Effects of team gender composition on women´s leadership aspirations

Andreas Born[a] , Eva Ranehill[b], and Anna Sandberg[c]

[a] Department of Economics, Stockholm School of Economics
[b] Department of Economics, University of Gothenburg
[c] Swedish Institute for Social Research, Stockholm University

April 30, 2018

Despite increased female educational attainment, labor markets remain vertically and horizontally segregated. In this study, we explore whether male dominated environments, in and of themselves, adversely affect women. We employ a laboratory experiment with a total of 580 participants of both genders, randomly assigning participants to either female majority teams (three women and one man) or male majority teams (three men and one woman). In line with our primary hypothesis, we find that women are significantly more reluctant to become the team leader in male majority teams than in female majority teams. Analyses of potential mechanisms show that women in male majority teams are also less confident, less influential, more swayed by the team discussion, and (accurately) believe that they will receive less support from their team members. These findings suggest that the absence of women in male dominated contexts may be a self-reinforcing process, whereby the women who end up in male majority environments find themselves less confident in their own ability and, thus, discouraged from taking on a leadership position.

Keywords: leadership; gender differences; experiment

1. Introduction

Despite important advancement in women´s educational attainment over the last decades, labor markets around the world remain characterized by vertical and horizontal gender segregation. A large literature shows that women remain underrepresented in leading positions (e.g. Bertrand and Hallock 2001; Bertrand 2011). Further, while recent evidence suggests that the gender wage gap continues to decline, the decline is not uniform across the wage distribution. The gender wage gap is noticeably larger, and appears the most persistent, at the top of the wage distribution (Blau and Kahn 2016; Albrecht, Bjorklund, and Vroman 2003; Arulampalam, Booth, and Bryan 2007). Today, conventional human capital variables explain only a small part of the gender wage gap, while occupation and industry remain important factors (Blau and Kahn 2016). This raises the question why, despite important female advancement in education and human capital accumulation, women remain in educations with lower earnings potential, in lower paid industries, and at lower paid levels.

While many factors are likely to contribute to labor market gender gaps, one thing holds true for all areas where women are underrepresented, independent of the nature of the work, the industry, and the level – they are male dominated. In this paper, we contribute to previous literature by asking whether this, in itself, may have an adverse impact on women's careers. In particular, we ask if, and if so why, women´s leadership aspirations are negatively influenced by environments where men are in the majority.

If, for example, women´s confidence, wellbeing, or behavioral strategies are adversely affected by being surrounded by many men in professional contexts, the absence of women in certain sectors, and at certain positions in the organizational hierarchy, may become a self-perpetuating cycle consistent with various gender related labor market gaps. For example, in addition to being consistent with both horizontal and vertical gender segregation, this phenomenon may also help explain the "gender tipping points" observed in several occupations and sectors – i.e., the fact that the share of women in a sector often increases only slowly until a critical mass is reached, and then starts increasing rapidly towards a majority of women.[1] Many studies also suggest that women's job satisfaction is lower in workplaces dominated by men. Thus, a negative effect of male dominated environments on women would be in line with several important labor market phenomena.

[1] England et al. (2007) identifies tipping points in the gender distribution of PhD students receiving doctorates in different fields, and Pan (2015) identifies tipping points in the gender distribution of occupations.

Confounding factors and data limitations make it very difficult to nail down the *causal* impact of an environment's gender composition on women's behavior and well-being using observational data. Most importantly, selection (by employees and/or employers) into different industries and organizations, or other types of naturally occurring groups, is far from random. Women who work in male dominated environments differ from other women along a number of unobservable characteristics that may correlate with behavior, productivity and well-being (Kirn et al. 2018). In the current study, we avoid these selection issues through random assignment of participants to groups with varying gender composition in an economic experiment. Importantly, the experimental approach allows us to stratify the randomization, making it possible to study group compositions that are scarce in naturally occurring data (for example, in the field, there is often a lack of observations of executive committees or company boards with a female majority). Finally, in the laboratory we can systematically explore a number of mechanisms through which male dominated environments may impact women's behavioral strategies and outcomes. Specifically, in addition to leadership aspiration, we also measure whether the gender composition of groups impact performance, confidence, influence, and actual and expected support from the other group members.

In the experiment, 580 participants were randomized to either female majority teams, comprising three women and one man, or male majority teams, comprising one woman and three men. All participants were first asked to solve a task individually, before they sat down together with their team for 10 minutes to discuss the task and come up with a joint solution. Then, before solving a second, similar task, each team elected a leader. The leader was to decide, after receiving input from the other team members, on a joint team solution for the second task. Before the election, all team members were required to indicate how much they wanted to become the leader. Then, everyone voted for their preferred leader by ranking the other three team members. The two team members with the highest willingness to become the leader became candidates in an election, and the candidate with the most votes became the team leader.[2]

In accordance with previous literature, we find that women are significantly more reluctant than men to become the team leader. On average, on the 1-10 response scale, women state a willingness to become the team leader that is 1.63 units lower than that of men (Cohen's d = 0.56). The modal response for men is 10, indicating the highest possible interest in leading

[2] Participants thus provided their votes before the candidates were announced. In order to avoid strategic voting, only the votes of the two group members who were *not* candidates in the election counted. The procedures of the election, and their implications, were carefully explained to the participants in advance, and are described in detail in Section 2.2.

their team, while the modal response for women is one, indicating the lowest possible interest in leadership. In support of our main hypothesis, we also find that women are significantly more reluctant to become leaders in male majority teams than in female majority teams. On average, women in male majority teams state a willingness to become the team leader that is 1.39 units lower than that of women in female majority teams (Cohen's d = 0.46). Thus, the impact of team gender composition on female leadership aspirations is relevant in size and comparable in magnitude to the overall gender gap.

Neither the general gender gap in leadership aspirations, nor the effect of team gender composition on female leadership aspirations, can be explained by gender gaps in task related ability. In fact, according to our point estimates, the men who perform the worst in their team state, on average, a higher willingness to lead than the women who perform the best.

Women´s lower leadership aspirations translates, in our setting, to a lower likelihood of becoming a candidate in the election, and, subsequently, a lower likelihood of becoming the team leader. Further, in line with the results presented above, women in male majority teams are significantly less likely than women in female majority teams to become a candidate in the election. Overall, the average likelihood of becoming a candidate is 50 percent. For women, the corresponding number is 44 percent in female majority teams, but only 29 percent in male majority teams.

Our exploratory analyses of potential mechanisms show that, compared to women in female majority teams, women in male majority teams believe that they perform worse relative to the other team members. They are also less influential, more swayed by the team discussion, receive fewer votes in the election, and are less optimistic about how many votes they will receive. Low relative performance beliefs and low expectations of electoral support seem to be particularly important factors discouraging women in male majority teams from aspiring for leadership. Controlling for relative performance, women in male majority teams believe, on average, that they are ranked 0.37 positions lower in terms of performance and 0.26 positions lower in terms of election outcomes, compared to women in female majority teams. The negative impact of male majority teams on female leadership aspirations is reduced by roughly half when controlling for relative performance beliefs in a regression framework, and it is reduced by 64% when controlling for beliefs about electoral outcomes.[3]

Our results show that women are, at least in some respects, adversely influenced by being in numerical minority. Thus, we contribute to the understanding of the surprising

[3] Note, however, that these decreases are not statistically significant.

persistence of gender gaps by proposing a mechanism that works as a self-reinforcing process, whereby women become reluctant to enter, and prone to leave, male dominated environments.

These results contribute to our understanding of gender gaps in labor markets in several different ways. In addition to the research previously mentioned regarding vertical and horizontal gender gaps, tipping, and job satisfaction, our findings speak directly to the current debate about gender quotas at top levels.[4] If under representation causes women to become less confident and lower their aspirations, policies that increase the share of women in, for example, corporate boards, may yield benefits for all women on the board. Thus, in addition to the effects that are currently brought up in the public debate, affirmative action policies may have other important long run consequences for the retention of women on corporate boards.

Further, during recent decades, production arrangements in most industries have shifted towards teamwork-based solutions. For example, between 1987 and 1999, the share of large US firms with more than 20% of employees working in teams increased by 65 percent (Lazear and Shaw, 2007). Thus, the behavioral effects of team gender composition on employees should, if anything, become more relevant over time. From a broader perspective, our results speak to the recognition, retention, and promotion of competence in organizations, and our findings should be of interest to policy makers and employers aiming to attract and retain more women, and the most competent individuals, to top positions and male dominated occupations.

The remainder of the paper is organized as follows. Section 2 presents the experiment design. In the third section, we present our main results, exploring whether women in male dominated teams are less willing to lead their team than women in female majority teams. Section 4 presents the more exploratory analysis on what mechanisms may contribute to the impact of team gender composition, and Section 5 concludes.

2. Experiment Design

The main part of the experiment comprised ten stages, during which participants worked on two tasks: the "Lost at Sea" task and the "Desert Survival" task. Below, we first describe the two tasks, and then the ten stages of the experiment.

[4] Since Norway was the first country to pass a gender representation law for corporate boards in 2003, Belgium, France, Ireland, Iceland, Italy, Malaysia, the Netherlands, and Spain have adopted similar measures. Further, The European Commission has recently proposed legislation with the aim that, by 2020, 40% of non-executive directors shall be women.

2.1 Tasks

In the Lost at Sea task and the Desert Survival task, participants were asked to rank ten items in terms of their importance for survival in a hypothetical scenario. The Lost at Sea task briefly described a boat accident leaving a group of people on an inflatable life raft in mid Atlantic. Similarly, the Desert Survival task described a plane accident, leaving a group of people stranded in the desert, far away from any human settlements, at an imprecise location. In each scenario, the survivors had ten items that were left undamaged in the accident. The task of the participants in the study consisted of ranking these ten items based on their importance for the group´s survival.[5]

In order to calculate a participant's payoff for either of these tasks, the participant's answer was compared to an answer provided by survival experts. The closer the participant's answer was to that of the expert panel, the higher payoff did the answer generate. For each item that the participant ranked differently than the experts, they lost points. The number of points lost for an item corresponds to the number of ranks between the participant's proposed rank and that of the experts. The total number of lost points in the task was the sum of the points lost over all ten items. Thus, the total number of lost points could range between 0 (perfect solution) and 50 (worst possible solution). The final payment for each of these tasks, in CHF, was calculated according to the formula:

$$Payment = (100 - Total\ number\ of\ lost\ points)/2$$

Thus, the lowest possible payment in the task was CHF 25.

We chose these particular tasks for several reasons. First, since we are interested in the behavior of women in male dominated areas, we wanted tasks that are gender neutral or have a slight male stereotype.[6] We also needed tasks that are suitable for both team work and individual work, and where the correct answer is open for discussion while still perceived as meaningful and related to individual competence. Complete descriptions of the tasks are

[5] These tasks are based on team building exercises that originally consisted of 15 items. We simplified these tasks such that only 10 items remained to rank. The items in the Lost at Sea task were: *mosquito netting, a mirror, a container of water, a case of army rations, maps of the Atlantic Ocean, a floating seat cushion, a can of oil/petrol, a transistor radio, some plastic sheeting,* and *rope*. The items in the Desert Survival task were: *a mirror, an overcoat, water, a torch, a parachute, a folding knife, a pistol, a first-aid kit, a book about animals that can be eaten*, and *a bottle of salt tablets*.

[6] In the questionnaire after the experiment, participants were asked to indicate what gender, if any, was better at the task. Participants answered on a scale between 0 and 10, where 0 indicates that men are better and 5 indicates that both genders are equally good. The average answer (mean=4.65, s.d.=1.15) is significantly different from 5 and confirms that the task is considered to be somewhat stereotypically male.

available in the Online Appendix. No information on the participant's performance was provided before the end of the study.

2.2 Stages

The main experiment consisted of ten stages, as summarized in Figure 1. At the end of the experiment, one of the stages 1, 2, 3, 8 or 9 was randomly selected to count for the participant's final payment.

In ***Stage 1***, before any teams were formed, participants worked individually for 8 minutes on their computer, providing their answer to the Lost as Sea task. If this stage was chosen for payment, participants were paid based on their individually submitted answers.

Participants were then divided into teams of four for ***Stage 2***, such that each team comprised either three men and one woman, or one man and three women. Each team moved to a separate room and had 10 minutes to discuss the Lost at Sea task and agree on a joint ranking of the 10 items. After the team discussions, participants went back to their own computer.[7] If Stage 2 was chosen to count for payment, all team members were paid the same amount of money, based on their team's answer.

Then, in ***Stage 3***, participants were given the opportunity to update their previous, individual answer. To do so, they were presented with a screen displaying a table with both their own individual answer from Stage 1 and their team's answer from Stage 2. Participants were then asked to enter an individual answer once again, and were free to make updates as they saw appropriate. Any payment for Stage 3 was based on this individual, possibly updated, answer.

After the three initial stages of individual and team performance, participants were, in ***Stage 4***, asked to guess how well they performed in Stage 1, compared to the other participants in their team. The participants answered on a scale from 1 to 4, where 1 implied that they believed that, compared to their team members, their own answer was closest to the answer provided by the expert panel. These performance beliefs were incentivized such that, were Stage 3 randomly chosen for payment, participants received an additional CHF 2 in case their guess in Stage 4 was correct.

[7] During the team discussion, all team members were asked to fill in an answer sheet, ensuring that they had their team's answer on paper when they returned back to their computers. After the discussion, each participant had to enter their team's answer on their own computer, and the study did not proceed until all team members had filled in the same answer.

Our main outcome variable, how much participants wanted to become the leader of their team, was elicited in ***Stage 5***. Participants were informed that they would once again perform a task that resembled the Lost at Sea task, but involving a different survival situation and 10 different objects. The new task would be performed in the same teams, but differed from the previous task in how the teams were to decide on a joint answer. Instead of a face-to-face discussion, the team would now elect a leader. The leader would be responsible for providing the final team answer, after seeing all the other team members' individual answers on his/her computer screen. This responsibility of the leader thus includes several important aspects of what is generally associated with a leadership role. For example, it includes taking in and synthesizing information from others, but also asks for personal expertise and the willingness to make and implement a final decision for the team. All team members were asked to indicate how much they wanted to become the team leader on a scale between 1 and 10. They were informed that the two team members who indicated the highest willingness to become the leader would become candidates in an election.

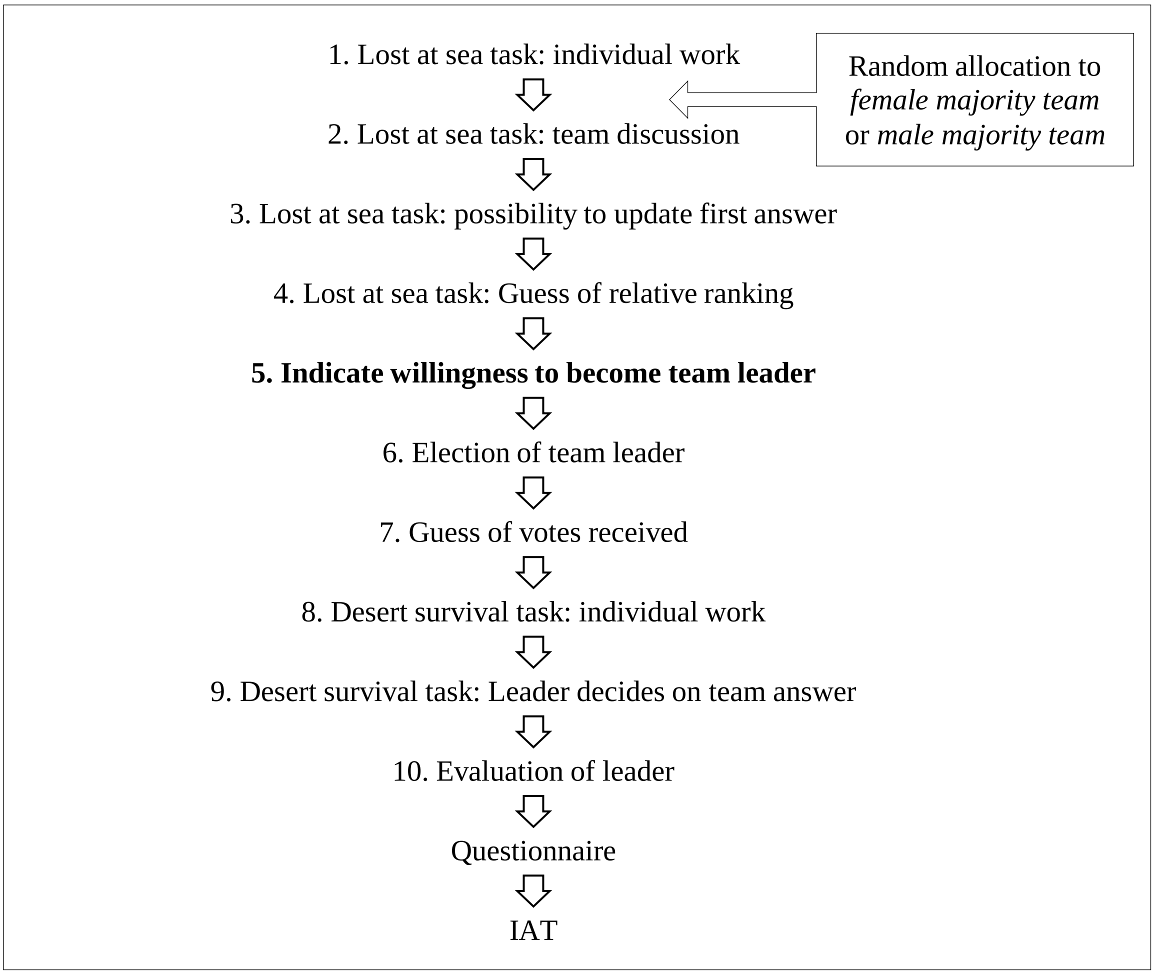


Figure 1. Overview of experiment.

The leader was elected in ***Stage 6***. The election was implemented such that all team members provided their anonymous vote through a ranking of the other three team members, with their most preferred leader at position 1, and their least preferred leader at position 3. Participant provided their ranking before the two candidates were revealed, and they were not allowed to vote for themselves. To determine the outcome of the election, the two candidates' ranking points were compared and the candidate with the lowest sum was elected leader. In order to limit strategic voting, the votes from the candidates themselves were not counted in the election, and ties were broken randomly. The instructions carefully explained the procedures of the election, and their implications.

In ***Stage 7***, before participants were informed about the outcome of the election, they were asked to guess how their average rank in the election compared to the other team members. All participants answered by providing a number between 1 (implying they were ranked first) and 4 (implying they were ranked last). If the participant's guess was correct, CHF 2 were added to the participant´s payment for Stage 8.[8]

After the election, in ***Stage 8***, participants were introduced to the Desert Survival task. This stage resembled Stage 1 in that each participant, including the team leader, individually ranked the 10 items in the Desert Survival scenario in terms of their importance for survival. If this stage was chosen for payment, participants were paid according to the same principle as in the Lost at Sea scenario in Stage 1.

When all team members, including the leader, had provided an individual ranking, their rankings were transmitted to the team leader in ***Stage 9***. The rankings were presented side by side in a comparable way in a large table on the leader´s screen. The leader then had 6 minutes to submit a final, joint answer for the team. When submitting the team answer, the leader was free to consider the other team members' proposals or not. If this stage was selected for payment, all participants were paid based on this team answer, and the payment for the team answer was calculated in the same way as before.[9]

Finally, in ***Stage 10***, all team members were informed about the leader´s answer, and the resulting payoff should this stage be chosen for payment. They were then asked to evaluate

[8] After having provided their guess, all team members were informed about the identity of the two candidates and who won the election. The exact number of votes that each team member received was not revealed.

[9] While the leader worked on the final team answer, the other three team members performed the same task, albeit unincentivized.

their leader's performance on a scale from 1 to 10, while the leaders were asked to evaluate their own performance.

The experiment ended with a short questionnaire and an IAT, eliciting the strength of the participants' implicit associations between leadership and being male. The questionnaire included questions about the participants' nationality, parental education, political orientation, willingness to take risks, beliefs about gender differences in task performance, and previous leadership experience. Participants were also asked to provide a brief motivation for why they wanted to become the leader or not.

2.3 Experiment Procedures

The experiment was conducted at the laboratory for experimental economics at the University of Zurich in May 2017. We ran 30 sessions, and recruited 20 subjects (i.e. 5 teams) to participate in each session. For each session we invited roughly the same amount of men and women, aiming for 9 men (women) and 11 women (men). If not enough participants showed up for a session, the session was run with 16 participants, or 4 teams, as worked best given the gender composition of the participants present. In total, 5 sessions were run with only 4 teams, yielding a total of 580 participants (145 teams) in the final sample. Table 1 provides an overview of the sample size by gender and treatment. All participants were students from the University of Zurich or the Swiss Federal Institute of Technology, and their average age was 23 years.

Table 1. Number of Observations

	Female dominated team	Male dominated team	Total
Women	207	76	283
Men	69	228	297
Total	276	304	580

All instructions can be found in the online appendix. Instructions were distributed directly before each relevant part of the experiment. Control questions were asked before participants performed the task for the first time in Stage 1 (related to the assessment of the task and the resulting payment), before the participants indicated their leadership aspirations in Stage 5 (related to how candidates were selected), and before the leader was elected in Stage 6 (related to the electoral procedures). The study only advanced after all participants had answered the questions correctly. The experiment was programmed in Ztree (Fischbacher, 2007). The average payment was CHF 49 and sessions lasted about 1.5 hours.

In the analysis below, the non-parametric tests reported are Mann-Whitney-Wilcoxon tests unless otherwise stated. All hypothesis tests reported in the paper are two-sided. Before exploring the data, we published a pre-analysis plan available at the Open Science Framework (https://osf.io).

3. Main Results

In this section, we first consider the general gender gap in leadership aspirations. Then, we turn to our main research question, and explore whether team gender composition affects female leadership aspiration. Section 4 presents an exploratory analysis of potential mechanisms and Section 5 discusses team outcomes.

3.1 The gender gap in leadership aspirations

Figure 2 presents the distribution of leadership aspirations by gender. On the 1-10 scale, 1 indicates the lowest possible interest in leading the team, and 10 indicates the highest possible interest. On average, men state a willingness to become the team leader that is 1.63 units higher than that of women (men: 7.27, women: 5.63, Cohen's d=0.56). This average gender difference in leadership aspirations is highly significant ($p<0.001$) and a Kolmogorov-Smirnov test further confirms that the distributions differ significantly by gender ($p<0.001$). In fact, the modal response for men (given by 27% of the male participants) is 10, indicating the highest possible interest in leading their team, while the modal response for women (given by 17 % of the female participants) is 1, indicating the lowest possible interest in leadership.

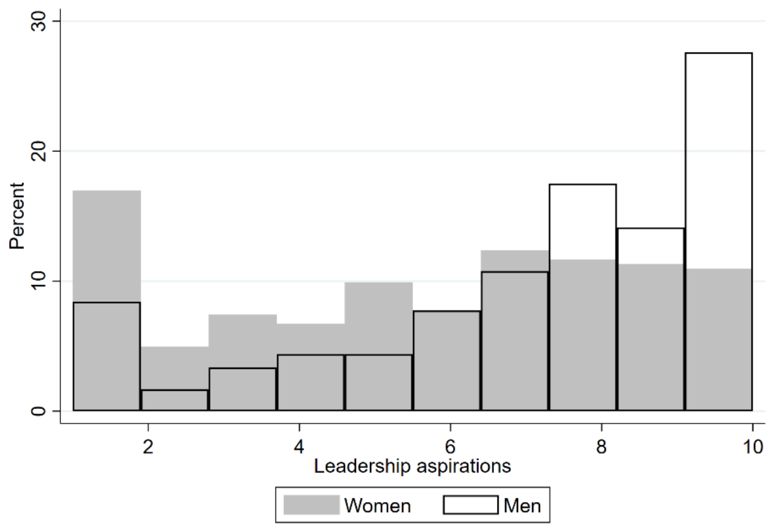


Figure 2: Distribution of leadership aspirations by gender

As illustrated in Table 2, we further analyze gender gaps in willingness to lead in an OLS regressions with leadership aspiration as the dependent variable.[10] The first column controls only for the gender of the participant and illustrates the large and highly significant gender gap in willingness to lead presented above. In the second column, we add a control variable indicating the participant's individual performance in the first Lost at Sea task relative to the other team members (i.e., whether the participant was the best, 2nd best, 3rd best, or worst in their team).[11] Controlling for relative performance is important, as male participants perform on average 1.8 points (0.26 standard deviations) better than women at this first, individual, task ($p=0.001$, see Figure A.1 in the Appendix for the distributions). However, also in this specification, the estimated gender gap in average leadership aspirations remains highly significant and sizable at 1.58 ($p<0.001$). Thus, the observed gender gap in willingness to become the team leader cannot be accounted for by gender differences in the ability to solve the task.[12]

[10] All results in Table 1 are robust to running Tobit regressions (left-censored at 1 and right-censored at 10) or ordered Probit regressions instead of OLS regressions. These results are available from the authors on request.

[11] When computing relative performance, the best performance is ranked 1 and there is no correction for ties. That is, the rank is 1 + the number of performances that are better than the participant's own performance.

[12] In Table 1, we control for relative performance using an ordinal variable taking the value 1, 2, 3 or 4. All results are robust to instead using dummy variables for relative performance as controls, or controlling for absolute performance (number of penalty points in the task). These results are available from the authors on request.

Table 2: Differences in leadership aspirations across gender and team composition

	Dependent variable: Leadership Aspiration (1-10)			
	(1)	(2)	(3)	(4)
Female	-1.633***	-1.584***	-2.491***	-2.123**
	(0.261)	(0.262)	(0.470)	(0.657)
Female majority team			0.673	0.650
			(0.355)	(0.353)
Female X Female majority team			0.713	0.706
			(0.593)	(0.588)
Constant	7.266***	7.887***	7.110***	7.559***
	(0.166)	(0.312)	(0.195)	(0.355)
N	580	580	580	580
Controls:				
Relative performance Stage 1		YES		YES
Female X Relative perf. Stage 1				YES
F-test:				
Female majority team + ‘Female X Female majority team':			1.386*** (F=11.816)	1.355*** (F=11.377)

* p<0.05; ** p<0.01; *** p<0.001.

Note: OLS regressions using leadership aspirations (1-10) as the dependent variable. Standard errors are clustered at the team level. The final row shows results from an F-test, testing the treatment effect for women.

3.2 The effect of team gender composition

Our main research question is whether team gender composition affects women’s leadership aspirations. Figure 3 presents the average leadership aspiration by gender and treatment. Women in female majority teams state an average willingness to become the leader of 6.0, whereas the corresponding number for women in male majority teams is 4.6, and this difference is highly statistically significant (p=0.001, Cohen´s d=0.46). We thus find support for our first hypothesis that women are significantly less willing to become the leader of male majority teams than of female majority teams. Further, the impact of team gender composition on women’s leadership aspirations is relevant in size, and, although slightly smaller, comparable in magnitude to the average gender gap between men and women.

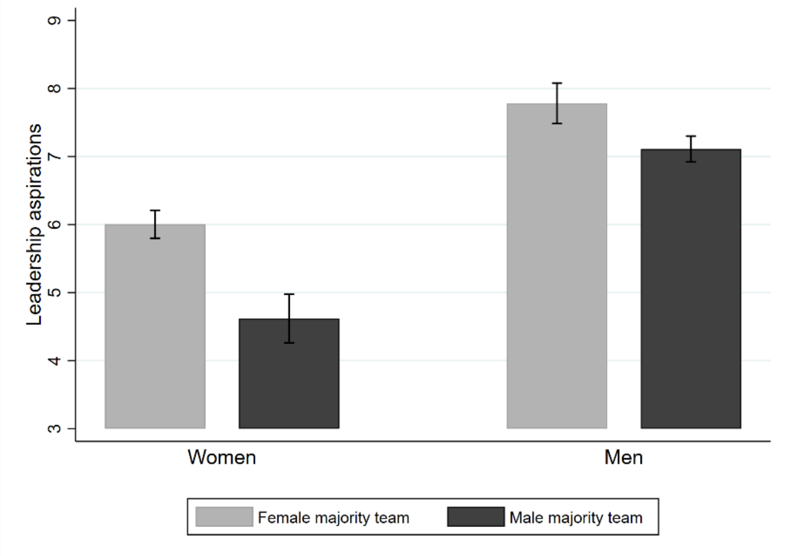


Figure 3: Distribution of leadership aspirations by gender and treatment

Note: Error bars represent standard errors.

A Kolmogorov-Smirnov test confirms that the distribution of women's leadership aspirations differs by team gender composition in important ways (p=0.005). For example, only 14% of women in female majority teams state the lowest possible willingness to become the team leader, compared to 25% in male majority teams. Further, in female majority teams the modal response among women is not 1, as in male majority teams, but 8 (see Figure A.2 (a) in the Appendix).

In the two final columns of Table 2, we add as explanatory variables to the first two specifications an indicator variable for team gender composition, and an interaction between this variable and the previously included gender dummy. The estimated interaction between *Female* and *Female majority team* is positive, indicating that the gender gap in leadership aspirations is smaller in female majority teams than in male majority teams. However, the effect is not statistically significant (p=0.231). Thus, even though the point estimate goes in the hypothesized direction, we do not find support for our second hypothesis, that the gender gap in leadership aspirations is significantly smaller in female majority teams compared to male majority teams.

One of the reasons why the gender gap in leadership aspiration is not significantly smaller in female majority teams than in male majority teams is that men are also more willing to become the leader of female majority teams. Men in female majority teams state an average willingness to become the leader of 7.78, whereas the corresponding number for men in male majority teams is 7.11 However, the effect for men is smaller than for women and not statistically significant (p=0.133). We further cannot reject the null hypothesis that the distributions of male leadership aspirations are the same in both treatments (p=0.102, Kolmogorov-Smirnov test, see Figure A.2 (b) in the Appendix).

As a consequence of men´s slightly higher performance in the first task, women’s individual performance relative to that of the other team members is, on average, slightly worse in male majority teams than in female majority teams. Although this difference is quite small (average rank of 2.57 vs. 2.47) and insignificant (p=0.50), specification 4 in Table 2 accounts for this by controlling for relative performance.[13] As shown, the estimated impact of team gender composition on women’s leadership aspirations (given by the sum of the coefficients of *Female majority team* and *Female X Female majority team*) remains stable (1.39 vs. 1.36) and significant ($p<0.001$).[14] Thus, the observed impact of team gender composition on women’s willingness to lead cannot be accounted for by performance differences.

In order to get an impression of whether the gender and treatment differences in willingness to lead vary by relative performance, Figure 5 shows the average leadership aspirations by gender, treatment and the participants’ performance ranking. The figure illustrates that, regardless of relative performance and treatment team, the average leadership aspiration of men is always higher than that of women. In other words, in point estimates, the men who perform the *worst* in their team are, on average, more willing to lead than the women who perform the *best* in their team.[15] Moreover, for each level of relative performance, men in female majority teams have the highest willingness to lead, followed by men in male majority teams, and women in female majority teams. Women in male majority teams consistently state the lowest willingness to lead.

[13] In column 4, we also interact the measure of relative performance with *Female*. Thus, we allow for gender-specific effects of relative performance on the outcome variable.
[14] For men, the estimated treatment effect (given by the coefficient of *Female majority team*) drops from 0.67 to 0.65, and remains insignificant, when controlling for relative performance.
[15] If we, for example, compare the leadership aspirations of the women who were best in their team (N=65) with the men who performed the worst (N=57) the difference is not significant (p=0.144).

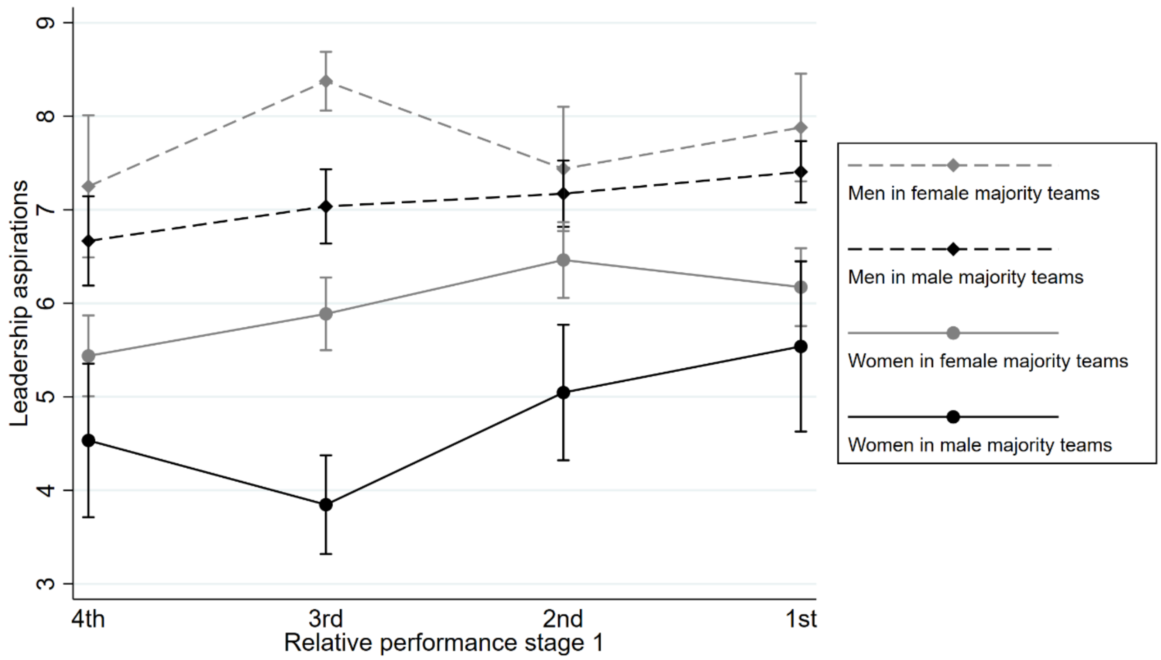


Figure 5: Willingness to lead by relative performance, gender and treatment

Note: Error bars represent standard errors.

To sum up, men are, on average, more willing to become the team leader than women. Both women and men are more inclined to aspire for leadership when assigned to female majority teams, although this effect is only statistically significant for women. Further, these differences are not driven by a gender gap in relative performance. In the next section we explore which mechanisms may drive these effects.

4. What drives the gender gap in leadership aspirations?

Why do women shy away from leading their teams, and why are women particularly averse to becoming the leader of male majority teams? Below, we discuss the six potential mediating factors that our experiment was designed to capture: (i) confidence in own ability to perform the task, (ii) actual electoral outcomes, (iii) anticipated electoral outcomes, (iv) influence in the team discussion, (v) performance in the second task, and (vi) gender stereotypes. We first explore whether there is a general gender gap in each of the mediating factors by asking, for example, whether women are less confident than men. We then turn to whether women in male and female dominated teams differ with respect to the mediating factors, asking, for example, whether women in male majority teams are less confident than women in female majority teams. Finally, we look at whether potential gender and treatment differences in the mediating factors can help explain the general gender gap in leadership aspirations as well as the impact of team gender composition on women's leadership aspirations.[16]

4.1 Measures of mediating factors

Based on previous literature, we identified six mediating factors that we judged relevant for the gender gap in leadership aspirations, and which we could measure in our design.

(i) Confidence. A number of previous studies indicate that women are often less confident in their own ability than men. A gender gap in confidence may account for at least part of the observed gender gap in leadership aspirations. Moreover, if women believe that men perform better at the task, or if other aspects of male dominated environments influence women´s confidence negatively, women in male majority teams may be even less confident in their relative ability than women in female majority teams.

Our design comprises two measures of confidence. Both measures were incentivized and elicited directly after the team discussion:

a. Guess of relative performance: In Stage 4, we asked all participants to guess how well they did in the first, individual, Lost at Sea task relative to the other members of their team. The resulting variable is categorical and ranges from 1 (worst in the team) to 4 (best in the team).

b. Updating: In Stage 3, we gave all participants the opportunity to update their individually submitted answers from Stage 1, allowing us to measure how much each participant altered her initial responses relative to the other team members after the team discussion. Participants who are less confident in their own ability may be more likely to be convinced by others in the team discussion and, thus, update their individual answers after the team discussion to a larger extent than more confident participants. To obtain a measure of updating, we computed, for each item in the Lost at Sea task, the absolute difference in rank between the participant's original answer and their updated answer. Then, for each participant, we summed these absolute differences over all ten items. Finally, to get a measure of the participant's degree of updating relative to the other team members, we divided each participant´s sum of differences by the sum of all team members' differences. The resulting variable is continuous and ranges between 0 (the participant did not update at all) and 1 (the participant was the only one in their team updating).

[16] While we present some results also pertaining to the male participants, in this section we focus on the mechanisms behind the impact of team gender composition on female leadership aspirations. As outlined in our pre-registered pre-analysis plan, this is the focus of our study.

(ii) Actual electoral outcomes. Another reason for the gender gap in leadership aspirations may be that women receive less support from their team members than men do. Previously cited research also indicates that it may be harder for women, and that women may believe it to be harder, to lead teams consisting predominantly of men. We approximate the support a participant receives form their team members by calculating their rank in the election. In the election, each participant was asked to rank the other three team members in terms of who should win. The most preferred team member was ranked first (3 points), the second most preferred was ranked second (2 points), and the least preferred was ranked third (1 point). To determine each participant's rank in the election, we compare the team members' summed ranking points. The resulting variable ranges between 1 (the participant was ranked last) and 4 (the participant was ranked first). In case of ties, the participants with equal points are assigned the average rank.[17]

(iii) Anticipated electoral outcomes. It is also possible that women are less confident than men about the support they will receive from their team members. Thus, in Stage 7, before the results of the election were revealed, we asked each participant to guess how well they would do in the election. The variable indicating the participant's guess is categorical, taking a value between 1 (the participant thought that they would come last in the election) and 4 (the participant thought that they would win the election).

(iv) Influence. A further potential mediating factor may be gender differences in influence. If women are less influential in team discussions in general, and in male majority teams in particular, this may discourage them from trying to become the team leader.

To assess each team member's relative influence in the team discussion, we computed the absolute difference in rank between the participant's individual answer from Stage 1 and their team's joint answer from Stage 2, for each item in the Lost at Sea task. Then, we sum these absolute differences over all ten items, and divide the participant´s sum by the total sum of all team members. Finally, to facilitate interpretation, we reversed the scale of the variable. Thus, the final variable indicating influence will be continuous, taking on values between 0 (the participant had no influence in the team's solution) and 1 (the team's solution was identical to the participant's individual solution).

(v) Performance in the second task. Previous literature indicates that, at least sometimes, male majority environments seem to have a negative impact on women´s

[17] In the actual election, ties were broken randomly.

performance. We therefore also include individual performance in the second task (i.e., the Desert Survival task of Stage 8) as a possible mediating variable.

(vi) Gender stereotypes. We also elicited participants' gender-leader associations through the IAT test, administered at the end of the study, before the final questionnaire. Our thought was that the average IAT score of the other team members may inform us about the general team climate with respect to gender stereotypes and leadership that participants faced during the team discussions. Our primary IAT measure is the average IAT score of the participant's team members (excluding the participant's own score). The IAT score is measured on a scale between -2 and +2, where a positive (negative) score indicates that the team members are biased in the sense that they find it easier (more difficult) to associate men with leadership than women.

4.2 Gender gaps in the mediating measures

To explore whether there is a general gender gap in the mediating factors, Table 3 presents regressions with each mediating measure as dependent variable, controlling for participant gender and relative performance. The first two columns explore whether there is a gender gap in confidence among our participants. In line with previous studies, we find that women are, on average, less confident than men. The first column indicates that, controlling for their actual relative performance, women believed that they were ranked, on average, 0.44 positions lower than men ($p<0.001$). Similarly, the second column indicates that, controlling for actual relative performance, women were more prone than men to change their individual answers following the team discussion. Compared to their team members, women updated their answers 4.3 percentage points more than men ($p<0.001$).

Columns 3 and 4 explore whether women´s actual and expected electoral outcomes are different from those of men. The third column indicates that women were ranked, on average, 0.42 positions worse than similarly performing men in the election ($p<0.001$). In line with this finding, the fourth column indicates that women believed that they would be ranked, on average, 0.58 positions worse than men with a similar relative performance did ($p<0.001$).

Table 3: Overall gender gaps

Outcome variable:	Guess rank Stage 1	Updating Stage 3	Rank election	Guess rank election	Influence	Performance Stage 8	IAT scores of team members
	1=worst 4=best	*0=minimum 1=maximum*	*1=lowest 4=highest*	*1=lowest 4=highest*	*0=minimum 1=maximum*	*0=worst 50=best*	*-2=female bias 2=male bias*
Female	-0.441***	0.043***	-0.424***	-0.578***	-0.029***	-1.005	-0.014
	(0.066)	(0.008)	(0.088)	(0.067)	(0.007)	(0.589)	(0.016)
Constant	3.304***	0.133***	3.027***	3.227***	0.846***	23.396***	0.216***
	(0.087)	(0.011)	(0.114)	(0.090)	(0.009)	(0.791)	(0.019)
N	580	580	580	580	580	580	577

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Note: Standard errors are clustered on the team level. All regressions are OLS regressions including controls for relative performance in Stage 1 (1=best, 4=worst). ***'Guess rank Stage 1'*** denotes the participant's guess (from Stage 4) of their within-team ranking in terms of performance in Stage 1 (1=worst, 4=best). ***'Updating Stage 3'*** indicates how much the participant updated their individual answer from Stage 1 in Stage 3, relative to the other team members (0=the participant did not update, 1=the participant was the only one in the team who updated). ***'Rank election'*** denotes the participant's rank in the election based on all votes (1=last, 4=first, equal observations are assigned the average rank). ***'Guess rank election'*** denotes the participant's guess of their rank in the election (1=last, 4=first). ***'Influence'*** denotes how close the team answer in Stage 2 was to the participant's individual answer from Stage 1, relative to the other team members (0-1, where 1=the team solution was identical to the participant's solution). ***'Performance Stage 8'*** indicates the participant's individual performance in the desert survival task in Stage 8 (0=worst possible, 50=best possible)**.**

The coefficient of *Female* in Column 5 indicates that, controlling for relative performance, women are less influential than men in the team discussion ($p<0.001$). The second to last column illustrates that the gender difference in individual performance in the second task is insignificant.[18] Finally, the last column shows that women and men are exposed to team members with, on average, similar implicit gender-leadership associations ($p=0.398$).[19] To sum up, controlling for relative performance, we find evidence in support of gender gaps in all mediating factors explored in our study, apart from performance in the second task and the IAT scores of the team members. Most important appears to be that women are less confident about their relative performance and are more pessimistic about their endorsement in the election. The next section explores whether our mediating factors differ between women in female majority teams and male majority teams.

4.3 Do male majority environments impact women negatively?

In addition to the baseline gender gaps in our mediating variables, illustrated in Table 3, it is possible that women in male majority environments face different challenges than women in female majority environments with respect to, for example, confidence, support, or influence. In Table 4 we regress each of the potential mediating factors on variables indicating participant gender, team gender composition, an interaction between the two, and relative performance.[20]

Consistent with our hypotheses, women in male majority teams have a more negative perception of their relative performance than women in female majority teams. The first column indicates that, controlling for *actual* relative performance, women in male majority teams *believed* that they were ranked, on average, 0.37 positions worse than women in female majority teams did ($p<0.001$) as indicated by the sum of the coefficients of *Female team* and *Female X Female team*. Women in male majority teams were also more prone to update their individual answers following the team discussion than women in female majority teams ($p=0.006$). As with the gender gap in leadership ambition, it is notable that the impact of team gender composition on female confidence appears relevant, and while somewhat smaller, is comparable in magnitude to the average gap between men and women in the whole sample.

---

[18] The regression reported in Column 5 from Table 3 controls for relative performance in the first task. The raw gender gap (including no controls) in performance in the second task is 1.19 ($p=0.055$).

[19] When using the participant's own IAT score as outcome variable in the same regression, we find that women express less of an implicit association between maleness and leadership than men do (difference=0.092, $p<0.001$).

[20] In Table 4, as in column 4 of Table 2, we also interact the measure of relative performance with *Female*. Thus, we allow for gender-specific effects of relative performance on the outcome variable.

Table 4: Effects of gender and team gender composition

Outcome variable:	Guess rank Stage 1 *1=worst* *4=best*	Updating Stage 3 *0=minimum* *1=maximum*	Rank election *1=lowest* *4=highest*	Guess rank election *1=lowest* *4=highest*	Influence *0=minimum* *1=maximum*	Performance Stage 8 *0=worst* *50=best*	IAT scores of team members *-2=female bias* *2=male bias*
Female	-0.658***	0.051*	-0.292	-0.591**	-0.034	-1.331	0.006
	(0.180)	(0.022)	(0.257)	(0.216)	(0.017)	(1.624)	(0.036)
Female majority team	0.221	-0.027*	0.346**	0.463***	0.016	0.054	-0.074*
	(0.113)	(0.012)	(0.131)	(0.113)	(0.010)	(1.046)	(0.028)
Female X Female majority team	0.150	-0.005	-0.120	-0.201	0.007	-0.387	-0.014
	(0.161)	(0.021)	(0.231)	(0.169)	(0.018)	(1.398)	(0.022)
Constant	3.239***	0.145***	2.837***	3.072***	0.838***	28.652***	0.248***
	(0.121)	(0.015)	(0.150)	(0.126)	(0.012)	(1.052)	(0.024)
N	580	580	580	580	580	580	580.000
F-test: 'Female maj. team' + 'Female X Female maj. team'	0.371*** (F=14.445)	-0.032** (F=7.937)	0.227 (F=3.100)	0.262* (F=5.549)	0.023* (F=6.100)	-0.333 (F=0.114)	-0.088** (F=8.587)

* p<0.05; ** p<0.01; *** p<0.001

Note: Standard errors are clustered on the team level. All regressions are OLS regressions including controls for relative performance in Stage 1 (1=best, 4=worst), and relative performance in Stage 1 interacted with *Female*. The final row presents results from an F-test, testing the treatment effect for women. ***'Guess rank Stage 1'*** denotes the participant's guess (from Stage 4) of their within-team ranking in terms of performance in Stage 1 (1=worst, 4=best). ***'Updating Stage 3'*** indicates how much the participant updated their individual answer from Stage 1 in Stage 3, relative to the other team members (0=the participant did not update, 1=the participant was the only one in the team who updated). ***'Rank election'*** denotes the participant's rank in the election based on all votes (1=last, 4=first, equal observations are assigned the average rank). ***'Guess rank election'*** denotes the participant's guess of their rank in the election (1=last, 4=first). ***'Influence'*** denotes how close the team answer in Stage 2 was to the participant's individual answer from Stage 1, relative to the other team members (0-1, where 1=the team solution was identical to the participant's solution). ***'Performance Stage 8'*** indicates the participant's individual performance in the desert survival task in Stage 8 (0=worst possible, 50=best possible)**.**

Figure 6 presents the mediating measures separated by gender, treatment and actual performance ranking. Panels (a) and (b) present participants' relative performance beliefs and their relative updating, respectively. Consistent with the picture that emerged with respect to willingness to lead, men in female majority teams are, in point estimates, the most confident in their ability, followed by men in male majority teams, and women in female majority teams. Within all levels of relative performance, women in male majority teams are the least confident.

In addition to the impact on confidence described above, we find some support that team gender composition impacts women´s actual, and expected, electoral outcomes. The third column of Table 4 suggests that, controlling for relative performance, women in male majority teams are ranked, on average, 0.23 positions worse by their team members than women in female majority teams. However, this difference fails to reach statistical significance at conventional levels (p=0.080). The fourth column indicates that, on average, women in male majority teams also believe that they will be ranked 0.26 positions worse in the election than do similarly performing women in female majority teams (p=0.020).

Panels (c) and (d) in Figure 6 illustrate the actual and expected election outcomes separated by gender, treatment and relative performance. The patterns that arise are very similar to those presented for confidence (Panels (a) and (b)) and leadership aspirations (Figure 5). For each level of relative performance, men in female majority teams receive, and expect to receive, the most votes, followed by men in male majority teams. With one minor exception, women in male majority teams receive, and expect to receive, the least votes. Noticeably, also with one minor exception, men receive, and expect to receive, more votes than women regardless of treatment or relative performance.

The impact of team gender composition on women´s influence further reinforces the pattern presented above. Column 5 of Table 4 indicates that, controlling for relative performance, male majority teams appear to generate team answers farther from the individual answers of female team members than do female majority teams (p=0.015). The overall pattern illustrated in Panel (e) in Figure 6 also suggests that men in female majority teams are the most influential, followed by men in male majority teams, and women in female majority teams. Least influential are women in male majority teams.

The last two columns of Table 4 explore whether women in male majority teams perform worse at the second task, or face team members who associate maleness with leadership to a larger extent, than women in female majority teams do. While we find little evidence that women in male majority teams perform worse (p=0.736), we find that women in male majority teams face teams with a stronger male-leadership stereotype (p=0.004).

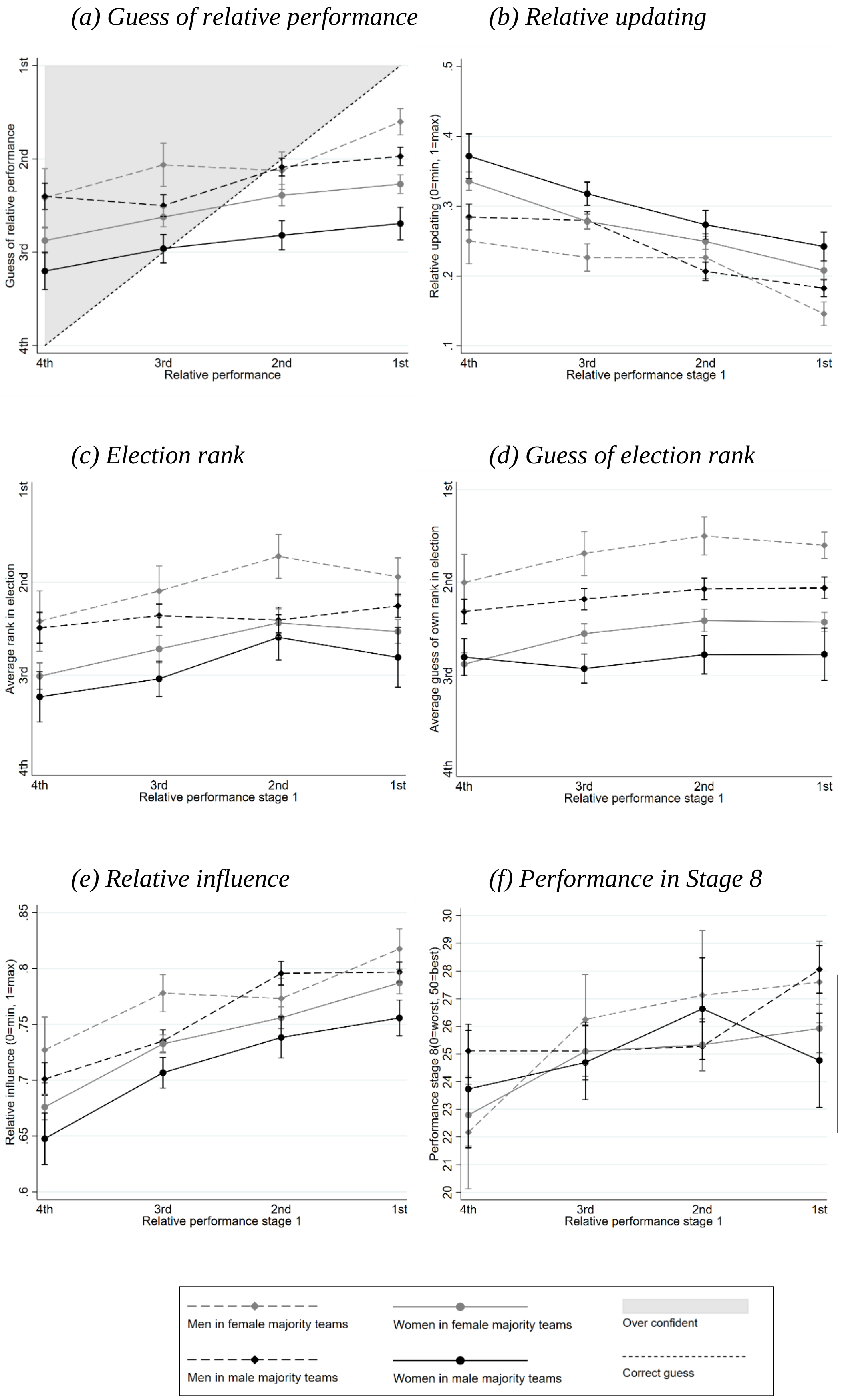


Figure 6: Mediating measures, by relative performance, gender and treatment

Note: Error bars represent standard errors.

To sum up, we find evidence in support of male majority teams having negative consequences for women. Compared to women in female majority teams, women in male majority teams are less confident in their own ability, accurately expect worse electoral outcomes, have less influence in the team decision making, and experience more gender bias.

4.3 Do the mediating measures drive gender gaps in leadership aspiration?

While our study was not primarily powered to explore whether the mediating factors can account for the gender gap in leadership aspiration, or the treatment effect, it is still interesting to report this analysis to inform future studies..

To asses to what extent our mediating variables can explain part of the overall gender gap in willingness to lead, we use the regression from column 2 of Table 2 as a starting point (regressing leadership aspirations on gender and relative performance). Then, we add our mediating variables one by one as controls, and measure how much the estimated gender gap in leadership aspirations (i.e., the coefficient of *Female*) decreases. The dotted line in panel (a) of Figure 7 illustrates the estimated gender gap without controlling for any mediating variables (coefficient=1.584, s.e.=0.262). The black dots illustrate the estimated gender gap after controlling for the mediating variable indicated on the x-axis.[21]

The estimated gender gap in leadership aspiration decreases when controlling for each of the mediating measures related to confidence, electoral outcomes, and influence. The decrease is statistically significant in two cases. The introduction of a control for relative performance beliefs in the regression implies a reduction of the estimated gender gap in leadership aspirations by almost half, from 1.58 to 0.87. This decrease is statistically significant at the five percent level (p=0.046, Z=2.00). Further, controlling for participants' *expectations* about how many votes they will receive, the estimated gender gap in leadership aspirations decreases significantly, by 64%, from 1.58 to 0.56 (p=0.004, Z=2.92). Thus, beliefs about relative performance, our most direct measure of confidence, and beliefs about electoral outcomes seem to explain a sizeable part of the gender gap in leadership aspirations among our participants.

[21] The regressions underlying panel (a) of Figure 7 are presented in Table A1 in the Appendix.

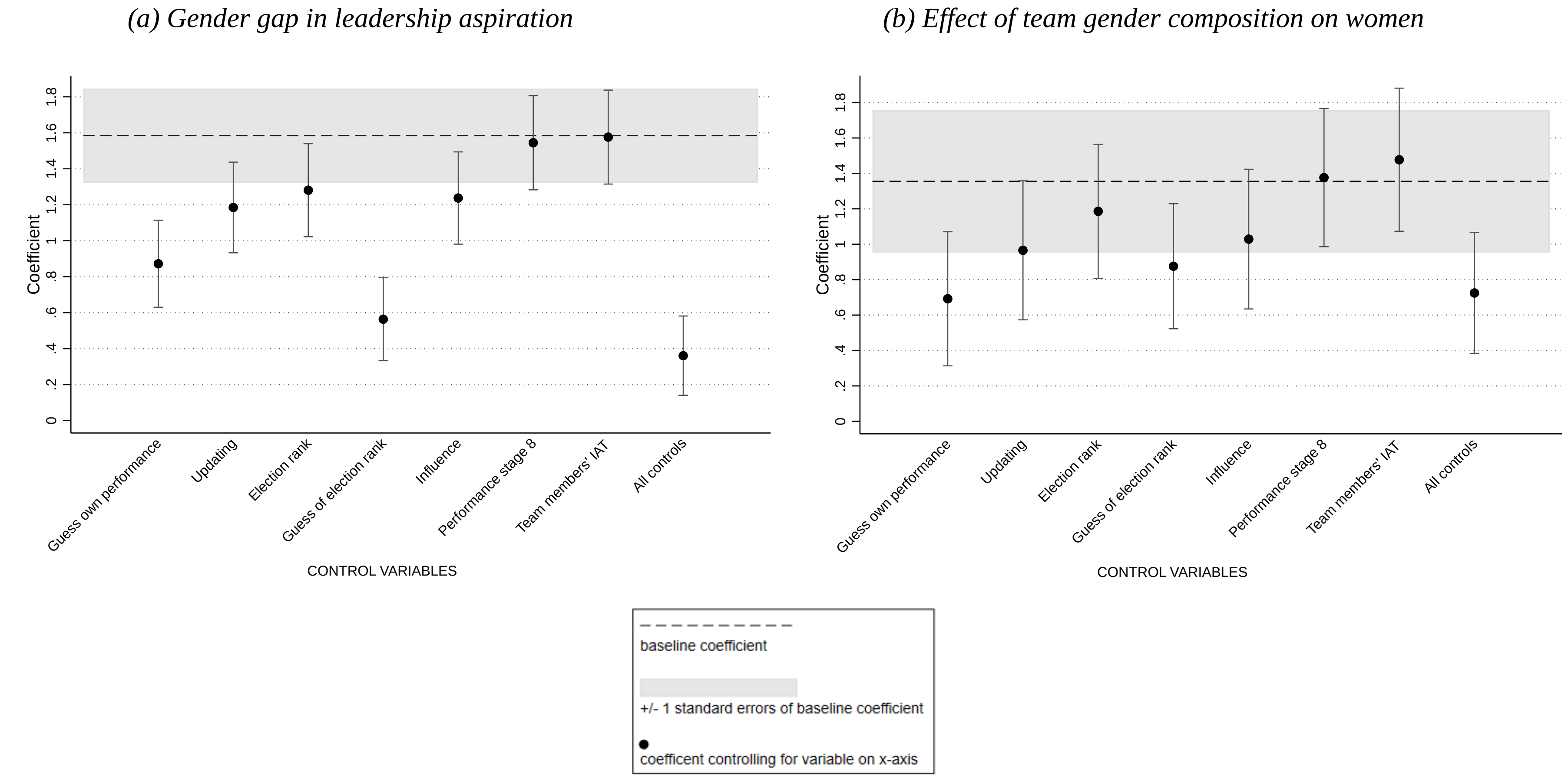


Figure 7: Controlling for mediating variables

Note: The dotted line in panel (a) shows the coefficient of *Female* from column 2 of Table 1. The dotted line in panel (b) shows the sum of the coefficients of *Female majority team* and *Female X Female majority team* from column 4 of Table 1. The shaded area represents +/- 1 standard errors of these coefficients. The black dots show the same coefficients after controlling for the variable indicated on the x-axis (in panel (b), we also control for the variable indicated on the x-axis interacted with *Female*). The error bars represent +/- 1 standard errors of these coefficients. The regressions underlying panels (a) and (b) are presented in Tables A1 and A2 in the Appendix.

Panel (b) in Figure 7 is based on the regression from column 4 of Table 2 (regressing leadership aspirations on gender, team gender composition, an interaction between the two, and relative performance), and illustrates how team gender composition affects women´s leadership aspiration through each of the mediating measures. The dotted line shows the estimated impact of team gender composition on women's leadership aspirations (i.e., the sum of the coefficients of *Female majority team* and *Female X Female majority team*) without controlling for any mediating variables (1.355, s.e.=0.402). The black dots illustrate how the estimated impact of team gender composition changes after controlling for each of the mediating variable indicated on the x-axis.[22]

As for the general gender gap in leadership aspiration, the impact of team gender composition for women´s leadership aspirations decreases when controlling for most of the mediation measures. However, while some of these changes appear sizeable, they are imprecisely measured and are not statistically significant.[23]

4. Outcomes

Although not the primary aim of our study, in this section we briefly evaluate whether male and female majority teams elect different types of leaders, and how such differences may impact the teams' outcomes. As illustrated in previous sections, men receive more votes, and have a higher average willingness to become the team leader, than women. This tendency cannot be accounted for by individual task performance and holds for both male and female majority teams. In addition, both men and women are more willing to become the team leader in female majority teams than in male majority teams, and – controlling for relative ability – men in female majority teams receive more votes than any other sub-team ($p<0.001$).[24] As a result, despite being more willing to become the leader, women in female majority teams do not become leaders more frequently than women in male majority teams.

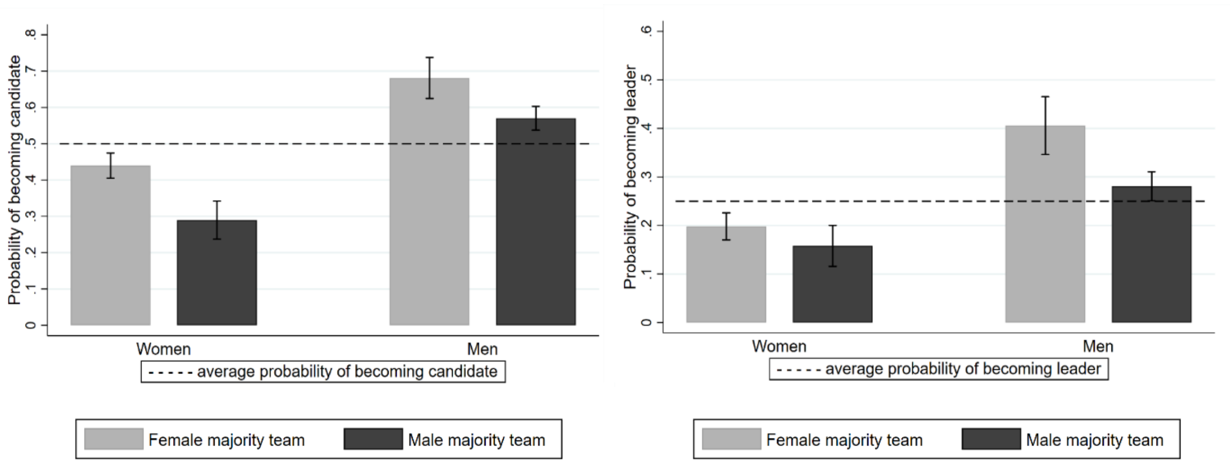
---

[22] The regressions underlying panel (b) of Figure 8 are presented in Table A2 in the Appendix. In each regression, we interact the measure of relative performance, and each control variable, with *Female*. Thus, we allow for gender-specific effects of these variables on the outcome variable.

[23] Controlling for participants' guess about their own relative performance, the effect of team gender composition on women's leadership aspirations (i.e., the sum of the coefficients of *Female majority team* and *Female X Female majority team* in the regression presented in Table A2) decreases by 49%, from 1.36 to 0.69 (p=0.230, Z=1.20). The corresponding decrease controlling for updating is 29% (p=0.490, Z=0.69), for rank in the election 12% (p=0.757, Z=0.31), for participants' electoral expectations 35% (p=0.368, Z=0.90), and for influence 24%, from 1.36 to 1.03, (p=0.562, Z=0.58).

[24] This p-value is based on OLS regressions with the average votes a participant received as the dependent variable, controlling for relative performance in Stage 1, and clustering standard errors at the team level. The p-value is less than 0.001 regardless of whether we compare men in female majority teams to (1) everyone else, (2) men in male majority teams, (3) women in female majority teams, or (4) women in male majority teams.

Figure 8 shows the average, individual, probability of becoming a candidate (Panel a), and of becoming the leader (Panel b), by gender and treatment (see Table A3 for regression analyses). In both types of teams, women's probability of becoming candidate and leader is lower than average, while men's is higher than average. In female majority teams, a man is 1.6 times more likely than a woman to become a candidate (68.1% vs. 43.9%), and 2.1 times more likely to become the leader (40.6% vs. 19.8%). In male majority teams, a man is 2.0 times more likely than a woman to become a candidate (57.0% vs. 28.9%), and 1.8 times more likely to become the leader (28.1% vs. 15.8%).[25] As illustrated in Table A3, these numbers do not change importantly if we control for relative performance.



(a) Probability of becoming candidate *(b) Probability of becoming leader*

Figure 8: Probability of becoming candidate and leader, by gender and treatment
Note: Error bars represent standard errors.

In terms of electing the most able leader, based on individual performance in the first task, only 32% of the teams manage to do so. While male and female teams are fairly similar in this regard, we see a small tendency of female majority teams to elect more able leaders, and, in particular, to elect fewer leaders among the least capable team members (see Figure A4).

[25] Although the number of participants in each cell becomes small we can also explore the probability of becoming the team leader given gender, team composition and relative performance. Figure A5 indicates a similar pattern as previous figures with men in female majority teams having the highest probability of becoming the team leader, in point estimates, for all levels of relative performance except when performing second best. On average, men in male majority teams fare second best, followed by women in female majority teams and women in male majority teams. Overall, women tend to have a probability of becoming the leader below the average of 25%, and men above. Interestingly, the best performing women in male majority teams have a very small chance of becoming the leader, but this comparison is based on very few observations.

5. Conclusion

In this study we use an economic experiment to identify the causal effect of team gender composition on women´s leadership aspirations through random assignment of participants to either male or female majority teams. Consistent with previous studies on gender gaps in leadership aspiration, we find that, compared to men, female participants are significantly less willing to become the team leader. This gender gap is not driven by gender differences in ability to perform the task, and remains regardless of relative performance and team gender composition. In line with our primary hypothesis, we also find that female participants are significantly less willing to become the leader of male majority teams than of female majority teams. A substantial share of the effect of team gender composition on women's leadership aspirations is explained by relative performance beliefs and beliefs about other team members' endorsement in the election of a leader. Women in female majority teams are more confident in their own task performance relative to the other team members, and believe that they will receive more votes in the subsequent election, than women in male majority teams.

Despite higher leadership ambition, women assigned to female majority teams are not more likely to become the team leader than women assigned to male majority teams. This is because men are also more willing to become the leader in female majority teams (albeit not significantly so), and – regardless of their relative performance – men receive a large share of the votes in female majority teams. As a result, men in female majority teams fare better than all other groups in many regards, women in male majority teams fare the worst. Thus, since numerical minority has the opposite effect on men compared to women, the increased female leadership ambitions in female majority teams does not translate into a larger share of women in leading positions. Still, we believe it is noteworthy that – after only ten minutes of team interaction – women randomly assigned to female majority teams are more confident and willing to take on a leadership role. This difference in attitudes is likely to correlate with other types of behaviors and feelings that may matter for women´s outcomes and gender equality in the long run. To explore the impact of team gender composition over a longer period of time, and what mechanisms may be at play, is an interesting avenue for future research.

Appendix A

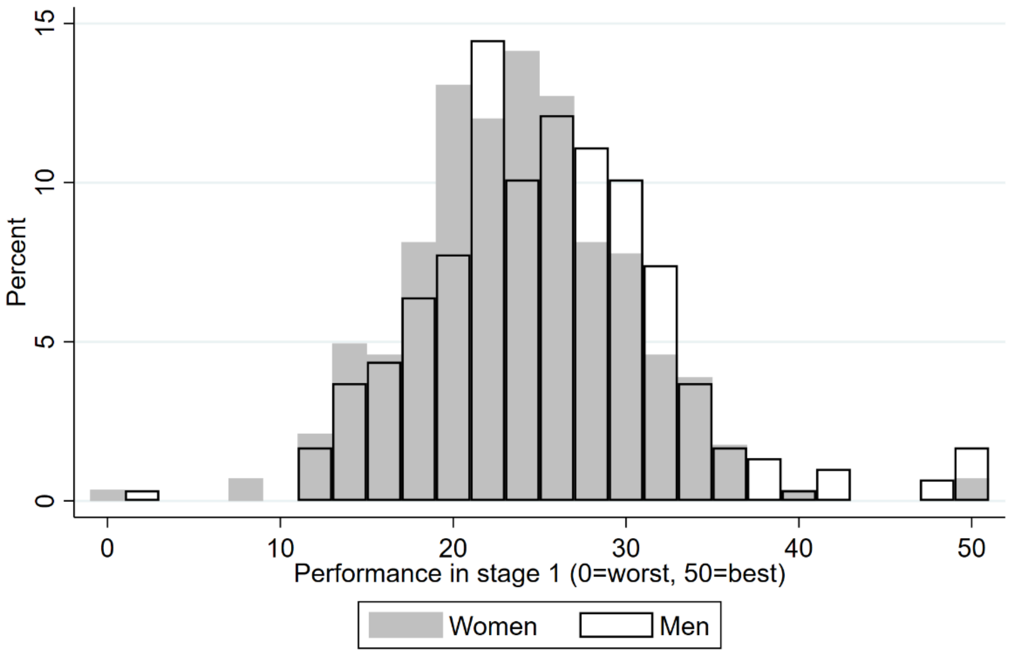


Figure A1: Distribution of male and female performance in Stage 1

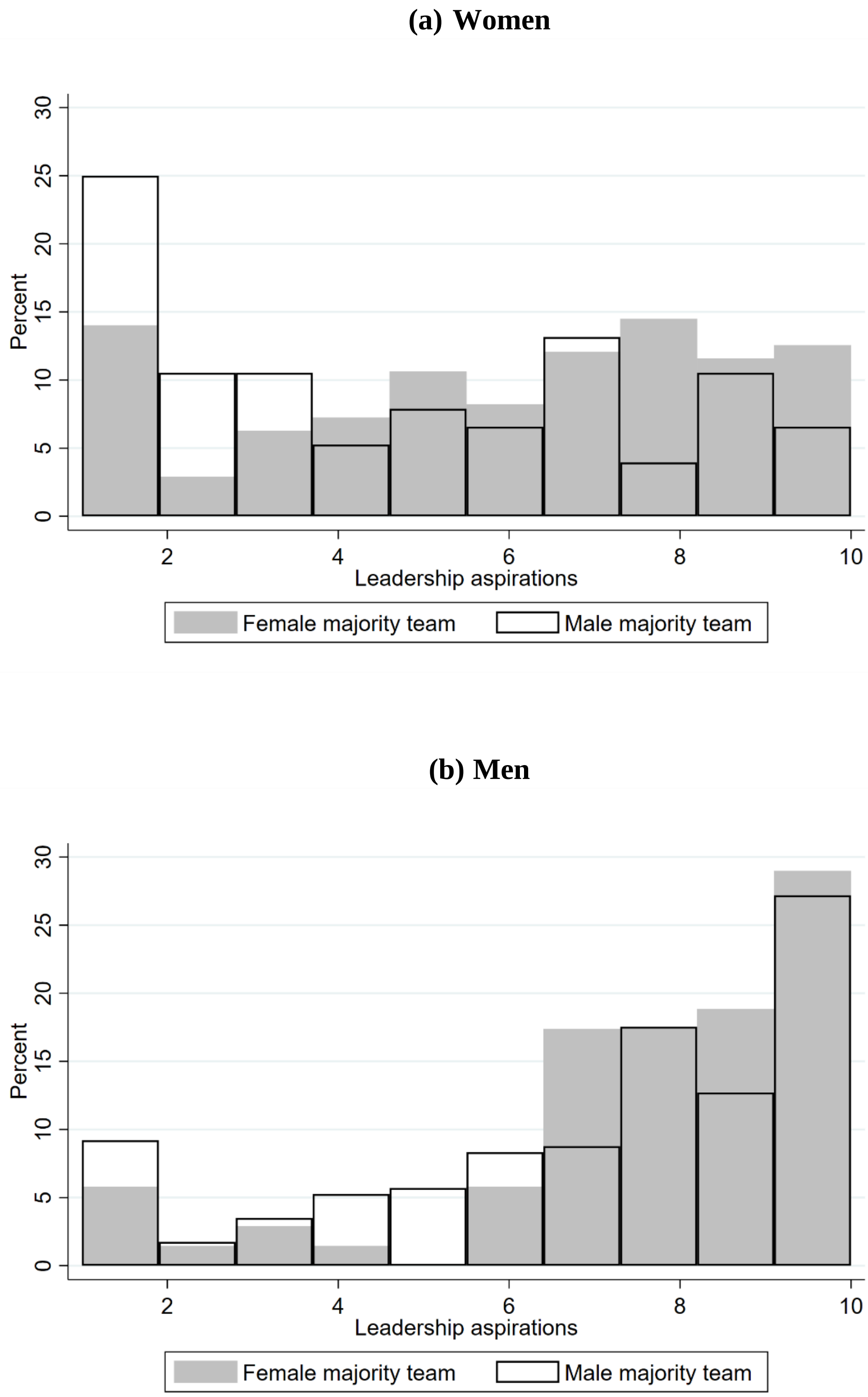


Figure A2: Distribution of leadership aspiration by gender and treatment

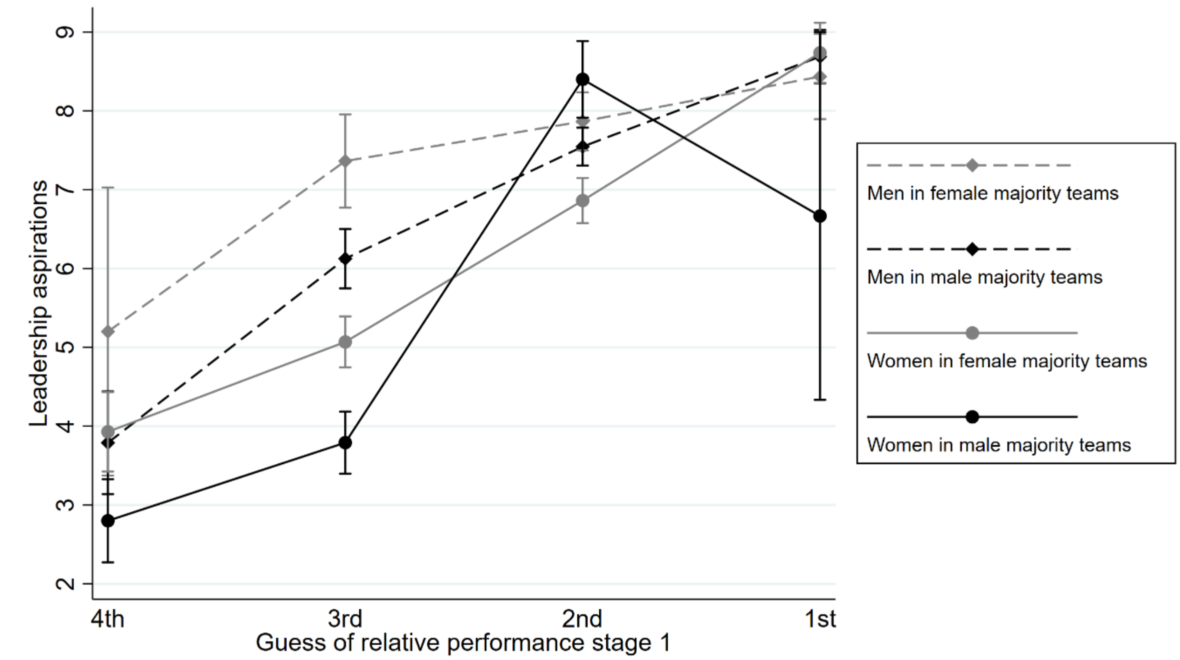
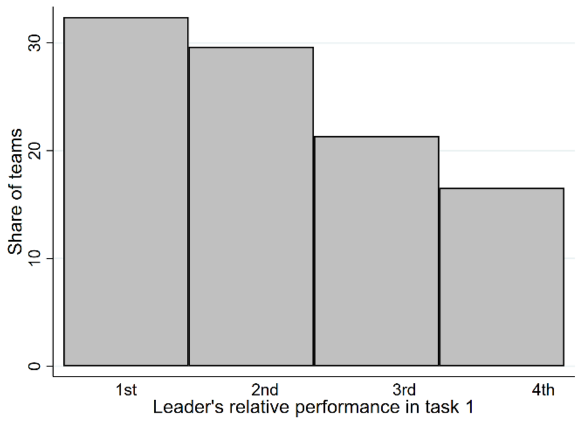


Figure A3: Leadership aspirations by participants' guess of their own relative performance

(a) Overall



(b) By gender composition of team

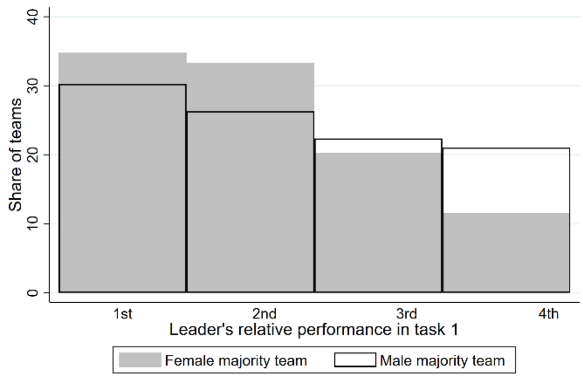


Figure A4: Leader’s relative performance in task 1

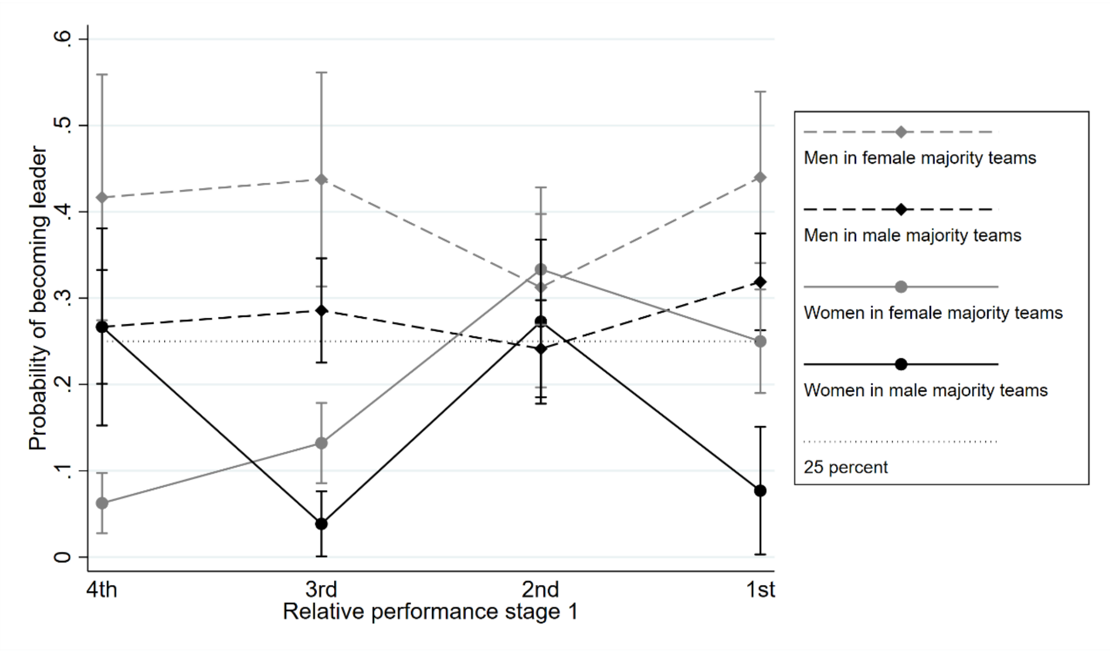


Figure A5: Probability of becoming team leader, by relative performance in Stage 1

Table A1: Changes in estimated gender difference after adding control variables

Control variable	Coefficient (overall gender gap)	Change compared to baseline
Baseline (no controls)	**-1.584*** (0.262)**	
Guess rank Stage 1	-0.872*** (0.242)	-45%* (z=2.00)
Updating Stage 3	-1.185*** (0.252)	-25% (z=1.10)
Rank election	-1.281*** (0.259)	-19% (z=0.82)
Guess rank election	-0.564* (0.231)	-64%** (z=2.92)
Influence	-1.237*** (0.256)	-22% (z=0.95)
Performance Stage 8	-1.545*** (0.262)	-2.5% (z=0.11)
IAT score of other team members	-1.576*** (0.261)	+0.5% (z=0.02)
All control variables	-0.361 (0.220)	-77%*** (z=3.57)

Notes. Each reported coefficient is the coefficient of *Female* in a separate OLS regression using the following specification: ***Leadership aspiration* = β_0 + β_1 *Female* + β_2 *Relative Performance* + β_3 *Control variable***. The bold coefficient in the first row is obtained using no control variables, and the coefficients reported in the subsequent rows are obtained controlling for the variable indicated in the first column. Standard errors are clustered on the team level. The reported change in the third column indicates the change in the coefficient of Female compared to the specification using no control variables.

Table A2: Changes in estimated treatment effect after adding control variables

	(a) WOMEN		(b) MEN	
Control variable	**Coefficient (treatment effect)**	**Change compared to baseline**	**Coefficient (treatment effect)**	**Change compared to baseline**
Baseline (no controls)	**1.355*** (0.402)**		**0.650 (0.353)**	
Guess rank Stage 1	0.692 (0.379)	-49% (z=1.20)	0.347 (0.341)	-47% (z=0.62)
Updating Stage 3	0.965 (0.392)	-29% (z=0.69)	0.480 (0.346)	-26% (z=0.34)
Rank election	1.186 (0.379)	-12% (z=0.31)	0.440 (0.350)	-32% (z=0.42)
Guess rank election	0.876 (0.353)	-35% (z=0.90)	-0.122 (0.338)	-119% (z=1.59)
Influence	1.029 (0.394)	-24% (z=0.58)	0.496 (0.352)	-24% (z=0.31)
Performance Stage 8	1.376 (0.390)	+1.5% (z=0.0)	0.649 (0.354)	-0.0% (z=0.00)
IAT score of other team members	1.477 (0.404)	+9.0% (z=0.21)	0.698* (0.338)	+7.4% (z=0.10)
All control variables	0.725 (0.342)	-46% (z=1.19)	-0.137 (0.319)	-121% (z=1.65)

Notes. All reported coefficients are obtained in separate OLS regressions using the following specification: ***Leadership aspiration*** **=** β_0 **+** β_1 ***Female*** **+** β_2 ***Team Female*** **+** β_3 **(*Female X Team Female*) +** β_4 ***Relative Performance*** **+** β_5 **(*Female X Relative performance*) +** β_6 ***Control variable*** **+** β_7 ***(Female X Control variable)***. Panel (a) reports the sum of the coefficients of *Team Female* and *Female X Team Female* (standard errors are obtained using STATAs 'lincom' command), while panel (b) reports the coefficient of *Team Female*. The bold coefficients in the first row are obtained using no control variables, and the coefficients reported in the subsequent rows are obtained controlling for the variable indicated in the first column. Standard errors are clustered on the team level. The changes reported in the third and fifth columns indicate the changes in the coefficients of interest compared to the specifications using no control variables.

Table A3: Impact of gender and team gender composition on likelihood of becoming candidate and leader

(a) Gender gaps in likelihood of becoming candidate and leader

Dependent variable:	*Indicator for if the participant became a candidate*				*Indicator for if the participant became the leader*			
	Female majority teams		Male majority teams		Female majority teams		Male majority teams	
Female	-0.242**	-0.231**	-0.281***	-0.274***	-0.208*	-0.193*	-0.123*	-0.121*
	(0.075)	(0.077)	(0.070)	(0.072)	(0.080)	(0.082)	(0.056)	(0.056)
Constant	0.681***	0.772***	0.570***	0.639***	0.406***	0.533***	0.281***	0.297***
	(0.057)	(0.087)	(0.017)	(0.069)	(0.060)	(0.068)	(0.014)	(0.068)
N	276	276	304	304	276	276	304	304
Controls:								
Relative performance Stage 1		YES		YES		YES		YES

(b) Impact of team gender composition on likelihood of becoming candidate and leader

Dependent variable:	*Indicator for if the participant became a candidate*				*Indicator for if the participant became the leader*			
	Women		Men		Women		Men	
Female majority team	0.150**	0.145**	0.111	0.108	0.040	0.035	0.125*	0.124*
	(0.056)	(0.056)	(0.059)	(0.059)	(0.046)	(0.047)	(0.061)	(0.061)
Constant	0.289***	0.416***	0.570***	0.622***	0.158***	0.299***	0.281***	0.304***
	(0.052)	(0.094)	(0.017)	(0.062)	(0.042)	(0.070)	(0.014)	(0.068)
N	283	283	297	297	283	283	297	297
Controls:								
Relative performance Stage 1		YES		YES		YES		YES

* $p<0.05$; ** $p<0.01$; *** $p<0.001$
Note: Standard errors are clustered on the team level