
Controls	No	Yes	No	Yes
R^2	0.10	0.25	0.12	0.24
Adj. R^2	0.10	0.21	0.11	0.20
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.

Table C.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)
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 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.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 the difference in choices

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	0.32	0.21	0.02	0.03
	(0.25)	(0.29)	(0.25)	(0.29)
Female follower	−0.15	−0.27	0.46*	0.55*
	(0.26)	(0.26)	(0.25)	(0.28)
Constant	0.51**	2.31	0.48**	0.68
	(0.23)	(1.50)	(0.22)	(2.05)
Controls	No	Yes	No	Yes
R^2	0.02	0.25	0.03	0.25
Adj. R^2	−0.00	0.02	0.01	0.02
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

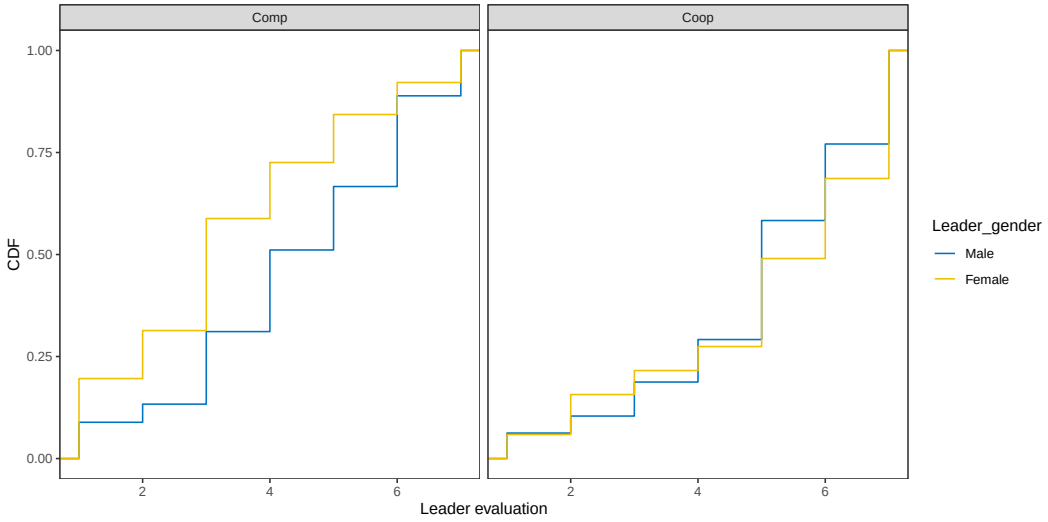
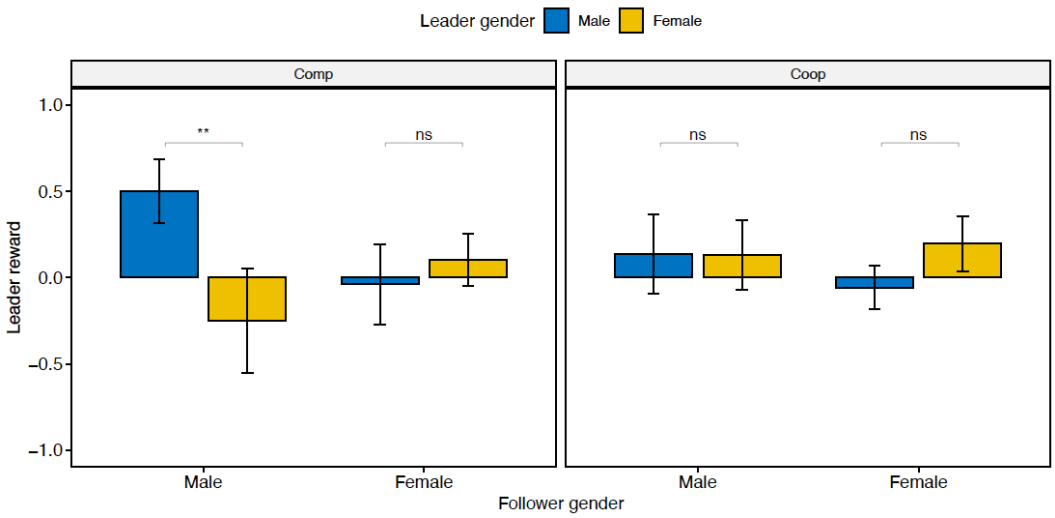


Figure D.2: Rewards by leader and follower gender



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

	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.

Table D.4: 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.5: Determinants of payment adjustment

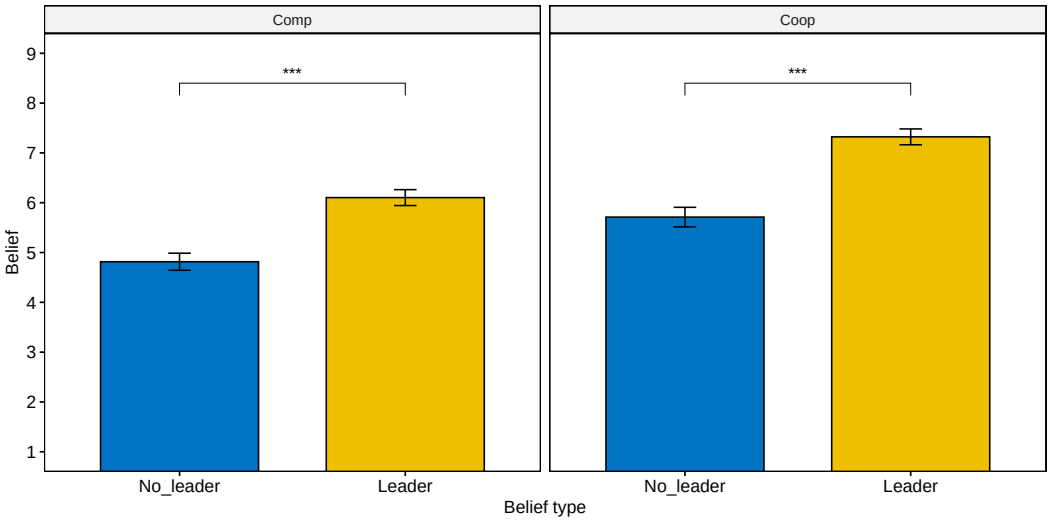
	Comp		*Coop*	
	Model 1	Model 2	Model 3	Model 4
Leader evaluation	0.19***	0.16**	0.05	0.05
	(0.07)	(0.08)	(0.05)	(0.05)
Female leader	0.32	0.27	0.25	0.26
	(0.23)	(0.29)	(0.21)	(0.23)
Male follower	0.54**	0.50*	0.29	0.28
	(0.25)	(0.27)	(0.28)	(0.32)
Belief	0.06	0.09	0.07	0.09
	(0.08)	(0.08)	(0.05)	(0.07)
Female leader x Male follower	−0.84**	−0.75	−0.25	−0.23
	(0.39)	(0.48)	(0.36)	(0.40)
Constant	−1.23**	−1.69	−0.86**	1.32
	(0.50)	(1.07)	(0.42)	(1.32)
Controls	No	Yes	No	Yes
R^2	0.17	0.31	0.06	0.27
Adj. R^2	0.13	0.06	0.01	0.00
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 payment adjustment amount, between -$3 and $3. The control variables are: risk preferences, SVO, age, education, ethnicity, mother's education, and father's education.

E Beliefs

Figure E.1: 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 E.1: 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 E.2: 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.

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 2).

Table G.1: Effectiveness of leaders by advice

	(1)	(2)
Profit max. advice	1.12***	1.20***
	(0.23)	(0.24)
Female leader		−0.11
		(0.15)
Female follower		0.08
		(0.14)
Constant	3.18***	2.68***
	(0.21)	(0.84)
Controls	No	Yes
R^2	0.16	0.26
Adj. R^2	0.16	0.21
Num. obs.	390	390

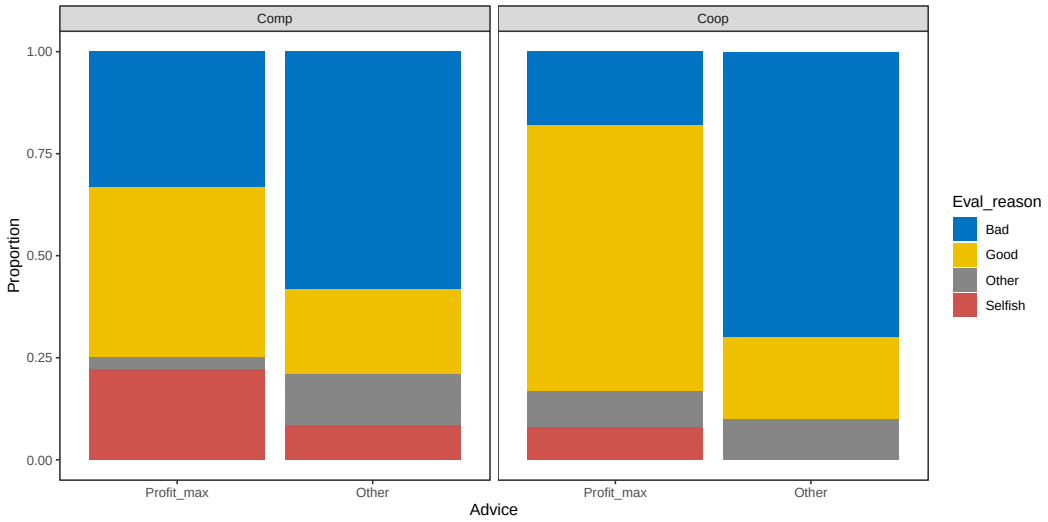
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

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 Other 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

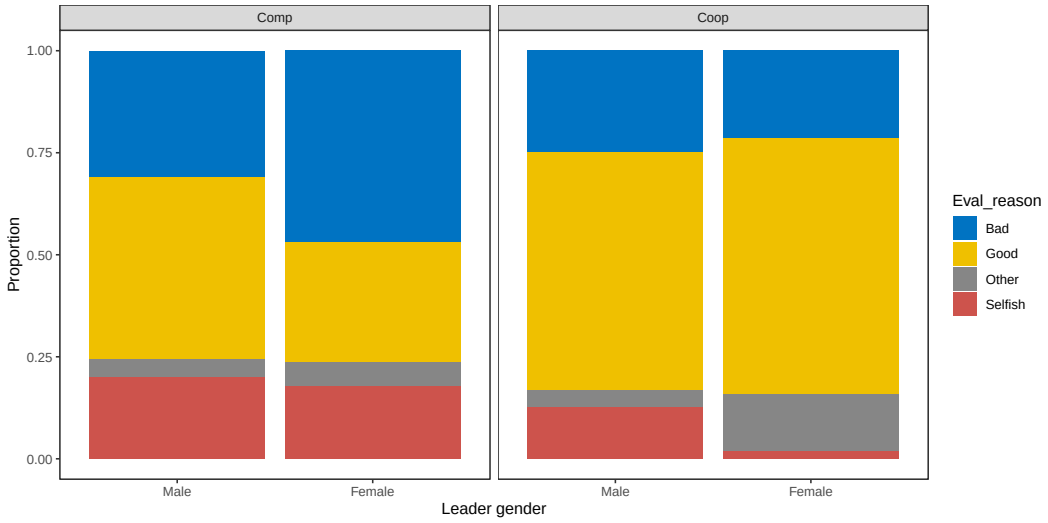


Table H.1: Determinants of leader evaluations by evaluation reason

	Comp		*Coop*	
	(1)	(2)	(3)	(4)
Female leader	−0.27	−0.32	0.33	1.16*
	(0.63)	(0.57)	(0.57)	(0.61)
Bad advice	−1.81**	−1.58**	−2.95***	−1.74**
	(0.83)	(0.77)	(0.73)	(0.82)
Selfish advice	0.44	0.47	−2.67	−0.74
	(0.46)	(0.84)	(1.65)	(1.20)
Other advice	−3.56***	−3.38***	−3.67***	−5.37***
	(0.78)	(0.83)	(0.45)	(0.71)
Female follower	0.08	0.46	0.27	1.05**
	(0.58)	(0.60)	(0.49)	(0.49)
Female leader x Bad advice	−1.10	−1.19	−1.25	−2.88***
	(1.05)	(1.21)	(0.83)	(1.11)
Female leader x Selfish advice	−3.73***	−3.12**	0.39	−1.31
	(0.81)	(1.37)	(0.58)	(0.92)
Female leader x Other advice	5.27***	4.76***	0.00	1.60
	(0.88)	(1.10)	(1.02)	(1.40)
Female leader x Female follower	0.01	−0.47	−0.22	−1.37*
	(0.88)	(1.07)	(0.64)	(0.74)
Female follower x Bad advice	−1.16	−1.63	2.01**	0.12
	(0.93)	(1.08)	(0.79)	(1.03)
Female follower x Selfish advice	−2.33***	−2.66***	1.23	−0.34
	(0.74)	(1.00)	(1.45)	(1.22)
Female follower x Other advice	−2.59***	−3.28***	4.73***	6.91***
	(0.79)	(1.26)	(0.49)	(0.83)
Female leader x Female follower x Bad advice	1.14	1.90	0.64	2.97
	(1.25)	(1.48)	(1.20)	(1.87)
Female leader x Female follower x Selfish advice	2.57**	2.34		
	(1.16)	(1.64)		
Female leader x Female follower x Other advice			−2.11***	−5.05***
			(0.45)	(1.81)
Constant	5.56***	7.92***	5.67***	4.19**
	(0.46)	(1.76)	(0.45)	(1.81)
R^2	0.60	0.66	0.58	0.73
Adj. R^2	0.54	0.47	0.51	0.58
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 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.

I Willingness to lead

Table I.1: Leader evaluations by WTL

	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.2: 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

Welcome

Thank you for participating in this experiment. You will be paid a show-up fee of $5.00 for completing the experiment, and the instructions explain how you can make decisions and earn more money so please read them carefully.

No communication is allowed, except with the experimenter. Please do not use your mobile phones, and make sure you can complete the experiment in a quiet place without distractions.

Today's experiment consists of four parts (Part 1A, Part 1B, Part 2A, and Part 2B). You will receive the instructions for each part after you have finished the previous part. At the end of the experiment, **either Part 1A or Part 1B** will be selected at random for payment, **AND either Part 2A or Part 2B** will be selected at random for payment.

You will be paid via PayID. The experiment will last approximately 60 minutes.

If you have any questions, you may send the experimenter a private message via the Chat function in Zoom.

Survey

Before proceeding to Part 1A, please answer the following questions.

What is your age?

What is your gender?

What is your field of study?

Which year of university are you in?

Next

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

1. The Roles | 2. The Project | 3. Your Decisions | 4. Practice Project 1

Part 1A - Overview

The Roles

In Part 1A, half the participants are **Blue** participants, and the other half are **Red** participants. Each Blue participant is randomly and uniquely paired with a Red participant and together they will participate in a project. You will be informed of your role before you start making any decisions.

If Part 1A is selected for payment, you will receive your earnings from the project.

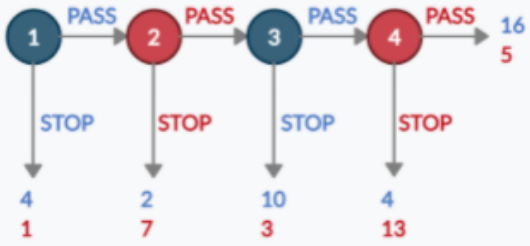
Click on the tabs at the top of the screen to navigate between the instruction tabs.

Part 1A - Overview

The Project

The project involves several rounds in which decisions are made alternately by the Blue and Red participants. For example, in the figure below there are **4 rounds** in total.

If you are a **Blue** participant, you will make decisions at odd numbered rounds (i.e., rounds 1 and 3) in the project. Your earnings are represented in blue. If you are a **Red** participant, you will make decisions at even numbered rounds (i.e., rounds 2 and 4) in the project. Your earnings are represented in red.



As the direction of the arrows indicates, the figure should be read from left to right. At round 1, the Blue participant chooses between STOP and PASS.

- If Blue chooses STOP, then Blue receives **$4** and Red receives **$1**, and the project ends.
- If Blue chooses PASS, then the project continues.

If Blue chooses PASS, it is then Red's turn to choose between STOP and PASS at round 2.

- If Red chooses STOP, then Blue receives **$2** and Red receives **$7**, and the project ends.
- If Red chooses PASS, then the project continues.

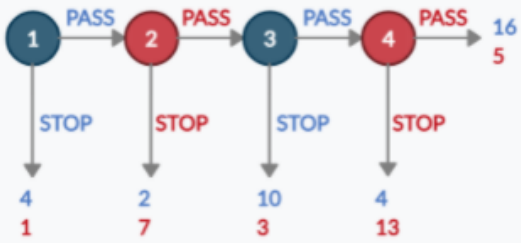
If Red chooses PASS, then Blue can choose again at decision 3 between STOP and PASS. If Blue chooses PASS, then the project continues, and it is Red's turn again to choose at decision 4 between STOP and PASS.

Part 1A - Overview

Your Decisions

Since both participants have two decisions to make in the project, all you have to do is decide **when to STOP, or to ALWAYS PASS**. This means you are choosing between three options:

1. **STOP** at your **first** decision round,
2. **STOP** at your **second** decision round, or
3. **ALWAYS PASS** (i.e., never STOP).



Both participants will make this decision before knowing the other participant's decision.

- If either participant chooses to STOP, the earnings will be determined according to whoever chooses to STOP at the earliest round.
- If both participants choose to ALWAYS PASS, then Blue receives **$16** and Red receives **$5**.

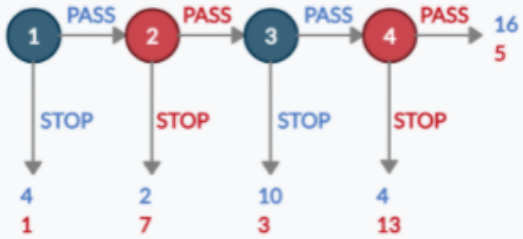
1. The Roles | 2. The Project | 3. Your Decisions | 4. Practice Project 1

Part 1A - Overview

Practice Project 1

You will now participate in a practice project to help you understand the project. This practice project will not be relevant for your earnings.

Suppose you are randomly assigned to be the **Blue** participant. Please make your decision.



I choose to:

- ○ STOP at round 1
- ◉ STOP at round 3
- ○ ALWAYS PASS

Next

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

Decision making in Part 1A

Your Decisions

The project in Part 1A consists of **8 rounds in total**, rather than the 4 rounds in the practice project you just completed.

As before, **Blue** participants decide between STOP and PASS in the odd numbered rounds (i.e. rounds 1, 3, 5, and 7) and **Red** participants decide between STOP and PASS in the even numbered rounds (i.e. rounds 2, 4, 6, and 8).

This means when you are making your decision about **when to STOP, or to ALWAYS PASS**, you are choosing between five options:

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

Both Blue and Red make decisions alternately as before until round 5. Beyond round 5, additional funding for the continuation of the project is uncertain and depends on the random choice of the Computer.

Click on the tabs at the top of the screen to navigate between the instruction tabs.

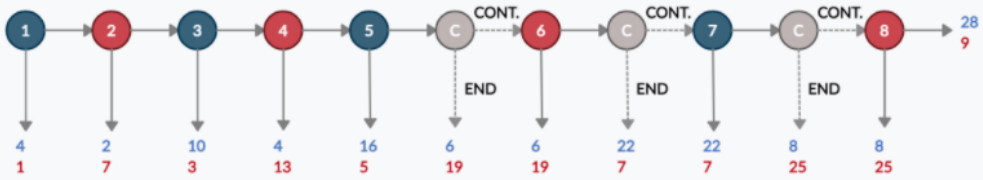
Competitive treatment

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.

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

Decision making in Part 1A

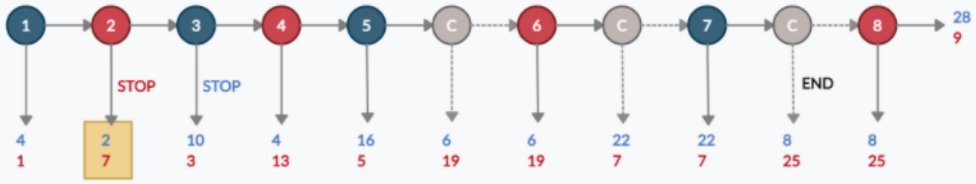
Examples

Below are two examples of possible scenarios to help you understand the project. These examples are for illustration purposes only and are not intended to suggest how anyone should or will behave in the project.

Example 1

Blue chooses to **STOP** at **round 3** and Red chooses to **STOP** at **round 2**. The Computer chooses **END** before **round 8**. Since Red chose to STOP at the earliest round (round 2), the earnings are as follows:

- Blue earns: **$2**
- Red earns: **$7**



Example 2

Blue chooses to **ALWAYS PASS** and Red chooses to **STOP** at **round 8**. The Computer chooses **END** before **round 6**. Since the Computer chose END at the earliest round (before round 6), the earnings are as follows:

- Blue earns: **$6**
- Red earns: **$19**

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

Decision making in Part 1A

Comprehension Questions

Before proceeding to the next screen, you are required to correctly answer the following comprehension questions.

1) If Blue chooses to STOP at round 1, what are the players' earnings?

- ○ Blue: 1, Red: 4
- ○ Blue: 2, Red: 4
- ○ Blue: 4, Red: 1
- ○ Blue: 4, Red: 2

2) The Computer may randomly choose to END funding for the project after which round?

- ○ Round 4
- ○ Round 5
- ○ Round 6
- ○ Round 7

3) If the Computer chooses to END funding for the project before round 7 and both Red and Blue choose to ALWAYS PASS, what are the players' earnings?

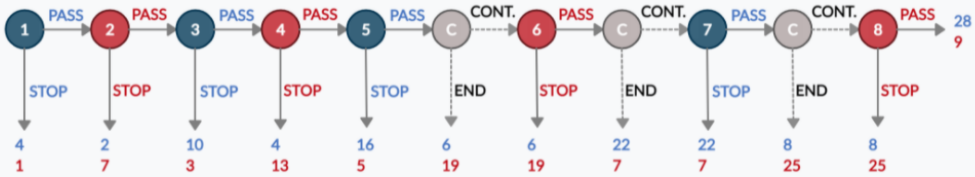
- ○ Blue: 22, Red: 7
- ○ Blue: 8, Red: 25
- ○ Blue: 6, Red: 19
- ○ Blue: 16, Red: 5

Next

Part 1A - Practice Project 2

You will now participate in a second practice project. Again, this practice project will not be relevant for your earnings.

Suppose you are randomly assigned to be the **Red** participant. Please make your decision.



I choose to:

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

Next

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

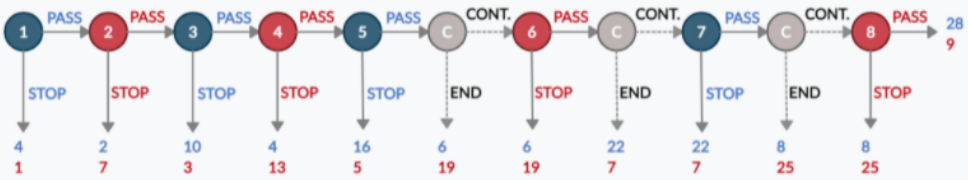
1. The Roles | 2. The Leader | 3. Summary | 4. Comprehension

Part 1B - Overview

The Roles

In Part 1B, you will be assigned into new groups of three participants. Within each group, there will be: one **Blue** participant, one **Red** participant and one **Leader**. You will either be assigned to be a Leader, or you will remain in the same role from Part 1A. You will be informed of your role before you start making any decisions.

Each group will participate in the same project from Part 1A. If Part 1B is selected for payment, you will receive your earnings from this project.



Click on the tabs at the top of the screen to navigate between the instruction tabs.

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

Part 1B - Overview

The Leader

Before the start of the project, the **Leader** will send a message to their group to explain strategies which could improve earnings. The Leader will not have any decisions to make in the project.

The Leader's earnings will equal the **average earnings** of the Blue participant and the Red participant, plus $3. For example, if Blue earns $16 and Red earns $5, then the Leader will earn $3 + ($16 + $5) / 2 = $13.50 from Part 1B.

At the end of the experiment, Blue and Red participants have the option to pay a fee to reward their Leader with an additional $3, or pay a fee to reduce their Leader's earnings by $3. Either the Blue or Red participant will be chosen at random and their decision will be implemented for their Leader.

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

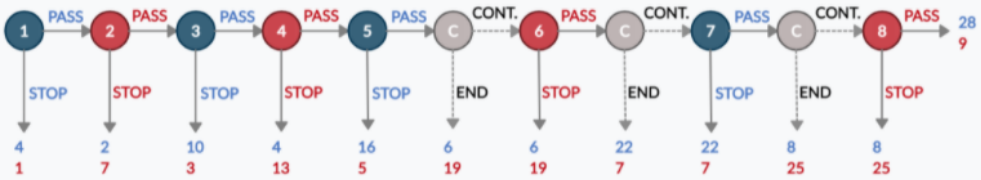
Part 1B - Overview

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:

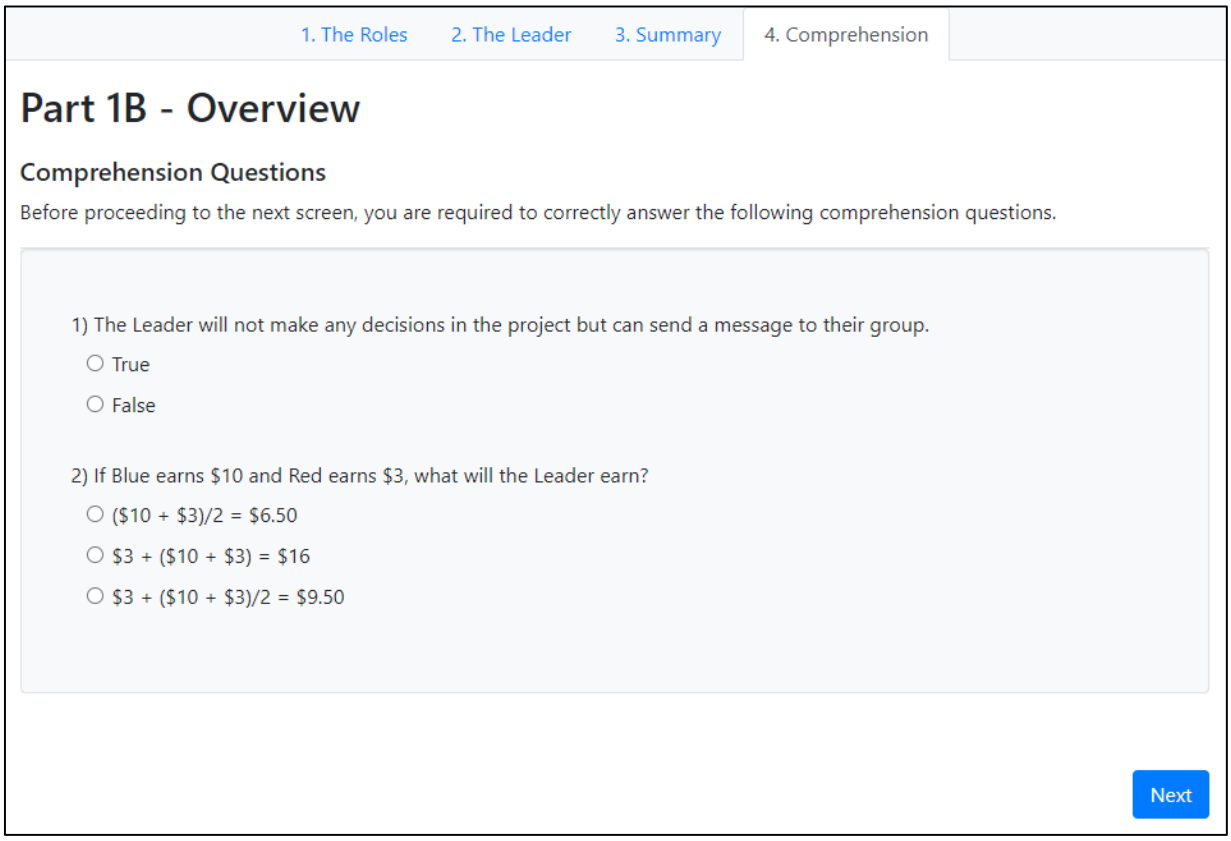
1. **STOP** at your **first** decision round (round **1** or **2**),
2. **STOP** at your **second** decision round (round **3** or **4**),
3. **STOP** at your **third** decision round (round **5** or **6**),
4. **STOP** at your **fourth** decision round (round **7** or **8**), or
5. **ALWAYS PASS** (i.e.,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

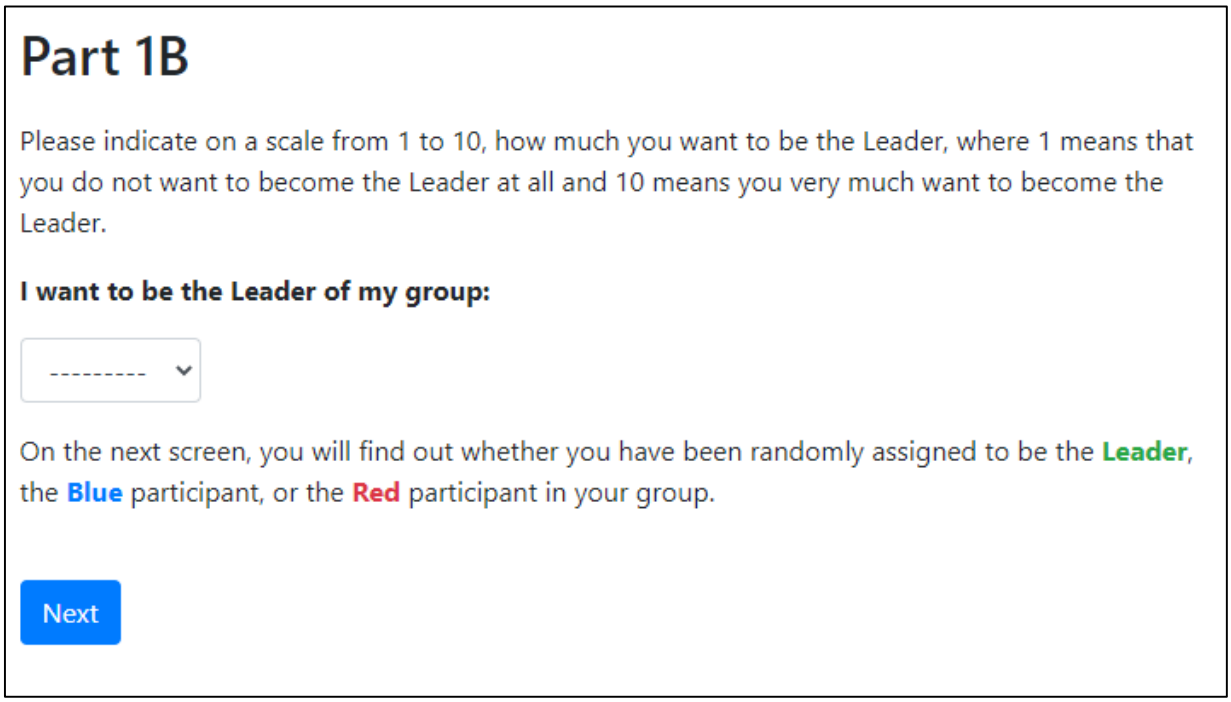


After **round 5**, the Computer may randomly choose to **END** funding for the project **before either round 6, 7, 8, or never**, with equal chances. If the Computer's choice is implemented, then this is equivalent to funding for the project reaching the next round and STOP is chosen. The project will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


1. The Roles
2. The Leader
3. Summary
4. Comprehension
Part 1B - Overview
Comprehension Questions
Before proceeding to the next screen, you are required to correctly answer the following comprehension questions.
1) The Leader will not make any decisions in the project but can send a message to their group.
True
False
2) If Blue earns $10 and Red earns $3, what will the Leader earn?
($10 + $3)/2 = $6.50
$3 + ($10 + $3) = $16
$3 + ($10 + $3)/2 = $9.50
Next


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.

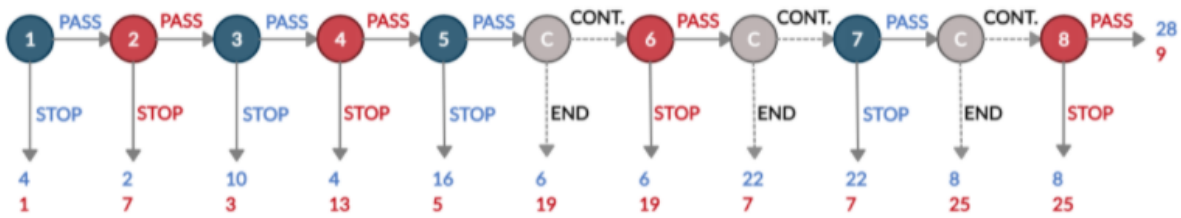
Your earnings will be maximized if your group members choose to **PASS** at every node. Please enter a message for your group.

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:

Next

SVO

Part 2A

In Part 2A, you have been randomly paired with another participant, whom we will refer to as the **Other**. You are asked to make a series of decisions about allocating resources (in units) between you and this person. Please indicate the distribution you prefer the most.

At the end of the experiment, one of the allocations will be randomly selected by the Computer. The Computer will also randomly select one participant in each pair and this participant's decision will determine your payoffs. If Part 2A is chosen for payment, we will use the following exchange rate: 15 units = $1.00.

Next

Part 2A - Item 1

For each of the following questions, please indicate the distribution you prefer the most where the first allocation is what you receive and the second is what Other receives, **i.e. (what You receive, what Other receives).**

Item 1:

- ○ (85, 85)
- ○ (85, 76)
- ○ (85, 68)
- ○ (85, 59)
- ○ (85, 50)
- ○ (85, 41)
- ○ (85, 33)
- ○ (85, 24)
- ○ (85, 15)

Next

Part 2B

In Part 2B, you have the chance to earn an additional payment, depending on your choice in this question. You will select one lottery from among the 6 different lotteries below.

Each lottery has two possible outcomes (Heads or Tails), chosen randomly by the Computer with each outcome equally likely to occur.

If Part 2B is chosen for payment, your earnings will depend on: 1) the lottery you choose, and 2) the outcome that occurs (Heads or Tails). We will use the same exchange rate of: 15 units = $1.00.

For example:

- If you select Lottery 2 and Heads occurs, you will receive 54 units. If Tails occurs, you will receive 36 units.
- If you select Lottery 5 and Heads occurs, you will receive 90 units. If Tails occurs, you will receive 18 units.

Lotteries	Heads	Tails
○ 1	42	42
○ 2	54	36
○ 3	66	30
○ 4	78	24
○ 5	90	18
○ 6	105	3

End of Part 2B. On the next page you will be asked to answer some survey questions.

Next

Survey

Were you born in Australia?

- ○ Yes
- ○ No

What is your ethnicity?

What is the highest level of education you have completed?

What is the highest level of education your mother has completed?

What is the highest level of education your father has completed?

Next

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:

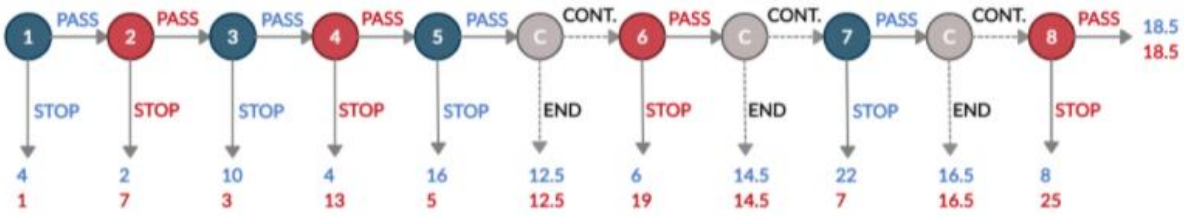
1. **STOP** at your **first** decision round (round **1** or **2**)
2. **STOP** at your **second** decision round (round **3** or **4**)
3. **STOP** at your **third** decision round (round **5** or **6**)
4. **STOP** at your **fourth** decision round (round **7** or **8**)
5. **ALWAYS PASS**

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.

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 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