

Working Paper in Economics No. 744

A man's world? – The impact of a male dominated environment on female leadership

Andreas Born, Eva Ranehill, Anna Sandberg

Department of Economics, November 2018

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

A man's world? – The impact of a male dominated environment on female leadership*

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

November, 2018

Despite the significant growth in female labor force participation and educational attainment over the past decades, few women reach leadership positions. In this study, we explore whether male dominated environments, in and of themselves, adversely affect women's willingness to lead a team. We find that women randomly assigned to male majority teams are less willing to become team leaders than women assigned to female majority teams. Analyses of potential mechanisms show that women in male majority teams are less confident in their relative performance, less influential, and more swayed by others in team discussions. They also (accurately) believe that they will receive less support from team members in a leadership election. Taken together, our results indicate that the absence of women in male dominated contexts may be a self-reinforcing process.

Keywords: leadership; gender differences; experiment

JEL codes: C92, J16

* We are grateful for helpful comments from Ingvild Almås, Fredrik Carlsson, Jeffrey Carpenter, Anna Dreber, Tore Ellingsen, Karin Hederos, Randi Hjalmarsson, Magnus Johannesson, Mikael Lindahl, Erik Lindqvist, Åsa Löfgren, Johanna Rickne, David Strömberg, Joseph Vecchio, Lise Vesterlund, and Roberto Weber, as well as from seminar participants at the ESA Meetings in Berlin 2018, the University of Exeter, SOFI, the University of Gothenburg, the University of Konstanz, the University of Lund and the University of Stavanger. We thank the Royal Swedish Academy of Sciences (grant number SO2015-0036) for generous financial support.

1. Introduction

Despite important advancement in women’s labor market participation and educational attainment over the last decades, economic and political activity around the world remains characterized by high degrees of vertical and horizontal gender segregation. A large literature shows that women are persistently underrepresented in leading positions (Bertrand and Hallock 2001; Lawless and Fox, 2012; Blau and Kahn 2017; Thomas et al. 2017) and in high-paying occupations (Bettio and Verashchagina 2009; Blau, Brummund, and Liu 2013; Cohen 2013; Gobillon, Meurs, and Roux 2015; Olivetti and Petrongolo 2016).¹ As a result, conventional measures of human capital today explain only a small part of the gender wage gap, while occupation and industry remain important factors (Blau and Kahn 2017). This raises the question of why—despite the fact that female educational attainment now surpasses that of men in many countries—so few women reach the top, and instead persistently pursue educations and career trajectories with less career potential, in lower paid industries, and at lower paid levels within firms.

In this paper, we propose that male dominated environments may, in themselves, have an adverse impact on women’s careers. In particular, we focus on the question of why so few women reach the top, and ask if women’s willingness to take on a leadership role is negatively influenced by being surrounded by many men. Since men are more likely than women to advance to leading positions, the share of women within organizations tends to decrease further up the career ladder. Thus, if women are adversely affected by being in minority, the absence of women at the top may become a self-perpetuating cycle whereby women become reluctant to enter, and prone to leave, male dominated, high-level positions.

While the career trajectories of women in male dominated environments is a subject of increasing interest, confounding factors and data limitations make it difficult to precisely estimate the *causal* impact of minority status on women’s outcomes using observational data. Most importantly, selection (by employees and/or employers) into different industries and organizations, is far from random. Women who enter male dominated work environments may

¹ While women in the U.S. hold the majority of bachelor and master degrees in all age categories younger than 70 years, their access to leadership positions has increased only slowly over time. For example, women constitute 45% of the employees of S&P 500 firms, but hold only 20% of board seats and 5% of CEO positions (Catalyst 2018). OECD statistics reveal similar patterns in other countries (<http://www.oecd.org/mcm/documents/C-MIN-2017-7-EN.pdf>). Similarly, high levels of female political engagement are not manifested in women attaining leading political positions at similarly high rates—e.g., voter turnout among women has been higher in each presidential election in the US since 1980, but in 2018 women still held only 19.3% of the seats in the House of Representatives and 21% in the Senate.

differ from other women along a number of unobservable characteristics that correlate with behavior, productivity and well-being. In this study, we avoid this problem by using an economic experiment with random assignment of participants to teams with varying gender composition. Moreover, 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, executive committees or company boards with a female majority are the exception, and where they exist are likely to have very different characteristics than their male-majority counterparts). Finally, we can systematically and rigorously explore a number of mechanisms through which male dominated environments may impact women's behavioral strategies and outcomes. Specifically, in addition to willingness to lead, we measure whether the gender composition of teams impacts women's performance, confidence, influence, and actual and anticipated support from the other team members.

Our experiment employs a pre-registered design, hypotheses and analyses, and studies 580 participants. In the first stage of the experiment, participants were asked to solve a task individually. This provides us with a baseline measure of individual ability, in a context without a salient gender composition. Then, 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 teams sat down together for 10 minutes in a separate room to discuss the initial task face to face, and come up with a joint solution.

The key part of our experiment is the second stage. 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 state how much they wanted to become the leader. The team members then voted for their preferred leader by ranking all the other three team members. The two team members who indicated the highest willingness to become the leader became candidates in an election, and the candidate with the most votes became the team leader.² Team members' stated willingness to lead, which is our primary outcome variable, thus had a direct impact on their possibility to become the team leader.³

² Voting thus took place before the identity of the two candidates was revealed. In order to avoid strategic voting, participants were informed that only the votes from the two team members who were *not* candidates would count when determining which candidate would win the election. 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.

³ We thus capture participants' *stated* willingness to lead. Since women in our study are randomized to treatment, we expect women in male majority teams to have the same intrinsic motivation to lead as women in female majority teams. What our design focuses on is thus the impact of the gender composition of the team on women's display of leadership aspirations. Arguably, this is also the basis for any promotion decisions outside of the laboratory.

First, we document a substantial and significant gender gap in willingness to lead. 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 ($p < 0.001$, Cohen's $d = 0.56$). The modal response for men is 10, indicating the highest possible willingness to lead their team, while the modal response for women is 1, indicating the lowest possible willingness to lead.

Second, in support of our main hypothesis, we provide evidence that women are significantly less willing 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 ($p = 0.001$, Cohen's $d = 0.46$). Thus, the impact of team gender composition on women's willingness to lead is relevant in size and comparable in magnitude to the overall gender gap.

Neither the general gender gap in willingness to lead, nor the effect of team gender composition on female leadership motivation, can be explained by gender gaps in task related ability.⁴ Instead, exploratory analyses of mechanisms show that, compared to women in female majority teams, women in male majority teams *believe* they perform worse relative to the other team members. Women in male majority teams are also less influential, more swayed by the team discussion, receive fewer votes in the election, and are less optimistic about the electoral support they will receive. We show that low relative performance beliefs and low expectations of electoral support are particularly important factors discouraging women from trying to obtain a leading position in male majority teams.

Women's lower willingness to lead translates, in our setting, into 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, since two out of four team members become candidates, 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.

While not the focus of our study, we find that minority status affects men and women very differently. In contrast to women in minority, who fare the worst along almost all measured dimensions, men in minority fare very well. Men in female majority teams show the

⁴ In fact, according to our point estimates, men who are the *worst* performers in their team state, on average, a higher willingness to lead than the women who are the *best* performers.

highest willingness to lead, display the highest degree of overconfidence, are the most influential, receive the most votes, and expect this to be the case.

The leaky pipeline in male dominated areas has raised concerns ranging from fairness considerations to how group gender composition and the lack of female leaders impact performance (e.g. Apestegua, Azmat, and Iriberry 2011; Hoogendoorn, Oosterbeek, and van Praag 2013) and collective outcomes (e.g. Chattopadhyay and Duflo 2004; Matsa and Miller 2013; Eckel and Füllbrunn 2015; Ranehill and Weber 2017). Our study contributes to this literature by systematically exploring several specific features of male dominated environments that may impede women’s career motivations and contribute to the low retention and promotion rates observed for women (Marschke et al. 2007; Thomas, Poole, and Herbers 2015; Hunt 2016; Gumpertz et al. 2017; Thomas et al. 2017).

Our finding that women are, at least in some respects, adversely influenced by being in numerical minority contributes to the understanding of the surprising persistence of economic gender gaps. In addition to the leaky pipeline, this would be consistent with many well-documented labor market phenomena, such as the persistent horizontal gender segregation in most labor markets (e.g. Blau and Kahn 2017). Part of the reason why women select different educations, industries and occupations than men may simply be that they feel deterred by male majority surroundings and believe that they cannot succeed there. It would also be consistent with the “gender tipping points”⁵ observed in several occupations and sectors (England et al. 2007; Pan 2015) as well as with women reporting a lower job satisfaction in workplaces dominated by men (e.g. Usui 2008; Lordan and Pischke 2016; Griffith and Dasgupta 2018).

Our results also speak to the current debate about gender quotas at top levels.⁶ If underrepresentation causes women to become less confident and lower their career 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 that increase female representation in settings where women are traditionally underrepresented may have other important long run consequences for women already in those contexts. From a broader perspective, our results speak to the recognition, retention, and promotion of competence in organizations, and our findings should

⁵ Gender tipping points denote the phenomenon whereby 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.

⁶ 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, Spain and, most recently, California, have adopted similar measures. Further, The European Commission has proposed legislation with the aim that, by 2020, 40% of non-executive directors shall be women.

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. In the next section, we discuss previous research and how it motivates our main research question as well as the potential mechanisms we explore. Section 3 presents the experiment design. The main results are presented in Section 4, whereas Section 5 covers the more exploratory analysis of mechanisms and Section 6 present an overview of outcomes. Finally, Section 7 concludes.

2. Previous Literature and Potential Mechanisms

Our study relates to many different strands of research. We add to the growing literature on gender and leadership, documenting a lower willingness to strive for leading positions among women than men (e.g. Butterfield and Powell 2003; Fox and Lawless 2004; Litzky and Greenhaus 2007; Fox and Lawless 2014; Kanthak and Woon 2015; Preece and Stoddard 2015; Thomas et al. 2017; Chakraborty and Serra 2018).⁷ Our findings are also consistent with research documenting an association between being male and leadership, whereby characteristics associated with being female are different from those associated with effective leadership (e.g. Eagly and Karau 2002; Koenig et al. 2011; Hoyt and Murphy 2016).

Our main research question—whether women’s willingness to lead is negatively impacted by being in minority—was motivated by a number of observations in previous literature related to gender norms, behavioral gender gaps, and group gender composition. In her seminal work, Kanter (1977b, 1977a) argued that women in minority in professional settings experience increased visibility, represent a contrast to the prevailing gender norms, and are confined to gender stereotypic roles.⁸ Consistent with this, Griffith and Dasgupta (2018) find that women in male dominated work places report lower job satisfaction, a less collegial work climate, and less equitable gender relations. Further, Gloor et al. (2017) offer

⁷ It should be noted that not all evidence support a gender gap in leadership motivation. A survey by Catalyst (2012) finds that male and female executives report equal desires to reach the CEO position. Further, Bursztyn et al. (2017) find that while single female MBAs state lower career orientation in a public setting, there are no gender gaps in stated professional ambition and leadership abilities in a private setting. However, none of these studies elicit actual leadership choices.

⁸ There is some evidence that women in minority experience stereotype threat, whereby fear of confirming a negative stereotype of one’s group, such as low mathematical skills for women, cause increased stress and impaired performance (e.g. Spencer, Steele, and Quinn 1999; Inzlicht and Ben-Zeev 2000). Further, according to Kanter’s critical mass theory (Kanter, 1977), women in minority will have limited influence in group interactions as long as they are very few. In line with this, a small empirical literature indicates that women’s impact and performance seem to increase as their numbers reach a critical mass of at least three individuals, or around 30-40% of the group (see, for example, Konrad, Kramer and Erkut 2008; Torchia, Calabro and Huse 2011; Schwarz-Ziv 2016; or Rossi, Hu and Foley 2017).

experimental evidence that female leaders receive lower ratings in male-majority groups than in gender balanced groups, and findings by Gagliarducci and Paserman (2012) indicate that female mayors who head an entirely male municipal council are the least likely to survive until the end of their term. Previous studies also find that female superiors benefit women both in the private and the political sector (Kunze and Miller, 2017; Baskaran and Hessami, 2018).⁹

While a large body of research indicates that women behave less aggressively and assertively than men (e.g. Niederle and Vesterlund 2007; Bertrand 2011; Coffman 2014; Exley, Niederle, and Vesterlund 2016), a smaller, but growing, number of studies suggest that this tendency may be reinforced when in minority. For example, Bursztyn, Fujiwara, and Pallais (2017) show that single female students avoid signaling high career ambitions in public, but only when surrounded by male peers. Similarly, results from Bowles, Babcock and Lai (2007) indicate that women face a social cost from negotiating assertively, especially when facing a male counterpart.¹⁰ In addition, when surrounded by men, women have been shown to take less risk (Sjögren Lindquist and Säve-Söderbergh 2011; Booth and Nolen 2012; Booth, Cardona-Sosa, and Nolen 2014), opt out of competition to a greater extent (e.g. Booth and Nolen 2012; Hogarth, Karelaia, and Trujillo 2012; Burow et al. 2017), and contribute their ideas to a lower extent (Chen and Houser 2017).

In addition to having a direct effect on women's willingness to lead, team gender composition could potentially influence women's willingness to lead through different channels, which need not be mutually exclusive. Below, we outline the primary channels that we explore in our analyses.

First, *women's expertise may be recognized to a lesser degree in male majority teams than in female majority teams*. A number of studies suggest that while both genders tend to undervalue women's expertise compared to that of men, men may do so more than women (e.g. Eagly et al. 1992; Grunspan et al. 2016; Boring 2017; Mengel, Sauermann, and Zölitz 2018). In line with this, further evidence suggests that women's status, influence, and speaking time decrease with their number (Karpowitz and Mendelberg 2015) and that men hold a more masculine construal of leadership than women do (Koenig et al. 2011). Women in male dominated environments may thus face more challenges to make their voices heard, gain acceptance in leadership roles, and be recognized for their expertise (e.g. Goldin and Rouse

⁹ However, research on whether female applicants benefit from female evaluators in, for example, recruitment committees is inconclusive (Moss-Racusin et al. 2012; Casadevall 2015; Bagues, Sylos-Labini, and Zinovyeva 2017).

¹⁰ Though see also Hederos Eriksson and Sandberg (2012) and Buser, Ranehill, and van Veldhuizen (2017).

2000; Thomas-Hunt and Phillips 2004; Sarsons 2017; Shurchkov and van Geen 2017). In our exploration of mechanisms, we use the votes in the leadership election as a measure of the support received from team members. To quantify a participant's influence in the team discussion, we compute the proximity between the team's answer and the participant's individual answer. We also measure the relative speaking time of women compared to men in the group discussion. Further, as an indirect measure of gender stereotypes, we use an Implicit Association Test (IAT).¹¹

Second, *women may perform worse when surrounded by many men*, and this may dampen their willingness to lead predominantly male teams. Several studies find that a larger share of women in the classroom increases female students' educational attainment (Hoxby 2000; Lavy and Schlosser 2011; Black, Devereux, and Salvanes 2013; but see also Hill 2017 and Oosterbeek and van Ewijk 2014 who find no effect on women).¹² In addition, previous research on competitiveness sometimes finds that women perform (relatively) worse when surrounded by, or competing against, men (Gneezy, Niederle, and Rustichini 2003; Antonovics, Arcidiacono, and Walsh 2009; Kuhnen and Tymula 2011; de Sousa and Hollard 2015, Booth and Yamamura 2017). We use women's performance in the second task, performed after the team discussions, to evaluate whether male dominated environments have an adverse effect on female performance.

Third, even if the above mentioned mechanisms are not present, women may believe that they are. That is, *women in male majority teams may believe that they perform (relatively) worse, or that they are not supported by their team members*. One of the most robust gender gaps found in the behavioral literature is that women are often less confident in their own ability than men (e.g., Lundeberg, Fox, and Punócohař 1994; Barber and Odean 2001; Niederle and Vesterlund 2007). If selection into leadership is based on self-estimated competence, a gender gap in confidence may account for at least part of the observed gender gap in leadership motivation. Moreover, if women believe that men are more competent, or if other aspects of male dominated environments influence women's confidence negatively, women's leadership motivation will decrease further.¹³ To quantify the importance of confidence, we elicit

¹¹ The IAT measures each participant's association between maleness and leadership, allowing us to assess whether women in male majority teams face more implicit bias than women in female majority teams.

¹² Studies also find positive effects of increased female ratios on the classroom environment, girls' academic self-concept (Lavy and Schlosser 2011; Belfi et al. 2012) and dropout rates (Anil et al. 2016). To our knowledge, studies investigating the impact of peer gender composition on educational choice find conflicting results (e.g. Anelli and Peri 2017; Zölitz and Feld 2017)

¹³ A number of other mechanisms may also contribute to worsening the confidence of women surrounded by many men. Dasgupta, Scircle and Hunsinger (2015) find that female majority groups leads to less anxiety and more participation on behalf of women in engineering. In an economic experiment, Reuben, Sapienza and Zingales

participants' relative performance beliefs by asking them to guess, after the team discussion, how well they performed in the first task compared to their team members. Further, as a second measure of confidence, we elicit how much participants change their individual answer to the first task, when given the opportunity to update their initial answer after the team discussion. Finally, to determine if women in male majority teams anticipate less support as leaders, we elicit participants' beliefs about how many votes they will receive in the election.

In the next section, we describe the experimental design, and our measures of the different channels, in detail.

3. Experiment Design

The main part of the experiment comprised two parts, during which participants worked on two tasks, the "Lost at Sea" task and the "Desert Survival" task. Below, we first describe these tasks, and then the different parts of the experiment. All instructions, and complete descriptions of the tasks, are available in Appendix B.

2.1 The 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.¹⁴

(2014) find that a gender gap in hiring arises partly because men have very optimistic beliefs about their previous performance. If women in male majority environments are surrounded by individuals that overestimate their performance, this may of course have negative effects on their relative confidence. Sarsons (2017) finds that women receive less credit for their contribution to team work when working together with men. If the relative judgment of a woman's skill decreases in male majority teams, this may impact the signal she receives from her team members and, consequently, her confidence in her own ability. Another possible reason why women may face challenges in male dominated environments is put forward in Babcock, Recalde and Vesterlund (2017), who find that women in mixed, but not single, gender groups tend to accept non-promotable tasks to a larger extent than men. In male majority environments, a larger share of non-promotable tasks with low visibility may fall on women, leading to even more skewed estimations of relative ability. Further, male dominated environment may result in a lack of female role-models (e.g. Bettinger and Long 2005; Hoffman and Oreopoulos 2009; Carrel, Page and West 2010; Blau et al. 2010; Porter and Serra 2017).

¹⁴ 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*

We chose these tasks because they are suitable for both team work and individual work, and have answers that are open for discussion while still perceived as meaningful and related to individual competence.¹⁵ Further, since we are primarily interested in the behavior of women in male dominated areas, our aim was to include tasks that had, if anything, a slight male stereotype.¹⁶ These types of tasks have also been used in previous research on group work and gender (Thomas-Hunt and Phillips 2004).

The participants' submitted answers were compared to answers provided by a panel of survival experts. Answers closer to that of the expert panel generated higher points and, subsequently, higher payoff. For each item that the participants 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 ranged between 25 and 50 Swiss Francs¹⁷, and was calculated according to the formula:

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

3.2 Overview of the experiment

The main experiment consisted of ten stages, as summarized in Figure 1. All answers to the tasks, as well as the elicitation of relative performance and electoral beliefs were incentivized. At the end of the experiment, one part of the experiment was randomly selected to count for the participant's final payment.¹⁸

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.

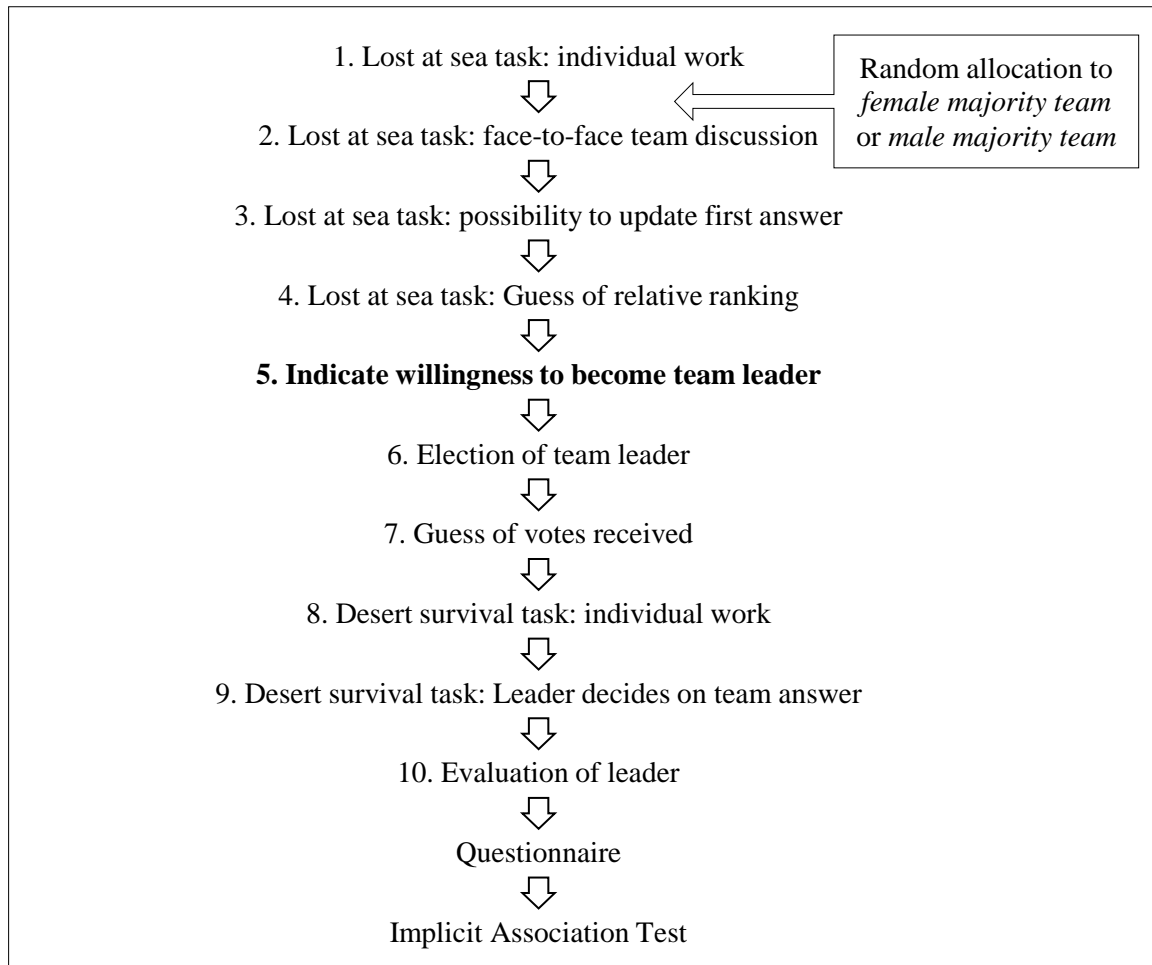
¹⁵ To assess whether participants' task performance is related to skill, and not only luck, we compare the participants' average performance to randomly generated answers. Randomly ordering the items in the Lost at Sea task would result in 33.0 penalty points while participants in our sample achieve, on average, 24.6 penalty points ($t=29.5$, $p<0.001$, t-test). Thus, participants perform significantly better than chance, indicating that performing well in the Lost at Sea task is a function of skill.

¹⁶ Our results confirm that the "Lost at Sea" and "Desert Survival" tasks have a slight male stereotype. In the questionnaire after the experiment, participants were asked to indicate what gender, if any, performed better at these tasks. 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. We did not aim for a task with a strong male stereotype for two reasons. First, we did not want the stereotype of the task to overrule the effect of group gender composition, or cause only very few women to aim for leadership. Further, most leadership positions comprise a mixture of tasks, including people management, which may have a more mixed gender stereotype.

¹⁷ At the time of the experiment 1 Swiss Franc corresponded to about USD 1.

¹⁸ At the end of the study, one of the Stages 1, 2, 3, 8, or 9 was randomly selected to count for the participant's final payment. If Stage 3 was chosen, the earnings from Stage 4 (relative performance beliefs) were added to those

Figure 1. Overview of the experiment.



Stages 1-4: Individual performance, influence and confidence

In the first stages of the experiment, we measured participants' individual task performance, influence in the team's decision making process and confidence in their own performance. First, before any teams were formed, participants worked with the Lost at Sea task at their computer for 8 minutes, and provided an individual answer. This gives us a measure of individual task-related ability. Then, participants were randomly allocated to teams of four, such that each team comprised either three men and one woman, or one man and three women.¹⁹ Each team moved to a separate room and had 10 minutes to discuss the Lost at Sea task face to face, and agree on a joint ranking of the 10 items. Before discussing the task, the

from Stage 3. Similarly, if Stage 8 was chosen, the earnings from Stage 7 (beliefs about electoral support) were added to those from Stage 8. At the end of Stage 9, when the leaders decided on the incentivized, joint group answer, we asked all non-leaders to provide the answer they would have provided if they had been the leader. These Stage 9 answers from non-leaders were elicited hypothetically. The reason for this was to keep the payoff associated with the leader's answer salient, and to avoid incentive differences between leaders and non-leaders.

¹⁹ We decided to only study these two types of teams in order to maximize statistical power.

team members were asked to introduce themselves, ensuring that they would be able to recognize each other via name during the subsequent parts of the experiment.²⁰ After the team discussion, participants went back to their own computer.²¹

Based on these first two stages in the experiment, we compute a measure of each team member's relative influence in the group decision making process. To obtain this measure, we first compute, for each item, the absolute difference in rank between the participant's individual answer and their team's joint answer. Then, we sum these differences over all ten items, and divide the participant's sum by the total sum of all team members. Thus, our measure of influence (inversed for ease of interpretation) is 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).²² We obtain an additional, related, measure from audio recordings of the team discussions, allowing us to register the share of female/male speaking time in each team.²³ Using these two measures, we can explore if male majority environments reduce women's influence and/or level of participation in the decision-making process.

After the team discussion, we asked participants to guess how well they performed individually in the first stage of the experiment, compared to the other participants in their team. The guess was incentivized such that participants received an additional 2 Swiss Francs in case their guess was correct.²⁴ The resulting variable is ordinal and ranges from 1 (the participant believed he/she was worst in the team) to 4 (the participant believed he/she was best in the team).

To obtain a second measure of confidence, participants were also given the opportunity to update their previous, individual answer. To do so, they were presented with a screen displaying their own individual answer from Stage 1 and their team's answer from Stage 2 in two adjacent columns. They were then asked to enter an individual answer once again in a third

²⁰ Audio recordings confirm that all teams complied with these instructions.

²¹ During the team discussion, all team members were asked to fill in an answer sheet, ensuring that they remembered their team's answer when they returned 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. 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.

²² To illustrate this measure of relative influence, consider an extreme example: Assume that the team's final answer is identical to the individual answers of participants 1-3 (total distance=0), but differs from that of participant 4 (total distance=X). In that case, the influence of participants 1-3 would be computed as $1 - [0/(0+0+0+X)] = 1$ and the influence of participant 4 would be computed as $1 - [X/(0+0+0+X)] = 0$.

²³ To obtain this measure, a research assistant listened to the audio tapes and recorded whether a man, a woman, or no-one was speaking. This allows us to compute a *team-level* measure of the share of female and male speaking time, indicating how much, on average, each gender participated in the discussion. However, we cannot link the data on speaking time to individual participants, unless the participant is of the minority gender in their team.

²⁴ Participants received this additional payment only if Stage 3 (the updated individual answer to the "Lost at Sea" task) was randomly chosen to count for payment.

column, and were free to make updates as they saw appropriate. The updated answer was incentivized, and was randomly chosen to count for payment at a similar rate as any other answer to the task. To measure how much a participant updates her answers after the team discussion, relative to the team members, we first compute the absolute difference in rank between the participant's original answer and their updated answer for each item. Then, for each participant, we sum these differences over all ten items and divide with the sum of all team members' differences. Thus, our measure of updating 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).

The measure of updating allows us to explore if women in male majority teams are less confident in their ability to answer the task, and thus more swayed by the group discussion, than women in female majority groups.

Stages 5-10: Leadership aspirations and voting

Our main outcome variable, how much participants wanted to become the leader of their team, was elicited in Stage 5 of the experiment. Participants were informed that they would once again perform a task that resembled the Lost at Sea task, but which involved a different survival situation and 10 different items. The new task would be performed in the same team, 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. The responsibility of the leader thus includes several important aspects of what is generally associated with a leadership role. It involves making and implementing a final decision on behalf of the whole team, and requires both personal expertise and the ability to take in and synthesize information from the other team members. In this design, it is in the economic interest of the group, and each participant, to elect the most capable leader.²⁵

To elicit willingness to lead, we asked all team members to indicate how much they wanted to become the team leader on a scale between 1 and 10.²⁶ The answer to this question

²⁵ We did not provide any additional payment to the leader. Thus, apart from the incentives to elect the most capable leader there are no monetary incentives for becoming the leader. We chose this design since we wanted participants' indicated willingness to lead to reflect only intrinsic motivation to lead and relative performance beliefs. We also wanted to keep the design simple and comprehensible to all participants, avoiding complex strategic components.

²⁶ When eliciting this measure, 1 denoted the highest willingness to lead and 10 the lowest. In the following analysis, for a more intuitive interpretation, we reversed the scale so that higher numbers indicate a higher

constitutes our primary outcome variable. Participants were informed up front that the two team members who indicated the highest willingness to become the leader would become candidates in the subsequent election. Thus, participants were aware that their stated willingness to lead would have a direct impact on their probability of becoming the team leader. We use the 1-10 response scale, rather than a binary choice, to obtain a more precise measure of the strength of participants' willingness to lead. This measure can be thought of as a proxy for how much effort participants would exert in trying to become the team leader, or how strongly they would argue in favor of themselves in a leadership selection process.²⁷

In order to elect a leader, all team members provided their anonymous vote through a ranking of the other three team members (excluding themselves), with their most preferred leader at position 1, and their least preferred leader at position 3. Participants provided their ranking before the names of the two candidates were revealed. 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.

We use the participant's rank in the election to approximate support from the other team members, exploring if women receive less support in male majority environments than in female majority environments. The variable used for these analyses ranges between 1 (the participant was ranked last) and 4 (the participant was ranked first).²⁸

Before participants were informed about the outcome of the election, they were asked to guess their rank by providing a number between 1 and 4. The guess was incentivized such that participants received an additional 2 Swiss Francs in case their guess was correct.²⁹ The resulting measure allows us to explore if women in male majority teams are less confident than women in female majority teams in how much support they will receive from their team members in the election. The variable used for this analysis is ordinal, taking a value between

willingness to lead. A similar, but binary, approach to elicit willingness to lead is used in Erkal, Gangadharan, and Xiao (2018) who explore whether changing the default from "opting in for leadership" to "opting out from leadership" matters for the share of female leaders.

²⁷ A participant's likelihood of becoming a candidate in the election will depend on both the participant's own willingness to lead, and that of the other team members. However, importantly, choosing a higher number will always weakly increase the likelihood that a participant becomes a candidate.

²⁸ In these analyses, the participants with equal points are assigned the average rank. In the actual election, however, ties were broken randomly.

²⁹ Participants received this additional payment only if the upcoming Stage 8 (the individual answer to the "Desert Survival" task) was randomly chosen to count for payment. 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.

1 (the participant thought that they would come last in the election) and 4 (the participant thought that they would win the election).

After the identity of the leader was revealed, participants were introduced to the second task, the Desert Survival task. As in the first Lost at Sea task, each participant, including the team leader, individually ranked the 10 items of 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. Since previous literature sometimes suggests that male majority environments influence women's performance negatively, we use the individual performance in the second task to assess whether any treatment differences may be driven by women in male majority teams performing worse.

When all team members had provided an individual ranking, their rankings were transmitted to the team leader and presented side by side in a comparable way in a 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 part of the experiment was selected for payment, all participants were paid based on this team answer. While the leader worked on the final team answer, the other three team members performed the same task, albeit unincentivized.

Finally, 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 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 Implicit Association Test (IAT), eliciting the strength of the participants' implicit associations between leadership and being male. The IAT score is measured on a scale between -2 and +2, where a positive (negative) score indicates bias in the sense that the respondent finds it easier (more difficult) to associate men than women with leadership. Our primary IAT measure is the average IAT score of the participant's team members (excluding the participant's own score), which will inform us about the general team climate with respect to gender stereotypes and leadership that participants faced during the team discussions. 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.

3.3 Hypotheses

As described in our pre-analysis plan, based on the previous research outlined in the introduction, we formed two primary hypotheses about the impact of group gender composition on women's willingness to lead and the related gender gap.³⁰

H1 Women are more willing to become the leader in female majority teams than in male majority teams.

*H2 The gender difference in willingness to become the leader is smaller in female majority teams than in male majority teams.*³¹

3.4 Experiment Procedures

The experiment was conducted at the laboratory for experimental economics at the University of Zurich in May 2017.³² 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). All sessions thus included both male and female majority groups. If not enough participants showed up for a session, the session was run with 16 participants (i.e. 4 teams) instead. In total, 25 sessions were run with 5 teams and 5 sessions with 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.

³⁰ In order to keep the pre-analysis plan concise and focused, we specified only two tests as our primary analysis – the regressions presented in specifications 2 and 4 in Table 2. As secondary analysis, to explore the mediating factors, we specified the regressions presented in Panel b of Table 3 and Table A.2. The latter analysis is also illustrated in Panel b of Figure 6. The reason for our focus on regression analyses in the pre-analysis plan is that we considered it important to control for relative performance. To facilitate interpretation of the regression analyses in the paper, we reversed the scale of the gender dummy compared to what was specified in the pre-analysis plan (we use a dummy for “Male” instead of “Female”). Apart from the heterogeneity analysis presented in footnote 39, our pre-analysis plan did not specify any other test.

³¹ Since previous literature provides little support for a directional hypothesis regarding how group gender composition influences male behavior, we refrained from an explicit hypothesis in this case. However, implicit in our second hypothesis is the assumption that team gender composition does not influence men's behavior strongly in the same direction as women's behavior.

³² In Switzerland, the majority of adolescents go through vocational education and training, and less than 15% of the population obtains a degree from a university of institute of technology. Our sample is thus drawn from a selected part of the population relevant for our research question. Further, we capture this population before they have any extensive experience of working in male or female majority settings.

Table 1: Number of Observations

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

Instructions were distributed directly before each relevant part of the experiment.³³ 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 willingness to lead 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 control questions correctly. The experiment was programmed in Ztree (Fischbacher, 2007). The average payment was 49 Swiss Francs 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.

4. Main Results

In this section, we first consider the general gender gap in willingness to lead. Then, we turn to our main research question, and explore whether team gender composition affects how much women want to lead the team.

4.1 The gender gap in willingness to lead

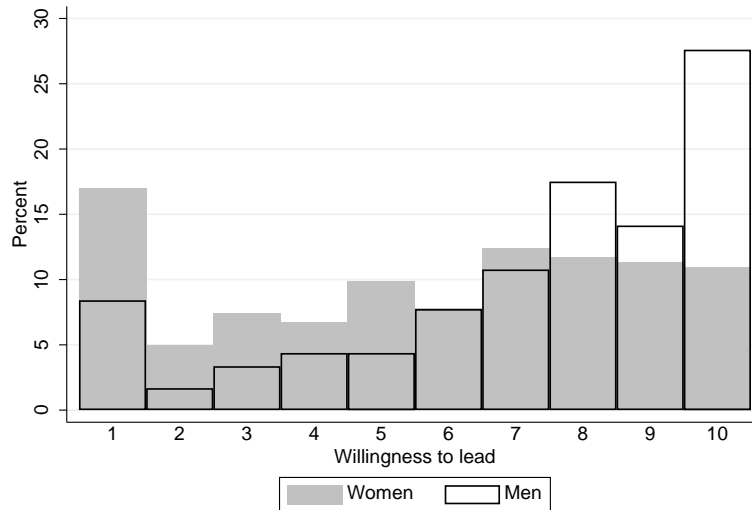
Figure 2 presents the distribution of willingness to lead by gender. On the 1-10 scale, 1 indicates the lowest possible interest in leading the team, and 10 indicates the highest possible interest. On average, men state a willingness to become the team leader that is 1.63 units higher than that of women (men: 7.27, women: 5.63, Cohen's $d=0.56$).³⁴ This average gender difference in willingness to lead 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

³³ All instructions can be found in Appendix B.

³⁴ As a point of reference, an effect size of a Cohen's d of 0.56 implies that there is a 65 percent probability that a randomly chosen man is more willing to become the team leader than a randomly chosen woman.

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 team leadership.

Figure 2: Distribution of willingness to lead by gender



Male participants perform on average 1.8 points (0.26 standard deviations) better than women in the first, individual, Lost at Sea task ($p=0.001$, see Figure A.1 in the Appendix for the distributions). To assess if this gender gap in task performance may account for part of the gender gap in willingness to lead, Table 2 presents OLS regressions with willingness to lead as the dependent variable. 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). Controlling for relative performance, the estimated gender gap in willingness to lead remains highly significant and sizeable at 1.58 units ($p<0.001$), indicating that the observed gender gap in willingness to lead cannot be accounted for by gender differences in the ability to solve the task.³⁵

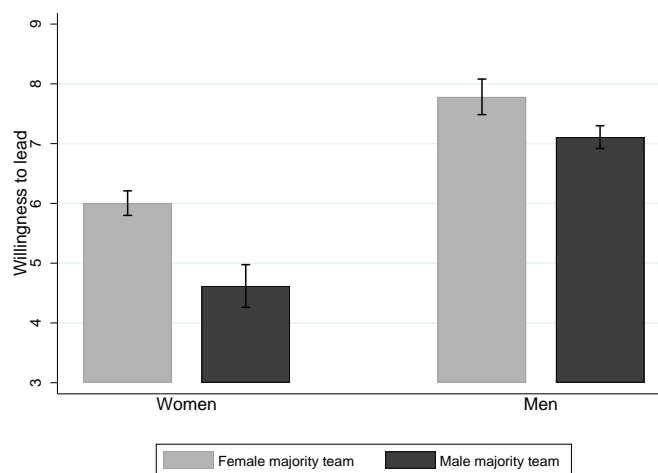
4.2 The effect of team gender composition

Our main research question is whether team gender composition affects women's willingness to lead the team. Figure 3 presents the average willingness to lead by gender and treatment.

³⁵ All results in Table 2 are robust to instead using dummy variables for relative performance as controls, or controlling for absolute performance (number of penalty points in the task). The results are also 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.

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 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 willingness to lead is relevant in size, and, although slightly smaller, comparable in magnitude to the average gender gap in willingness to lead.³⁶

Figure 3: Distribution of willingness to lead by gender and treatment



Note: Error bars represent standard errors.

A Kolmogorov-Smirnov test confirms that the distribution of women’s willingness to lead differs by team gender composition ($p=0.005$, see Figure A.3 (a) in the Appendix). 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.

Men also state a higher average willingness to lead in female majority teams than in male majority teams (7.78 compared to 7.11). However, this effect of team gender composition for men is somewhat smaller than for women and not statistically significant ($p=0.133$). Similarly, while the distribution of men’s willingness to lead shifts toward higher values when

³⁶ One way to evaluate the size of this effect is to rank all team members based on how willing they are to become the team leader (1=highest, 4=lowest). Based on their willingness to lead, women rank, on average, 0.63 positions lower than men ($p<0.001$). The average rank of women in male majority teams is 0.51 positions lower than that of women in female majority teams ($p=0.001$). Noticeably, as shown in Figure A.2 in the Appendix, more than half of women in male majority teams indicate the lowest willingness to lead out of all team members.

moving from male to female majority teams, this shift is slightly weaker than for women and not significant ($p=0.102$, Kolmogorov-Smirnov test, see Figure A.3 (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. The difference is quite small (women rank, on average, 2.57 vs. 2.47 in male and female majority teams respectively) and insignificant ($p=0.50$), and controlling for relative performance does not impact the treatment effect in any important way. Specification 3 in Table 2 presents an OLS regression showing the estimated impact of team gender composition on women's willingness to lead (given by the coefficient of *Male majority team*). When controlling for relative performance, in specification 4, this effect remains stable (1.39 vs. 1.36) and significant ($p<0.001$).^{37,38} Thus, the observed impact of team gender composition on women's willingness to lead cannot be accounted for by performance differences.³⁹

As further illustrated in specifications 3 and 4, the estimated interaction effect between *Male* and *Male majority team* is positive, indicating that the gender gap in willingness to lead the team is larger in male majority teams than in female majority teams. However, this effect is not statistically significant ($p=0.231$). Thus, we do not find support for our second hypothesis that the gender gap in leadership motivation is significantly smaller in female majority teams compared to male majority teams. Part of the explanation why we do not observe a smaller gender gap in willingness to lead in female majority teams is that men are also more willing to become the leader of these teams.

³⁷ In column 4, we also interact the measure of relative performance with *Male*. Thus, we allow for gender-specific effects of relative performance on the outcome variable.

³⁸ For men, the estimated treatment effect (given by the sum of the coefficients of *Male majority team* and *Male X Male majority team*) drops from 0.67 to 0.65, and remains insignificant, when controlling for relative performance.

³⁹ In our pre-analysis plan we specified an exploratory heterogeneity analysis, investigating whether some characteristics were associated with a stronger impact of team gender composition among our female participants. The characteristics specified in our pre-analysis plan were IAT score, business school, mother's education, father's education, political orientation, risk taking, perceived gender stereotype of the task, perceived gender stereotype of leadership, and leadership experience. Restricting the sample to women, and interacting these variables with *Male majority team*, the only characteristic that significantly influences the treatment effect is the perceived gender stereotype of the task. The negative impact of male majority teams on women's willingness to lead is larger among women who believe that men are, on average, better at the task. These results are available from the authors on request.

Table 2: Differences in willingness to lead across gender and team composition

	<i>Dependent variable: Willingness to lead (1-10)</i>			
	(1)	(2)	(3)	(4)
Male	1.633*** (0.261)	1.584*** (0.262)	1.778*** (0.361)	1.417* (0.600)
Male majority team			-1.386*** (0.403)	-1.355*** (0.402)
Male X Male majority team			0.713 (0.593)	0.706 (0.588)
Constant	5.633*** (0.174)	6.303*** (0.345)	6.005*** (0.188)	6.792*** (0.465)
N	580	580	580	580
Controls:				
Relative performance first task		YES		YES
Male X Rel. perf. first task				YES
F-test:				
Male majority team + 'Male X Male majority team':			-0.673 (F=3.586)	-0.650 (F=3.380)

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

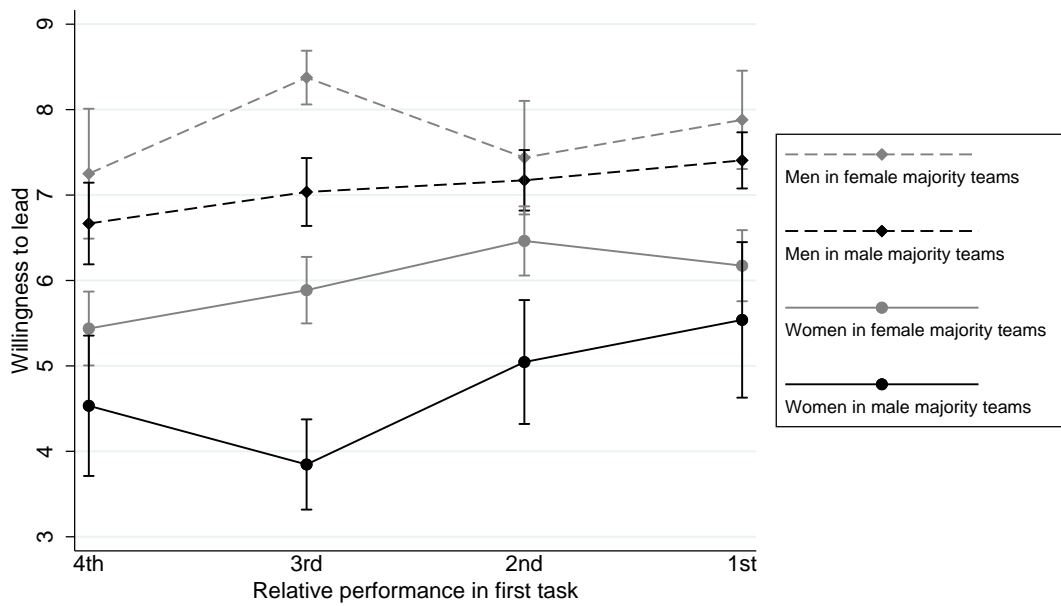
Note: OLS regressions using willingness to lead (1-10) as the dependent variable. Standard errors are clustered at the team level. The final row shows results from an F-test, testing the treatment effect for men.

Figure 4 shows the average willingness to lead by gender, treatment and the participants' performance ranking within their team. The figure illustrates that, regardless of relative performance and team gender composition, men's average willingness to lead 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.⁴⁰ Moreover, for each level of relative performance, men in female majority teams indicate 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.

To sum up, men are, on average, more willing to become the team leader than women. Both women and men are more inclined to aim for leadership when assigned to female majority teams, but 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 a set of mechanisms that may account for these effects.

⁴⁰ However, comparing the willingness to lead of the women who were best in their team (N=65) with the men who performed the worst (N=57), this difference is not statistically significant ($p=0.144$).

Figure 4: Willingness to lead by relative performance, gender and treatment



Note: Error bars represent standard errors.

5. Mechanisms

Why do women shy away from leading their teams, and why are women particularly averse to becoming the leader of male majority teams? In Section 2 we described a number of potential mediating factors that our experiment was designed to capture: (i) confidence in own ability to perform the task (measured as relative performance beliefs and tendency to update one's individual answer after the team discussion), (ii) electoral outcomes, (iii) anticipated electoral outcomes, (iv) influence and relative speaking time in the team discussion, (v) performance in the second task, and (vi) gender stereotypes. Table A.1 in the Appendix summarizes how we define and measure each of these factors.

In this section, in an exploratory analysis, we systematically study to what extent these factors may explain gender gaps in willingness to lead. 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 majority 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 willingness to lead and the impact of team gender composition on women's willingness to lead.⁴¹

5.1 Gender gaps in the mediating factors

To explore whether there is a baseline gender gap in the mediating factors, panel (a) of Table 3 presents regressions with each mediating factor as the dependent variable, controlling for participant gender and relative performance. In line with previous studies, we find that men are, on average, more confident than women. The first column indicates that, controlling for their actual relative performance, men believed that their own performance ranked, on average, 0.44 positions higher than did women ($p < 0.001$).⁴² Similarly, the second column indicates that, controlling for relative performance, men were less prone than women to update their individual answers following the team discussion. The average man's share of the team's total updating was 4.3 percentage points lower than that of the average woman ($p < 0.001$).

Next, we explore whether women's actual and expected electoral outcomes are different from those of men. The third column indicates that men were ranked, on average, 0.42 positions better than similarly performing women in the election ($p < 0.001$). In line with this finding, the fourth column shows that men also believed that they would be ranked, on average, 0.58 positions better in the election than did similarly performing women ($p < 0.001$).

Further, as indicated by column 5, men appear to be more influential than women in the team's decision making process. ($p < 0.001$). Analyses of the sound recordings paint a similar picture, showing that the average man speaks more in the team discussion than the average woman, both in absolute terms (275 vs. 220 seconds) and in relative terms (28.2% vs. 22.7 % of the team's total speaking time). The average woman's share of the team's total speaking time is significantly lower than 25% ($t = 2.70$, $p = 0.008$, t-test) while the average man's share is significantly higher than 25% ($t = 3.75$, $p < 0.001$, t-test).⁴³

⁴¹ In this section we focus on the mechanisms behind the impact of team gender composition on *women's* willingness to lead (as outlined in our pre-registered pre-analysis plan). Although not discussed in the text, corresponding tests addressing the mechanisms behind the impact of team gender composition on *men's* willingness to lead can be found in the bottom rows of Table 4, and in Table A.2 in the Appendix.

⁴² The results presented in Table 3, using ordinal outcome variables (Guess rank first task, Rank election and Guess rank election), are robust to running ordered Probit regressions instead of OLS regressions. These results are available from the authors on request.

⁴³ Since we cannot link the data from the sound recordings to the individual level, all analyses of speaking time are conducted at the team level. Thus, for these analyses, we collapse the data on the team level, giving 139 observations (the sound recordings of 6 teams malfunctioned and could not be analyzed). In the pre-registered pre-analysis plan we assumed that we would be able to link speaking time to the individual participants. For this reason, the analyses of speaking time presented in the paper depart somewhat from the pre-analysis plan.

Finally, we do not find strong support for a gender gap in the two last mediating variables. The second to last column of panel (a) in Table 3 illustrates that the gender difference in individual performance in the second task is insignificant ($p=0.090$).⁴⁴ The last column shows that, as expected across the two treatments, women and men have team members with similar implicit gender-leadership associations ($p=0.398$).⁴⁵

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. The next section explores whether our mediating factors differ between women in female majority teams and male majority teams.

5.2 Do male majority environments impact women negatively?

In addition to the baseline gender gaps in our mediating variables, it is possible that women in male majority environments face different challenges than women in female majority environments. For example, as discussed in Section 2, women in male majority environments may struggle to remain confident, be influential, or to gain acceptance as leaders and experts.

In panel (b) of Table 3 we regress each of the potential mediating factors on variables indicating participant gender, team gender composition, an interaction between the two, and relative performance.⁴⁶ 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 coefficient of *Male majority team* in 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 with the gender gap in willingness to lead, the impact of team gender composition on female confidence appears relevant, and while somewhat smaller, it is comparable in magnitude to the average gender gap in confidence in the whole sample.

The negative impact of male majority teams on women's confidence cannot be accounted for by *actual* gender differences in task ability. However, what if participants – regardless of team gender composition – *believe* that the average gender gap in task ability is larger than it actually is? Could such beliefs mechanically account for the observed treatment effect on women's confidence? To address this question, Figure A.4 shows results from a

⁴⁴ The regression reported in Column 6 of panel (a) in 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 ($p=0.055$).

⁴⁵ 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, $p<0.001$).

⁴⁶In panel (b) of Table 3, as in column 4 of Table 2, we also interact the measure of relative performance with Male. Thus, we allow for gender-specific effects of relative performance on the outcome variable.

simulation exercise based on the assumption that both male and female ability is normally distributed. The figure illustrates how, on average, various gender performance gaps translate into different relative performance ranks of women in female vs. male majority teams.⁴⁷ The average self-estimated rank of women in female majority teams (2.53) is consistent with an underlying gender ability gap of about 0.10 standard deviations, while the average self-estimated rank of women in male majority teams (2.92) is consistent with a gender ability gap of about 0.51 standard deviations. Thus, under the assumption of normally distributed ability, the gap in relative performance beliefs between women in male and female majority teams cannot be consistently accounted for by pre-treatment beliefs about an overall gender gap in ability.

In line with the finding that women in male majority teams have lower relative performance beliefs, the second column of Table 3 shows that women in male majority teams are more prone to update their individual answers following the team discussion than women in female majority teams ($p=0.006$).

We further find some support that team gender composition impacts women's actual, and expected, electoral outcomes. The third column in panel (b) of Table 3 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 ($p=0.080$). As illustrated in the fourth column, women in male majority teams also anticipate less support from team members in the election. Compared to women in female majority teams, women in male majority teams expect that they will be ranked 0.26 positions worse in the election ($p=0.020$).

The impact of team gender composition on women's influence further reinforces the pattern presented above. Column 5 indicates that, controlling for relative performance, male majority teams generate team answers farther from the individual answers of female team members than do female majority teams ($p=0.015$). However, the analysis of the sound recordings of the team discussions reveal that women's average share of the total speaking time does not vary substantially or significantly between the two types of teams (22.5% in male

⁴⁷ In the simulations, task ability is assumed to be normally distributed with a standard deviation of 1 for both genders. The mean ability of women is kept constant at 0, while the mean ability of men, α , varies from 0 to 1 standard deviation higher than women's. Thus, α gives us the average gender gap in ability, expressed in standard deviations. In the simulations, we draw 100 000 teams consisting of three men and one woman, and 100 000 teams consisting of three women and one man, for each gender performance gap. Figure A.4 plots how the average performance ranking of women in the two types of teams vary with the gender performance gap.

majority teams vs. 23.0% in female majority teams, $p=0.239$). As illustrated in Figure A.5 in the appendix, men's share of the total speaking time reacts more strongly to team gender composition, with men in female majority teams speaking the most (25.8% in male majority teams vs. 30.1% in female majority teams, $p=0.007$).

The last two columns in panel (b) of Table 3 explore whether women in male majority teams perform worse at the second task, or face team members who, on average, associate maleness with leadership to a larger extent, than women in female majority teams do. We find no evidence that women perform worse in male majority teams ($p=0.736$). However, since men tend to express a stronger implicit association between maleness and leadership, we do find that women in male majority teams face team members with a stronger male-leadership stereotype ($p=0.004$).

Figure 5 shows the average of each mediating variable, separated by gender, treatment and actual performance ranking. The picture that emerges is consistent with that of willingness to lead from Figure 4. For each level of relative performance, with only a few minor exceptions, the point estimates indicate that men in female majority teams are the most confident in their ability, the most influential, get the most votes, and expect this to be the case. They are followed by men in male majority teams, and then women in female majority teams. The worst outcomes are observed for women in male majority teams who are the least confident, the least influential, receive the least support in the election, and expect this to be the case. Noticeably, as shown in panels (c) and (d), with one minor exception, men receive, and expect to receive, more votes than women regardless of treatment and relative performance. In fact, controlling for performance, men in female majority teams receive more votes than any other group ($p<0.001$).⁴⁸

⁴⁸ This p-value is based on OLS regressions with the average votes a participant received as the dependent variable, controlling for relative performance in the first task (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.

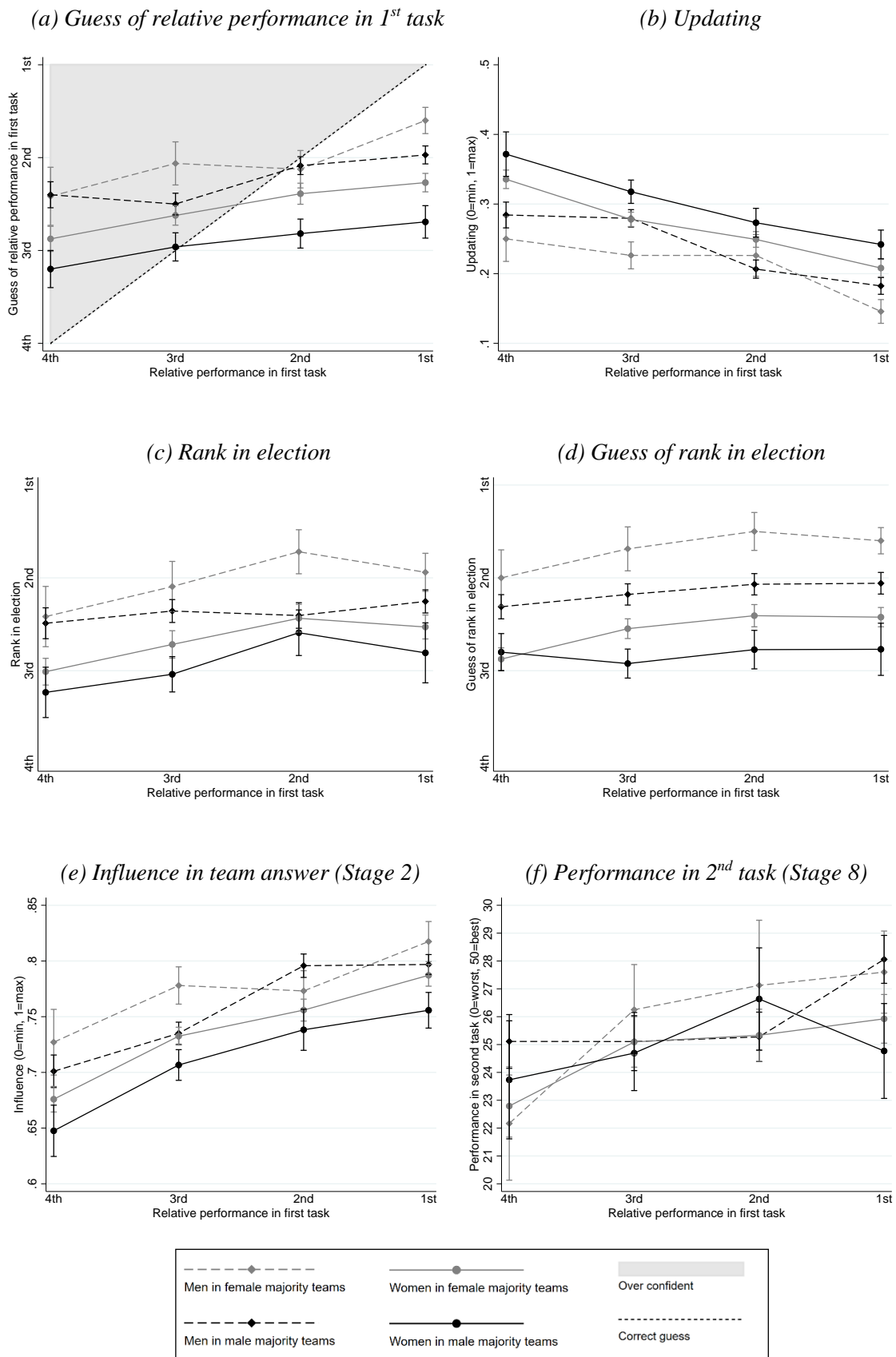
Table 3: Effects of gender and team gender composition on mediating variables

Outcome variable:	Guess relative perf. 1 st task <i>1=worst</i> <i>4=best</i>	Updating <i>0=minimum</i> <i>1=maximum</i>	Rank election <i>1=lowest</i> <i>4=highest</i>	Guess rank election <i>1=lowest</i> <i>4=highest</i>	Influence <i>0=minimum</i> <i>1=maximum</i>	Performance 2nd task <i>0=worst</i> <i>50=best</i>	IAT scores of team members <i>-2=female bias</i> <i>2=male bias</i>
(a) Overall gender gap							
Male	0.441*** (0.066)	-0.043*** (0.008)	0.424*** (0.088)	0.578*** (0.067)	0.029*** (0.007)	1.005 (0.589)	0.014 (0.016)
Constant	2.863*** (0.087)	0.176*** (0.012)	2.604*** (0.131)	2.649*** (0.095)	0.817*** (0.010)	27.390*** (0.745)	0.202*** (0.022)
<i>N</i>	580	580	580	580	580	580	580
(b) Effect of team gender composition							
Male	0.508** (0.171)	-0.046 (0.024)	0.412 (0.265)	0.792*** (0.192)	0.027 (0.019)	1.718 (1.666)	0.008 (0.035)
Male majority team	-0.371*** (0.098)	0.032** (0.011)	-0.227 (0.129)	-0.262* (0.111)	-0.023* (0.009)	0.333 (0.985)	0.088** (0.030)
Male X Male majority team	0.150 (0.161)	-0.005 (0.021)	-0.120 (0.231)	-0.201 (0.169)	0.007 (0.018)	-0.387 (1.398)	-0.014 (0.022)
Constant	2.952*** (0.103)	0.164*** (0.015)	2.771*** (0.164)	2.743*** (0.123)	0.827*** (0.011)	26.988*** (1.014)	0.167*** (0.032)
<i>N</i>	580	580	580	580	580	580	580
F-test:							
'Male majority team' +	-0.221	0.027*	-0.346**	-0.463***	-0.016	-0.054	0.074**
'Male X Male majority team'	(F=3.780)	(F=4.750)	(F=6.988)	(F=16.662)	(F=2.343)	(F=0.003)	(F=6.733)

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

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). The regressions in panel (b) also control for relative performance in Stage 1 interacted with *Male*. The final row presents results from an F-test, testing the impact of *Male majority team* on male participants. Table A.1 in the Appendix summarizes how all outcome variables are defined.

Figure 5: Averages of mediating variables, by relative performance, gender and treatment



Note: Error bars represent standard errors.

5.3 To what extent do the mediating factors drive gender and treatment differences in willingness to lead?

In this section, we assess how much the elicited mediating variables may contribute to the observed gender and treatment differences in willingness to lead. To do so we run regressions with and without the mediating variables as controls and observe to what extent the estimated gender and treatment effects decrease. This analysis rests on several strong assumptions, such as the mediating variables being uncorrelated with other, omitted, variables that may influence the relevant outcome. Further, the elicited mediating measures are likely to be noisy measurements of the true underlying factors. Thus, while this analysis may give an indication of why male majority environments impact women's willingness to lead, the results should be interpreted with caution.

To assess the impact of the mediating variables on the gender gap in willingness to lead, we use the regression from column 2 of Table 2 – regressing willingness to lead on gender and relative performance – as a starting point. Then, we add the mediating variables as controls, first one by one, and then all together, and measure the resulting decrease in the estimated gender gap in willingness to lead (i.e., the change in the coefficient of the gender dummy *Male*). As shown in Panel (a) of Figure 6, the largest decreases in the estimated gender gap occur when we control for relative performance beliefs and beliefs about electoral outcomes.⁴⁹ The estimated gender gap in willingness to lead decreases by almost half when introducing a control for relative performance beliefs, and by 64% when controlling for participants' expectations about how many votes they will receive. Thus, beliefs about relative performance, our most direct measure of confidence, and beliefs about electoral outcomes, seem to account for a sizeable part of the gender gap in willingness to lead among our participants. When controlling for all mediating factors in the same model, the gender gap in leadership motivation decreases by 77% and is no longer significant.⁵⁰

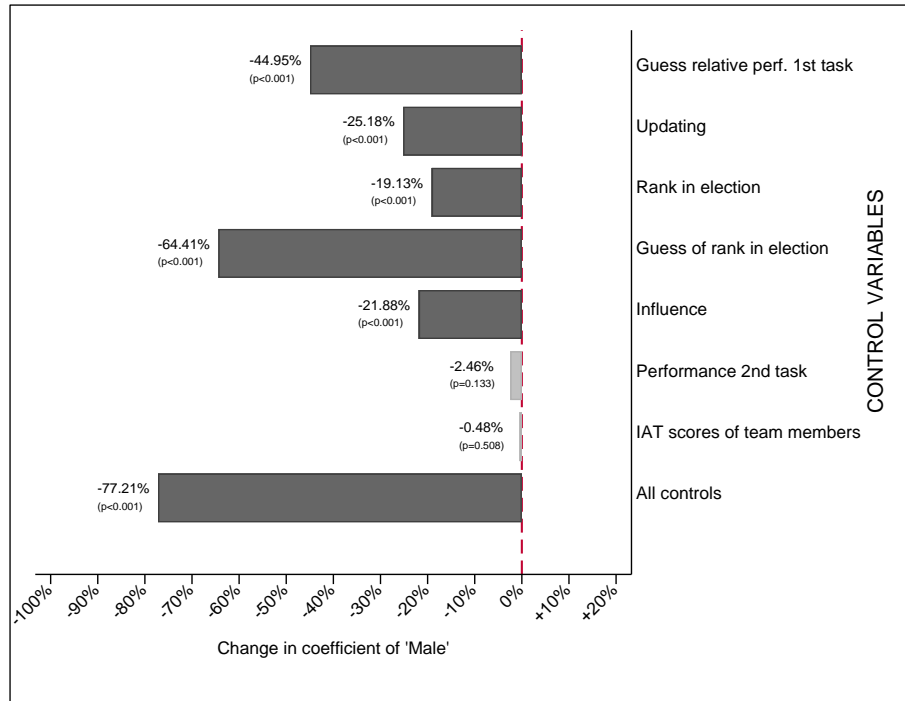
Panel (b) in Figure 6 illustrates the influence of the mediating factors on the effect of team gender composition on women's willingness to lead. This analysis is based on the regression in column 4 of Table 2, where willingness to lead is regressed on gender, team gender composition, an interaction between the two, and relative performance. The bars illustrate how the estimated impact of team gender composition on women's willingness to lead changes after controlling for each of the mediating variables (i.e., how the coefficient of

⁴⁹ The results of these regressions are reported in panel (a) of Table A2. The p-values presented in Figure 6 are obtained using STATA's *suest* command.

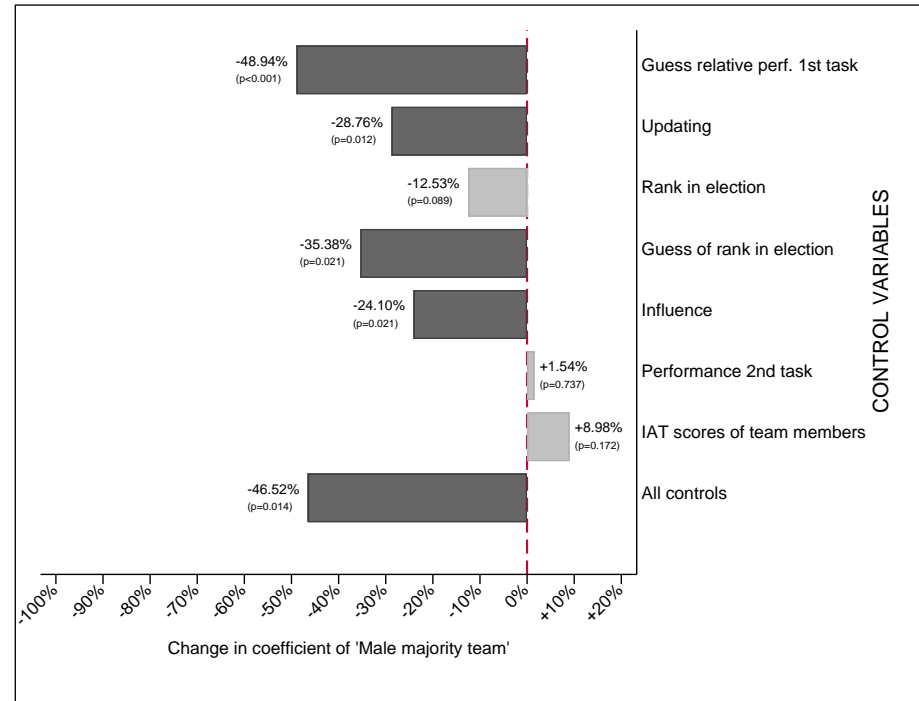
⁵⁰ Correlations between the mediating variables are reported in Table A4.

Figure 6: Changes in the estimated gender and treatment effects associated with additional controls

(a) Gender gap in willingness to lead



(b) Effect of team gender composition on women



Note: Panel (a) shows the change in in the coefficient of *Male* from column 2 of Table 2, when controlling for the variables indicated on the y-axis. Panel (b) shows the change in the absolute value of the coefficient of Male majority team from column 4 of Table 2, when controlling for the variable indicated on the y-axis (in panel (b), we also control for the variable indicated on the y-axis interacted with *Male*). The underlying regressions are presented in Table A.2 in the Appendix. The p-values are obtained from Chi2-tests using STATA's *suest* command, and the dark grey bars indicate statistically significant ($p < 0.05$) changes.

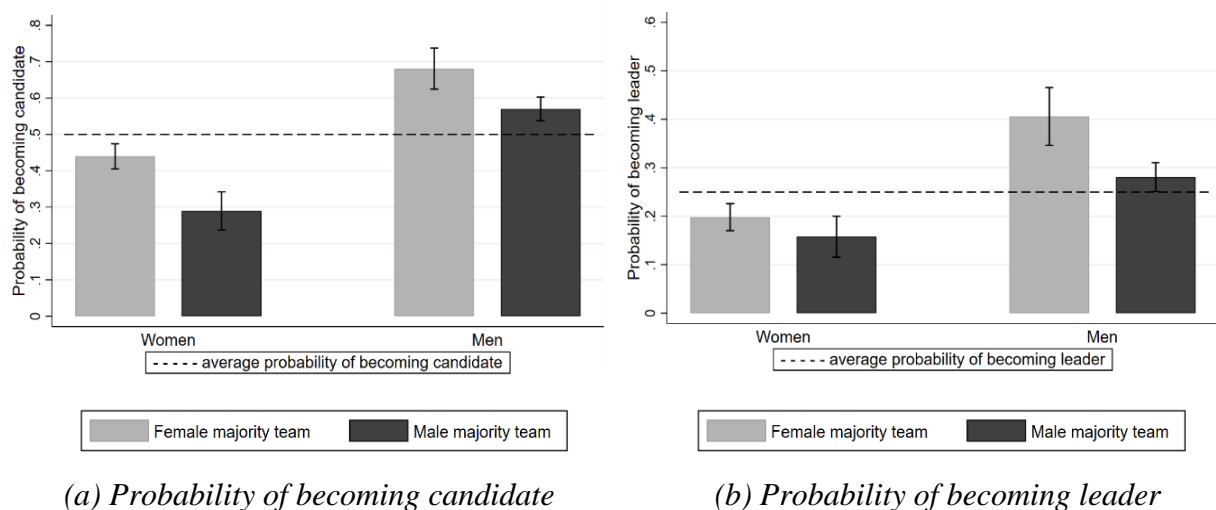
the dummy “*Male majority team*” changes).⁵¹ As for the general gender gap in willingness to lead, the impact of team gender composition on women’s willingness to lead decreases when controlling for most of the mediating variables, and the decrease is largest for relative performance beliefs (49 %) and beliefs about electoral outcomes (35%).

6. Outcomes

While 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 impact team outcomes.

Figure 7 shows the average, individual, probability of becoming a candidate (Panel a), and of becoming the leader (Panel b), by gender and treatment (see Table A.3 for regression analyses). In both types of teams, women’s probability of becoming a 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 to become a candidate (57.0% vs. 28.9%), and 1.8 times more likely to become the leader (28.1% vs. 15.8%). As illustrated in Table A.3, these numbers do not change importantly if we control for relative performance.

Figure 7: Probability of becoming candidate and leader, by gender and treatment

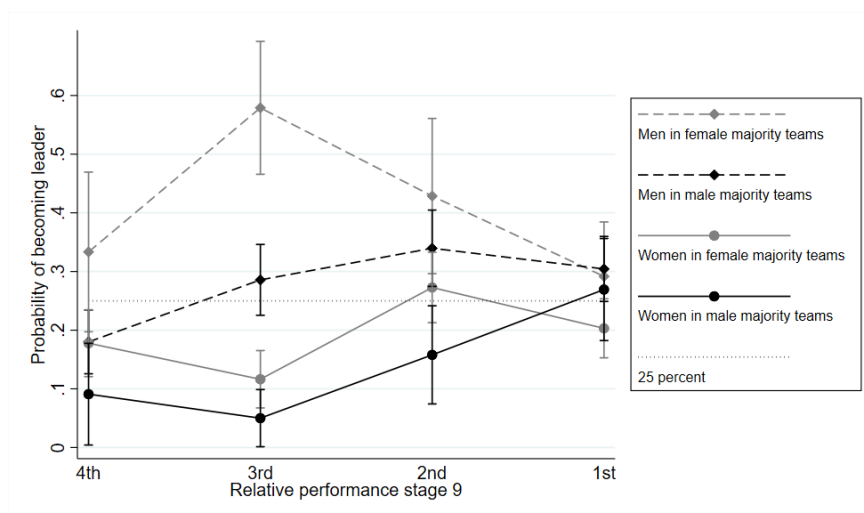


Note: Error bars represent standard errors.

⁵¹ The regression coefficients corresponding to panel (b) of Figure 6 are presented in panel (b) of Table A.2 in the Appendix. In each regression, we interact the measure of relative performance, and each control variable, with *Male*. Thus, we allow for gender-specific effects of these variables on the outcome variable.

In Stage 9, when the leaders worked on the team answer for the Desert Survival task, we asked all non-leaders to perform the same task as the leader, albeit unincentivized. That is, we asked them to provide the answer they would have provided if they had been the team leader. We use this information to obtain a measure of leadership performance for all team members. Based on this measure, only 33% of the teams select the most able leader.⁵² While male majority teams are slightly more successful than female majority teams in choosing the better performing leader (37% vs. 29%), this difference is not statistically significant ($\chi^2=1.01$, $p=0.315$). Figure 8 presents the probability of becoming the team leader by gender, treatment and relative performance in Stage 9. As illustrated by this figure, and Figure A.6 in the Appendix, one reason behind the tendency for male majority teams to elect slightly higher performing leaders appears to be that female majority teams elect low-performing men too often. While this analysis is based on rather small cells, it is noticeable that, in female majority teams, 58% of the men who perform third best, and 33% of the men who perform the worst, become the team leader. The corresponding numbers for women in male majority teams are 5% and 9%.⁵³

Figure 8: Probability of becoming the leader, by relative performance in Stage 9, gender, and treatment



Note: Error bars represent standard errors.

⁵² We chose to present the outcomes based on participant performance in Stage 9, because this is the task where the leader's performance determines the team's payoff. However, it may also be relevant to base this analysis on performance in Stage 1 or Stage 8. 32% and 34% of the teams select the most able leader in terms of individual performance in Stage 1 and Stage 8 respectively. There is no significant difference between female and male majority teams in any of the 3 cases.

⁵³ Given the small number of observations, some comparisons between teams differ when the numbers presented in Figure 8 are based on participants' performance in Stage 1 or Stage 8 instead of Stage 9. However, the main take-away remains.

How much money do participants lose from not choosing the best performing team member as the leader? On average, summing over all 145 teams, the difference in Stage 9 leadership performance between the leader and the best performing team member amounts to 4.2 penalty points, which translates to a loss of 2.1 Swiss Francs for each team member.⁵⁴ In female majority teams, these numbers are very similar for female and male leaders (4.2 vs. 4.0 penalty points, $p=0.945$). However, in male majority teams, the loss is on average smaller for the teams with a female leader than for those with a male leader (1.7 vs. 4.8 penalty points, $p=0.035$). One contributing reason for this is that, in male majority teams, female leaders are almost twice as likely as male leaders to be the best performer in the team (58% vs. 33%, $\chi^2=2.8$, $p=0.093$).

7. Conclusion

In this study, we identify the causal effect of team gender composition on women's willingness to lead through an economic experiment with random assignment of participants to either male or female majority teams. Consistent with previous studies on gender gaps in willingness to lead, we find that female participants are significantly less willing to become the team leader than male participants. 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 lead teams with a male majority than teams with a female majority. An analysis of mechanisms indicates that women in male majority teams are less confident in their own task performance, believe that they will receive less support in the subsequent election, are less influential, and face teams that are more gender biased than women in female majority teams. A substantial share of the effect of team gender composition on women's willingness to lead is due to the negative effect of male majority teams on women's confidence and anticipated electoral outcomes.

Despite being more willing to lead, women assigned to female majority teams are not significantly 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. We thus find that being in numerical minority

⁵⁴ If restricting the sample to the 97 teams who did not choose the best performer as their leader, the corresponding numbers are 6.3 penalty points and 3.2 Swiss Francs.

has a very different effect on men compared to women. Men in female majority teams fare better than all other groups in many regards, while women in male majority teams fare the worst. Hence, our findings point to the existence of a negative “glass ceiling” effect for women in numerical minority, and a positive “glass escalator” effect for men in numerical minority (Williams 1992).

Our results illustrate how gender stereotypes translate into gender gaps in outcomes and decision-making power. In particular, they highlight that women in male majority teams face both *external constraints*, in terms of, for example, lower influence and increased gender bias, and *internal constraints*, in terms of, for example, lower confidence in their relative ability to perform the task at hand. These external and internal constraints are likely to interact and work together, creating a challenging environment for women in male dominated contexts.

While our primary outcome variable is willingness to lead, speaking directly to the vertical gender segregation, we believe that this study highlights mechanisms that may be important determinants of the horizontal gender segregation as well. Today, most part of the raw gender wage gap can be attributed to men and women choosing different educations, industries and occupations (Blau and Kahn 2017). Thus, it is important to understand these large and persistent gender gaps in educational and occupational choices. We find that—after only ten minutes of team interaction—women randomly assigned to male majority teams are less confident and less willing to take on a leadership role. This difference in confidence and leadership motivation is likely to correlate with other types of behaviors and feelings that matter for women’s outcomes and well-being in the long run. Thus, our findings are in line with the conjecture that educational and occupational segregation by gender may be partly caused, and upheld, by an adverse effect of male majority environments on women’s outcomes and well-being.

Our results also have implications for the current debate about gender quotas. While both women and men are less confident and motivated to lead in teams with many men, male majority environments appear particularly negative for female candidates. This indicates that quotas may not only benefit the women directly impacted by them, but also any existing women in minority positions.

Further, our results contribute to research on how to structure team work, and how to attract, retain, and identify the most competent individuals to top positions. In our setting, point estimates indicate that the least able men are more willing to become the leader, and receive more votes, than the most able women.

Finally, a number of studies suggest that gender stereotypes and image concerns may cause women to shy away from career enhancing activities (see, for example, Bursztyn, Fujiwara, and Pallais 2017), and that, as a consequence, educational institutions and workplaces should make informed choices about what information regarding individuals' actions to make public. Our study instead indicates that the large majority of the gender gap in willingness to take on a leading role is driven by male over- and female under-confidence in task related ability, and lack of support for potential female leaders. This underlines the importance of objective and transparent performance measures and promotion criteria in organizations. One open question for future research is how organizational institutions may best balance these different needs and contribute to equal opportunity and the efficient choice of leaders. Another question is the extent to which mechanisms similar to those explored in this paper may help explain long lasting segregation based on, for example, race or religion.

DEPARTMENT OF ECONOMICS, STOCKHOLM SCHOOL OF ECONOMICS
DEPARTMENT OF ECONOMICS, UNIVERSITY OF GOTHENBURG
SWEDISH INSTITUTE FOR SOCIAL RESEARCH, STOCKHOLM UNIVERSITY

References

- Anelli, Massimo, and Giovanni Peri. 2017. "The Effects of High School Peers' Gender on College Major, College Performance and Income." *The Economic Journal*. Accessed August 17, 2018. <https://doi.org/10.1111/eoj.12556>.
- Anil, Bulent, Duygu Guner, Tuba Delibasi, and Gokce Uysal. 2016. "Does Classroom Gender Composition Affect School Dropout?" SSRN Scholarly Paper ID 2846329. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2846329>.
- Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2009. "The Effects of Gender Interactions in the Lab and in the Field." *The Review of Economics and Statistics* 91 (1): 152–62.
- Apestequia, Jose, Ghazala Azmat, and Nagore Iriberry. 2011. "The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field." *Management Science* 58 (1): 78–93. <https://doi.org/10.1287/mnsc.1110.1348>.
- Babcock, Linda, Maria P. Recalde, and Lise Vesterlund. 2017. "Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash." *American Economic Review* 107 (5): 131–35. <https://doi.org/10.1257/aer.p20171018>.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107 (4): 1207–38. <https://doi.org/10.1257/aer.20151211>.
- Barber, Brad M., and Terrance Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics* 116 (1): 261–92. <https://doi.org/10.1162/003355301556400>.
- Baskaran, Thushyanthan, and Zohal Hessami. 2018. "Does the Election of a Female Leader Clear the Way for More Women in Politics?" *American Economic Journal: Economic Policy* 10 (3): 95–121. <https://doi.org/10.1257/pol.20170045>.
- Belfi, Barbara, Mieke Goos, Bieke De Fraine, and Jan Van Damme. 2012. "The Effect of Class Composition by Gender and Ability on Secondary School Students' School Well-Being and Academic Self-Concept: A Literature Review." *Educational Research Review* 7 (1): 62–74. <https://doi.org/10.1016/j.edurev.2011.09.002>.
- Bertrand, Marianne. 2011. "New Perspectives on Gender" 4B: 1543–90.
- Bertrand, Marianne, and Kevin F Hallock. 2001. "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review* 55 (1): 3–21. <https://doi.org/10.2307/2696183>.
- Bettinger, Eric, P., and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review*, 95 (2): 152-157.
- Bettio, Francesca. and Alina Verashchagina. 2009, Gender Segregation in the Labour Market: Root Causes, Implications and Policy Responses in the EU , Luxembourg: Publications Office of the European Union.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2013. "Under Pressure? The Effect of Peers on Outcomes of Young Adults." *Journal of Labor Economics* 31 (1): 119–53.
- Blau, Francine D. Peter Brummund, and Albert Liu. 2013. "Trends in Occupational Segregation by Gender 1970–2009: Adjusting for the Impact of Changes in the Occupational Coding System." *Demography* 50 (2): 471–92.
- Blau, Francine D., Janet M. Currie, Rachel T. A. Croson and Donna K. Ginther, 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial," *American Economic Review*, 100(2): 348-52.

- Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55 (3): 789–865. <https://doi.org/10.1257/jel.20160995>.
- Booth, Alison, Lina Cardona-Sosa, and Patrick Nolen. 2014. "Gender Differences in Risk Aversion: Do Single-Sex Environments Affect Their Development?" *Journal of Economic Behavior & Organization* 99 (March): 126–54. <https://doi.org/10.1016/j.jebo.2013.12.017>.
- Booth, Alison L., and Patrick Nolen. 2012. "Gender Differences in Risk Behaviour: Does Nurture Matter?" *Economic Journal* 122 (558). <http://ideas.repec.org/a/ecj/econjl/v122y2012i558pf56-f78.html>.
- Booth, Alison, and Patrick Nolen. 2012. "Choosing to Compete: How Different Are Girls and Boys?" *Journal of Economic Behavior & Organization* 81 (2): 542–55. <https://doi.org/10.1016/j.jebo.2011.07.018>.
- Booth, Alison, and Eiji Yamamura. 2017. "Performance in Mixed-Sex and Single-Sex Competitions: What We Can Learn from Speedboat Races in Japan." *The Review of Economics and Statistics*, October. https://doi.org/10.1162/rest_a_00715.
- Bowles, Hannah Riley, Linda Babcock, and Lei Lai. 2007. "Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes It Does Hurt to Ask." *Organizational Behavior and Human Decision Processes* 103 (1): 84–103. <https://doi.org/10.1016/j.obhdp.2006.09.001>.
- Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145 (January): 27–41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>.
- Burow, Norma, Miriam Beblo, Denis Beninger, and Melanie Schröder. 2017. "Why Do Women Favor Same-Gender Competition? Evidence from a Choice Experiment." SSRN Scholarly Paper ID 2960062. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2960062>.
- Bursztyjn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–3319. <https://doi.org/10.1257/aer.20170029>.
- Buser, Thomas, Eva Ranehill, and Roel van Veldhuizen. 2017. "Gender Differences in Willingness to Compete: The Role of Public Observability." University of Zurich, Department of Economics, Working Paper No. 257. Available at SSRN: <https://ssrn.com/abstract=2997071> or <http://dx.doi.org/10.2139/ssrn.2997071>
- Butterfield, D. Anthony, and Gary N. Powell. 2003. "Gender, Gender Identity, and Aspirations to Top Management." *Women in Management Review* 18 (1/2): 88–96. <https://doi.org/10.1108/09649420310462361>.
- Carrell, Scott E., Marianne E. Page and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, MIT Press, vol. 125(3): 1101-1144. Casadevall, Arturo. 2015. "Achieving Speaker Gender Equity at the American Society for Microbiology General Meeting." *MBio* 6 (4): e01146-15. <https://doi.org/10.1128/mBio.01146-15>.
- Catalyst, *Pyramid: Women in S&P 500 Companies* (October 3, 2018).
- Chakraborty, Priyanka and Danila Serra. 2018. "Gender differences in top leadership roles: Does aversion to worker backlash matter?" Departmental Working Papers 1807, Southern Methodist University, Department of Economics.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–43. <https://doi.org/10.1111/j.1468-0262.2004.00539.x>.

- Chen, Jingnan, and Daniel Houser. 2017. "Gender Composition, Stereotype and the Contribution of Ideas." SSRN Scholarly Paper ID 2989049. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2989049>.
- Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *The Quarterly Journal of Economics* 129 (4): 1625–60. <https://doi.org/10.1093/qje/qju023>.
- Cohen, Philip N. 2013. "The Persistence of Workplace Gender Segregation in the US: Gender Segregation at Work." *Sociology Compass* 7 (11): 889–99. <https://doi.org/10.1111/soc4.12083>.
- Dasgupta, Nilanjana, Melissa McManus Scircle, and Matthew Hunsinger. 2015. "Female Peers in Small Work Groups Enhance Women's Motivation, Verbal Participation, and Career Aspirations in Engineering." *Proceedings of the National Academy of Sciences* 112 (16): 4988–93. <https://doi.org/10.1073/pnas.1422822112>.
- Eagly, Alice H., and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109 (3): 573–98. <https://doi.org/10.1037//0033-295X.109.3.573>.
- Eagly, Alice H., Mona G. Makhijani, and Bruce G. Klonsky. 1992. "Gender and the Evaluation of Leaders: A Meta-Analysis." *Psychological Bulletin* 111 (1): 3–22. <https://doi.org/10.1037/0033-2909.111.1.3>.
- Eckel, Catherine C., and Sascha C. Füllbrunn. 2015. "Thar SHE Blows? Gender, Competition, and Bubbles in Experimental Asset Markets." *American Economic Review* 105 (2): 906–20. <https://doi.org/10.1257/aer.20130683>.
- England, Paula, Paul Allison, Su Li, Noah Mark, Jennifer Thompson, Michelle J. Budig, and Han Sun. 2007. "Why Are Some Academic Fields Tipping toward Female? The Sex Composition of U.S. Fields of Doctoral Degree Receipt, 1971-2002." *Sociology of Education* 80 (1): 23–42.
- Erkal, Nisvan, Lata Gangadharan, and Erte Xiao. 2018. "Leadership Selection: Can Changing the Default Break the Glass Ceiling?" SSRN Scholarly Paper ID 3022386. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3022386>.
- Exley, Christine L, Muriel Niederle, and Lise Vesterlund. 2016. "Knowing When to Ask: The Cost of Leaning In." Working Paper 22961. National Bureau of Economic Research. <https://doi.org/10.3386/w22961>.
- Fox, Richard L., and Jennifer L. Lawless. 2004. "Entering the Arena? Gender and the Decision to Run for Office." *American Journal of Political Science* 48 (2): 264–80. <https://doi.org/10.2307/1519882>.
- . 2014. "Uncovering the Origins of the Gender Gap in Political Ambition." *American Political Science Review* 108 (03): 499–519. <https://doi.org/10.1017/S0003055414000227>.
- Gagliarducci, Stefano, and M. Daniele Paserman. 2012. "Gender Interactions within Hierarchies: Evidence from the Political Arena." *Review of Economic Studies* 79 (3): 1021–52.
- Gloor, Jamie L. and Morf, Manuela C. and Paustian-Underdahl, Samantha and Backes-Gellner, Uschi, Fix the Game–Not the Dame: A Team Intervention to Restore Gender Equity in Leadership Evaluations (January 1, 2017). University of Zurich, Institute of Business Administration, UZH Business Working Paper No. 375. Available at SSRN: <https://ssrn.com/abstract=3165925> or <http://dx.doi.org/10.2139/ssrn.3165925>
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118 (3): 1049–74. <https://doi.org/10.1162/00335530360698496>.

- Gobillon, Laurent, Dominique Meurs, and Sébastien Roux. 2015. "Estimating Gender Differences in Access to Jobs." *Journal of Labor Economics* 33 (2): 317–63. <https://doi.org/10.1086/678495>.
- Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90 (4): 715–41. <https://doi.org/10.1257/aer.90.4.715>.
- Griffith, Eric E., and Nilanjana Dasgupta. 2018. "How the Demographic Composition of Academic Science and Engineering Departments Influences Workplace Culture, Faculty Experience, and Retention Risk." *Social Sciences* 7 (5): 71. <https://doi.org/10.3390/socsci7050071>.
- Grunspan, Daniel Z., Sarah L. Eddy, Sara E. Brownell, Benjamin L. Wiggins, Alison J. Crowe, and Steven M. Goodreau. 2016. "Males Under-Estimate Academic Performance of Their Female Peers in Undergraduate Biology Classrooms." *PLOS ONE* 11 (2): e0148405. <https://doi.org/10.1371/journal.pone.0148405>.
- Gumpertz, Marcia, Raifu Durodoye, Emily Griffith, and Alyson Wilson. 2017. "Retention and Promotion of Women and Underrepresented Minority Faculty in Science and Engineering at Four Large Land Grant Institutions." *PloS One* 12 (11): e0187285. <https://doi.org/10.1371/journal.pone.0187285>.
- Hederos Eriksson, Karin, and Anna Sandberg. 2012. "Gender Differences in Initiation of Negotiation: Does the Gender of the Negotiation Counterpart Matter?" *Negotiation Journal* 28 (4): 407–28. <https://doi.org/10.1111/j.1571-9979.2012.00349.x>.
- Hill, Andrew J. 2017. "The Positive Influence of Female College Students on Their Male Peers." *Labour Economics* 44 (January): 151–60. <https://doi.org/10.1016/j.labeco.2017.01.005>.
- Hoffmann, Florian and Philip Oreopoulos. 2009. "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources*, 44(2): 479-494.
- Hogarth, Robin M., Natalia Karelaia, and Carlos Andrés Trujillo. 2012. "When Should I Quit? Gender Differences in Exiting Competitions." *Journal of Economic Behavior & Organization*, Gender Differences in Risk Aversion and Competition, 83 (1): 136–50. <https://doi.org/10.1016/j.jebo.2011.06.021>.
- Hoogendoorn, Sander, Hessel Oosterbeek, and Mirjam van Praag. 2013. "The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment." *Management Science* 59 (7): 1514–28. <https://doi.org/10.1287/mnsc.1120.1674>.
- Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." Working Paper 7867. National Bureau of Economic Research. <http://www.nber.org/papers/w7867>.
- Hoyt, Crystal, and Susan Murphy. 2016. "Managing to Clear the Air: Stereotype Threat, Women, and Leadership." *The Leadership Quarterly*, January, 387–99. <https://doi.org/10.1016/j.leaqua.2015.11.002>.
- Hunt, Jennifer. 2016. "Why Do Women Leave Science and Engineering? , Why Do Women Leave Science and Engineering?" *ILR Review* 69 (1): 199–226. <https://doi.org/10.1177/0019793915594597>.
- Inzlicht, Michael, and Talia Ben-Zeev. 2000. "A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males." *Psychological Science* 11 (5): 365–71.
- Kanter, Rosabeth Moss. 1977a. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82 (5): 965–90.

- . 1977b. “Men and Women of the Corporation,” January.
<https://www.hbs.edu/faculty/Pages/item.aspx?num=10807>.
- Kanthak, Kristin, and Jonathan Woon. 2015. “Women Don’t Run? Election Aversion and Candidate Entry.” *American Journal of Political Science* 59 (3): 595–612.
<https://doi.org/10.1111/ajps.12158>.
- Karpowitz, C. and Mendelberg, T.: *The Silent Sex: Gender, Deliberation, and Institutions*. (EBook, Paperback and Hardcover).” 2014.
<http://press.princeton.edu/titles/10402.html>.
- Koenig, Anne M., Alice H. Eagly, Abigail A. Mitchell, and Tiina Ristikari. 2011. “Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms.” *Psychological Bulletin* 137 (4): 616–42. <https://doi.org/10.1037/a0023557>.
- Konrad, A.M., V. Kramer, and S. Erkut. 2008. "Critical mass: The impact of three or more women on corporate boards." *Organizational Dynamics*, 37, pp. 145-164.
- Kuhnen, Camelia M., and Agnieszka Tymula. 2011. “Feedback, Self-Esteem, and Performance in Organizations.” *Management Science* 58 (1): 94–113.
<https://doi.org/10.1287/mnsc.1110.1379>.
- Kunze, Astrid, and Amalia R. Miller. 2017. “Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies.” *The Review of Economics and Statistics* 99 (5): 769–75. https://doi.org/10.1162/REST_a_00668.
- Lavy, Victor, and Analia Schlosser. 2011. “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics* 3 (2): 1–33.
<https://doi.org/10.1257/app.3.2.1>.
- Lawless, Jennifer, and Richard Fox. 2012. *Men rule: The continued under-representation of women in U.S. politics*. Washington, DC: Women & Politics Institute.
- Litzky, Barrie, and Jeffrey Greenhaus. 2007. “The Relationship between Gender and Aspirations to Senior Management.” *Career Development International* 12 (7): 637–59. <https://doi.org/10.1108/13620430710834404>.
- Lordan, Grace, and Jörn-Steffen Pischke. 2016. “Does Rosie Like Riveting? Male and Female Occupational Choices.” 10129. IZA Discussion Papers. Institute for the Study of Labor (IZA). <https://ideas.repec.org/p/iza/izadps/dp10129.html>.
- Lundeberg, Mary, Paul W. Fox, and Judith Punčohář. 1994. “Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments.” *Journal of Educational Psychology* 86 (March): 114–21. <https://doi.org/10.1037/0022-0663.86.1.114>.
- Marschke, Robyn, Sandra Laursen, Joyce McCarl Nielsen, and Patricia Rankin. 2007. “Demographic Inertia Revisited: An Immodest Proposal to Achieve Equitable Gender Representation among Faculty in Higher Education.” *The Journal of Higher Education* 78 (1): 1–26.
- Matsa, David A, and Amalia R Miller. 2013. “A Female Style in Corporate Leadership? Evidence from Quotas.” *American Economic Journal: Applied Economics* 5 (3): 136–69. <https://doi.org/10.1257/app.5.3.136>.
- Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2018. “Gender Bias in Teaching Evaluations.” *Journal of the European Economic Association*.
<https://doi.org/10.1093/jeea/jvx057>.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. “Science Faculty’s Subtle Gender Biases Favor Male Students.” *Proceedings of the National Academy of Sciences* 109 (41): 16474–79.
<https://doi.org/10.1073/pnas.1211286109>.

- Catalyst. 2012. "Women and Men in U.S. Corporate Leadership: Same Workplace, Different Realities?" Retrieved from <https://www.catalyst.org/knowledge/women-and-men-us-corporate-leadership-same-workplace-different-realities>. October 17, 2012.
- Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122 (3): 1067–1101.
- Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics* 8 (1): 405–34. <https://doi.org/10.1146/annurev-economics-080614-115329>.
- Oosterbeek, Hessel, and Reyn van Ewijk. 2014. "Gender Peer Effects in University: Evidence from a Randomized Experiment." *Economics of Education Review* 38 (February): 51–63. <https://doi.org/10.1016/j.econedurev.2013.11.002>.
- Pan, Jessica. 2015. "Gender Segregation in Occupations: The Role of Tipping and Social Interactions." *Journal of Labor Economics* 33 (2): 000–000. <https://doi.org/10.1086/678518>.
- Preece, Jessica, and Olga Stoddard. 2015. "Why Women Don't Run: Experimental Evidence on Gender Differences in Political Competition Aversion." *Journal of Economic Behavior & Organization* 117 (C): 296–308.
- Porter, Catherine and Danila Serra. 2017. "Gender differences in the choice of major: The importance of female role models," Departmental Working Papers 1705, Southern Methodist University, Department of Economics.
- Ranehill, Eva, and Roberto A. Weber. 2017. "Do Gender Preference Gaps Impact Policy Outcomes?" SSRN Scholarly Paper ID 3071682. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3071682>.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences* 111 (12): 4403–8. <https://doi.org/10.1073/pnas.1314788111>.
- Rossi, Fabrizio, Chengru Hu, and Maggie Foley. 2017. "Women in the boardroom and corporate decisions of Italian listed companies: Does the "critical mass" matter?", *Management Decision*, 55 (7): 1578-1595, <https://doi.org/10.1108/MD-01-2017-0029>
- Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107 (5): 141–45. <https://doi.org/10.1257/aer.p20171126>.
- Shurchkov, Olga, and Alexandra van Geen. 2017. "Why Female Decision-Makers Shy Away From Promoting Competition." SSRN Scholarly Paper ID 3155259. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3155259>.
- Sjögren Lindquist, Gabriella, and Jenny Säve-Söderbergh. 2011. "'Girls Will Be Girls', Especially among Boys: Risk-Taking in the 'Daily Double' on Jeopardy." *Economics Letters* 112 (2): 158–60.
- Sousa, José de, and Guillaume Hollard. 2015. "Gender Differences: Evidence from Field Tournaments." 1506. CEPREMAP Working Papers (Docweb). CEPREMAP. <https://ideas.repec.org/p/cpm/docweb/1506.html>.
- Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35 (1): 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Schwartz-Ziv, M. 2017. "Gender and Board Activeness: The Role of a Critical Mass." *Journal of Financial and Quantitative Analysis*, 52 (2): 751-780. doi:10.1017/S0022109017000059
- Thomas, Nicole R., Daniel J. Poole, and Joan M. Herbers. 2015. "Gender in Science and Engineering Faculties: Demographic Inertia Revisited." *PLOS ONE* 10 (10): e0139767. <https://doi.org/10.1371/journal.pone.0139767>.

- Thomas, Rachel, Marianne Cooper, Ellen Konar, Megan Rooney, Ashley Finch, Lareina Yee, Alexis Krivkovich, Irina Starikova, Kesley Robinson, and Rachel Valentino. 2017. *Women in the Workplace*, McKinsey&Company, Retrieved from <https://womenintheworkplace.com/> 12-10-2018.
- Thomas-Hunt, Melissa C., and Katherine W. Phillips. 2004. "When What You Know Is Not Enough: Expertise and Gender Dynamics in Task Groups, When What You Know Is Not Enough: Expertise and Gender Dynamics in Task Groups." *Personality and Social Psychology Bulletin* 30 (12): 1585–98.
<https://doi.org/10.1177/0146167204271186>.
- Torchia, M.; A. Calabro; and M. Huse. 2011. "Women Directors on Corporate Boards: From Tokenism to Critical Mass." *Journal of Business Ethics*, 102 (2): 299–317.
- Usui, Emiko. 2008. "Job Satisfaction and the Gender Composition of Jobs." *Economics Letters* 99 (1): 23–26.
- Zölitz, Ulf, and Jan Feld. 2017. "The Effect of Peer Gender on Major Choice." SSRN Scholarly Paper ID 3071681. Rochester, NY: Social Science Research Network.
<https://papers.ssrn.com/abstract=3071681>.

Appendix A: Additional figures and tables

Figure A.1: Distribution of male and female performance in 1st task (Stage 1)

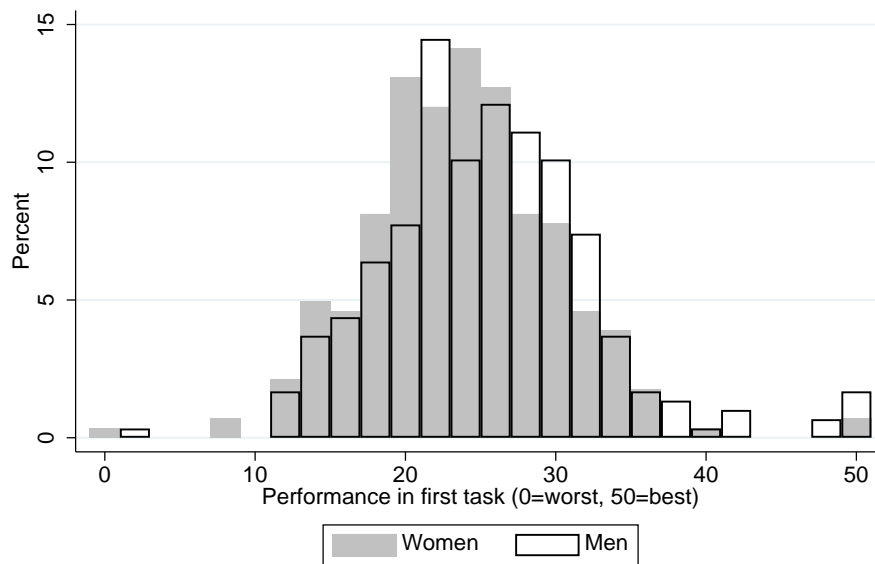


Figure A.2: Within-team rank in willingness to lead, by gender and treatment

(a) *Women*

(b) *Men*

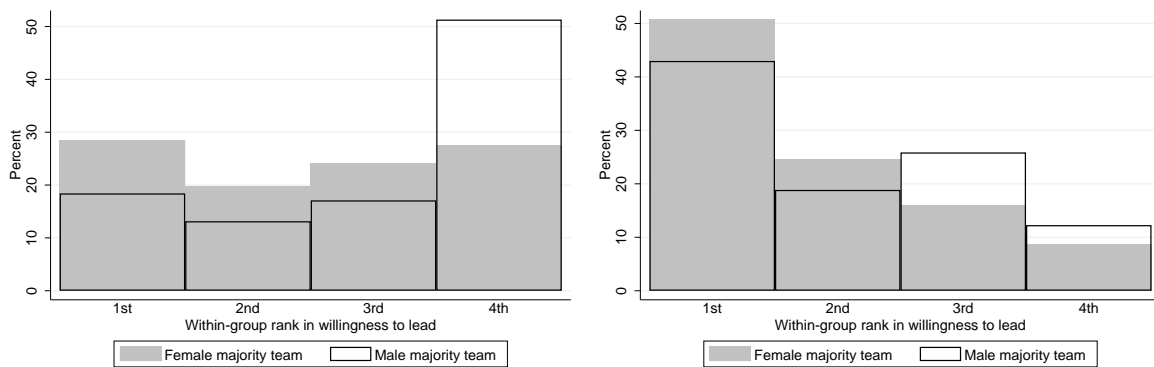


Figure A.3: Distribution of leadership aspiration by gender and treatment

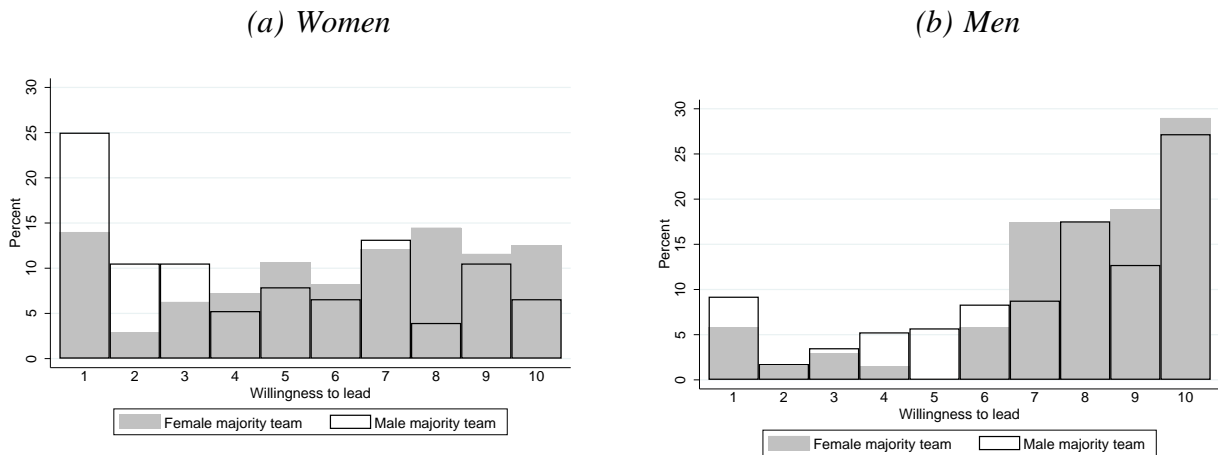
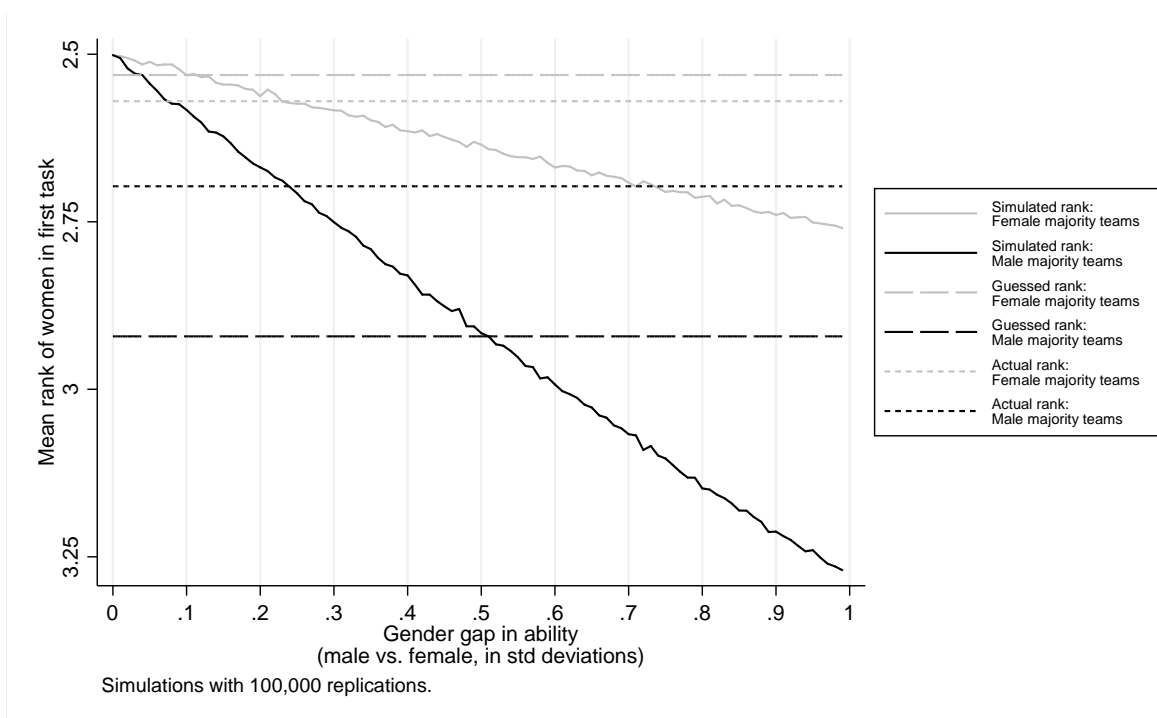
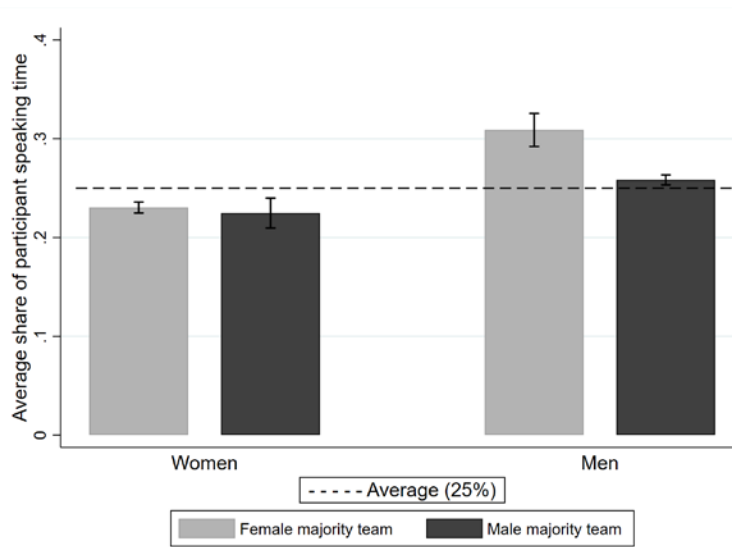


Figure A.4: Simulations of women’s relative performance rank as a function of gender gaps in ability



Note: The solid lines show the outcomes from a simulation exercise, illustrating how different sizes of the gender gap in task ability would translate into different average within-team performance rankings of female participants in male vs. female majority teams. In these simulations, we let task ability be normally distributed with a standard deviation of 1. The mean ability of women is 0, and the mean ability of men is α , ranging from 0 to 1. Thus, shown on the x-axis, α gives us the average gender gap in ability, expressed in standard deviations. The long dashed lines show the *average guessed rank* of female participants in male vs. female majority teams, and the short dashed lines show the *average actual rank* of female participants in male vs. female majority teams (ties are given the average rank value).

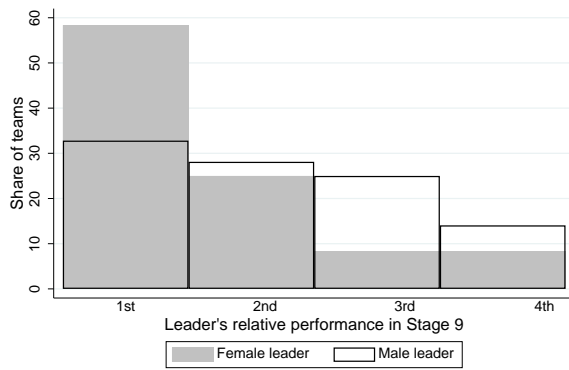
Figure A.5: Distribution of average share of total speaking time by gender and treatment



Note: Error bars represent standard errors.

Figure A.6: Leader's relative performance in Stage 9

(a) Male majority teams



(b) Female majority teams

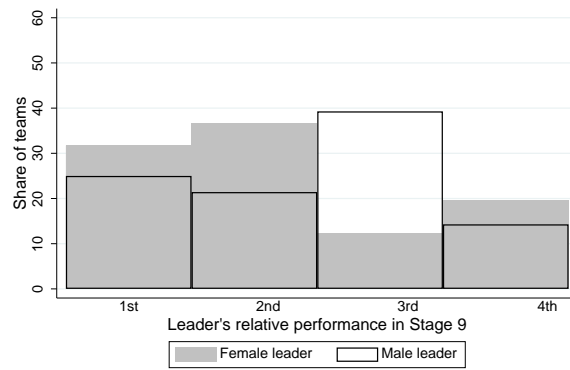


Table A.1: Description of mediating variables

Mediating Variable	Description
<i>Guess of relative performance in 1st task</i>	The participant's incentivized guess of their performance ranking, elicited in Stage 4 of the experiment. Discrete variable ranging from 1 (worst) to 4 (best).
<i>Updating</i>	Sum (over all 10 items) of absolute difference in rank between the participant's original answer (from Stage 1) and their updated answer (from Stage 3), divided by the sum of all team members' differences. Continuous measure ranging from 0 (the participant did not update at all) to 1 (the participant was the only one in their team updating).
<i>Rank in election</i>	The participant's rank in the election. Discrete variable ranging from 1 (the participant was ranked last) and 4 (the participant was ranked first).
<i>Guess of rank election</i>	The participant's incentivized guess of their rank in the election, elicited in Stage 7 of the experiment. Discrete variable ranging from 1 (the participant was ranked last) and 4 (the participant was ranked first).
<i>Influence</i>	Sum (over all 10 items) of absolute difference in rank between the participant's individual answer (from Stage 1) and their team's joint answer, divided by the sum of all team members' differences. Continuous measure ranging from 0 (the participant had no influence in the team's solution) to 1 (the team's solution was identical to the participant's individual solution).
<i>Speaking time</i>	The total (average) speaking time of the female team member(s) in the team discussion divided by the total speaking time of all team members. Continuous measure ranging from 0 (no woman spoke in the team discussion) to 1 (only women spoke in the team discussion).
<i>Performance in 2nd task</i>	Individual performance in the second task. Discrete variable ranging from 0 (worst possible solution) to 50 (best possible solution)
<i>IAT of other team members</i>	Average IAT score among team members, excluding the participant. Continuous variable ranging from -2 to 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.

Table A.2: Gender gap and treatment effects, controlling for mediating variables

<i>Dependent variable: Willingness to lead (1-10)</i>									
Control variable(s):	-	Guess rel. perf. 1 st task	Updating	Rank election	Guess rank election	Influence	Performance 2nd task	IAT scores of team members	All
(a) Overall gender gap									
Male	1.584*** (0.262)	0.872*** (0.242)	1.185*** (0.252)	1.281*** (0.259)	0.564* (0.231)	1.237*** (0.256)	1.545*** (0.262)	1.576*** (0.261)	0.361 (0.220)
Constant	6.303*** (0.345)	9.755*** (0.418)	7.924*** (0.360)	4.442*** (0.447)	10.452*** (0.404)	-3.550** (1.172)	7.179*** (0.478)	6.192*** (0.380)	10.076*** (2.121)
<i>N</i>	580	580	580	580	580	580	580	580	580
(b) Effect of team gender composition									
Male	1.417* (0.600)	-0.131 (0.704)	0.185 (0.635)	1.568 (0.911)	-0.277 (0.654)	4.796* (2.216)	0.385 (0.911)	1.534* (0.664)	-2.077 (4.210)
Male majority team	-1.355*** (0.402)	-0.692 (0.379)	-0.966* (0.392)	-1.186** (0.379)	-0.876* (0.353)	-1.029* (0.394)	-1.376*** (0.390)	-1.477*** (0.404)	-0.725* (0.342)
Male X Male majority team	0.706 (0.588)	0.346 (0.520)	0.486 (0.568)	0.745 (0.555)	0.997* (0.490)	0.533 (0.574)	0.727 (0.577)	0.779 (0.573)	0.862 (0.464)
Constant	6.792*** (0.465)	10.458*** (0.578)	8.773*** (0.496)	4.715*** (0.648)	10.929*** (0.505)	-4.830** (1.741)	8.232*** (0.674)	6.561*** (0.496)	11.594*** (3.416)
<i>N</i>	580	580	580	580	580	580	580	580	580
F-test:									
'Male majority team' + 'Male X Male majority team'	-0.650 (F=3.380)	-0.347 (F=1.035)	-0.480 (F=1.925)	-0.440 (F=1.582)	0.122 (F=0.130)	-0.496 (F=1.988)	-0.649 (F=3.357)	-0.698* (F=4.268)	0.137 (F=0.185)

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

Note: Standard errors are clustered on the team level. All regressions are OLS regressions including controls for: (1) relative performance in Stage 1 (1=best, 4=worst), and (2) the control variable(s) indicated in the column heading. The regressions in panel (b) also control for: (1) relative performance in Stage 1 interacted with *Male*, and (2) the control variable(s) indicated in the column heading interacted with *Male*. The final row presents results from an F-test, testing the impact of *Male majority team* on male participants. Table A.1 in the Appendix summarizes how all outcome variables are defined.

Table A.3: Impact of gender and team gender composition on likelihood of becoming candidate and leader

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

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

<i>Dependent variable:</i>	<i>Indicator for if the participant became a candidate</i>				<i>Indicator for if the participant became the leader</i>			
	Women		Men		Women		Men	
Female majority team	0.150** (0.056)	0.145** (0.056)	0.111 (0.059)	0.108 (0.059)	0.040 (0.046)	0.035 (0.047)	0.125* (0.061)	0.124* (0.061)
Constant	0.289*** (0.052)	0.416*** (0.094)	0.570*** (0.017)	0.622*** (0.062)	0.158*** (0.042)	0.299*** (0.070)	0.281*** (0.014)	0.304*** (0.068)
<i>N</i>	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.

Table A.4: Correlations between mediating measures

	Guess relative perf. 1 st task	Updating	Rank election	Guess rank election	Influence	Performance 2 nd task	IAT scores of team members
Guess relative perf. 1 st task							
<i>All</i>	1.00						
<i>[Women]</i>	[1.00]						
<i>(Men)</i>	(1.00)						
Updating							
<i>All</i>	0.63***	1.00					
<i>[Women]</i>	[0.59***]	[1.00]					
<i>(Men)</i>	(0.63***)	(1.00)					
Rank election							
<i>All</i>	0.26***	0.29***	1.00				
<i>[Women]</i>	[0.25***]	[0.26***]	[1.00]				
<i>(Men)</i>	(0.19***)	(0.26***)	(1.00)				
Guess rank election							
<i>All</i>	0.53***	0.44***	0.44***	1.00			
<i>[Women]</i>	[0.55***]	[0.44***]	[0.43***]	[1.00]			
<i>(Men)</i>	(0.42***)	(0.36***)	(0.38***)	(1.00)			
Influence							
<i>All</i>	-0.56***	-0.83***	-0.38***	-0.48***	1.00		
<i>[Women]</i>	[-0.53***]	[-0.84***]	[-0.33***]	[-0.46***]	[1.00]		
<i>(Men)</i>	(-0.54***)	(-0.81***)	(-0.37***)	(-0.44***)	(1.00)		
Performance 2 nd task							
<i>All</i>	0.15***	0.16***	0.11**	0.13**	-0.18***	1.00	
<i>[Women]</i>	[0.09]	[0.15*]	[0.04]	[0.11]	[-0.15**]	[1.00]	
<i>(Men)</i>	(0.16**)	(0.14*)	(0.14*)	(0.11)	(-0.17**)	(1.00)	
IAT scores of team members							
<i>All</i>	0.01	0.03	0.00	-0.01	-0.01	0.03	1.00
<i>[Women]</i>	[0.05]	[0.03]	[0.03]	[0.03]	[-0.02]	[0.03]	[1.00]
<i>(Men)</i>	(0.01)	(0.04)	(-0.01)	(-0.02)	(-0.02)	(0.04)	(1.00)

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

Appendix B: Instructions

[INITIAL INSTRUCTIONS]

Thank you for participating in today's experiment.

We will now explain the nature of today's experiment, as well as how to navigate the computer interface with which you will be working. I will use this script to make sure that the information given in all sessions of this experiment is the same.

Today's experiment consists of 4 different tasks. In each of these tasks you can earn money. How much money you earn, depends on your decisions and the decisions of other participants. Before each task we will describe in detail how your payment is determined for that specific task. At the conclusion of the experiment you will only be paid for one of the four tasks. Which task you will be paid for is decided randomly. In addition to your earnings of the selected task you will receive a show up fee of 10 CHF. You will be paid privately, in cash, at the conclusion of the experiment.

If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. If you violate these rules we have to ask you to leave the study and send you home without payment.

Due to the scientific standards of this laboratory, we are not allowed to lie to you or mislead you at any time in the experiment. Accordingly, we will not lie to you or mislead you.

At the beginning of today's experiment, all participants will be randomly assigned to a 4-person group. Throughout today's experiment, you will interact only with the other 3 people in your group.

Please click "Continue" to proceed.

[INDIVIDUAL INFORMATION]

Before the study starts, we would like you to fill in some information about yourself. If you have any questions raise your hand and an experimenter will come to you.

You confirm your entries by clicking the "Continue" button. Once everybody has answered these questions, we will go through the instructions for Task 1.

1. What is your name?
2. How old are you?
3. Are you male or female?
4. Are you a student at the University of Zurich, the ETH, or elsewhere?
5. What is your field of study?

If 'other' which one?

[1. LOST AT SEA TASK: INDIVIDUAL WORK]

TASK 1: LOST AT SEA

Your first task is called *Lost at Sea*. For this task, please imagine the following scenario:

You have chartered a yacht with three friends, for the holiday trip of a lifetime across the Atlantic Ocean. Because none of you have any previous sailing experience, you have hired an experienced skipper and two-person crew.

In the middle of the Atlantic a fierce fire breaks out in the ships galley. The skipper and crew have been lost whilst trying to fight the blaze. Much of the yacht is destroyed and is slowly sinking.

Your location is unclear because vital navigational and radio equipment have been damaged in the fire. Your best estimate is that you are many hundreds of miles from the nearest landfall.

You and your friends have managed to save 10 items, undamaged and intact after the fire. In addition, you have salvaged a four man rubber life craft and a box of matches.

Your task

Your task is to **rank the 10 items in terms of their importance for your survival, as you wait to be rescued.**

Payment

Your suggested solution will be compared to the solution provided by a panel of experts. **The closer you are to the experts' solution, the more money you will earn** if Task 1 is randomly selected to count for your payment at the end of the study. We will explain exactly how your payment is calculated shortly but will first show you the interface with which you will be working to answer the task.

Please click "Continue" to proceed.

HOW TO ENTER YOUR ANSWERS

We will now use a short example to explain how you enter your answer on the screen, and how to calculate your payoff. For simplicity, we assume that you only have 4 items to rank. In the real task you will rank 10 items.

A hypothetical example

Suppose that you have four items to rank according to their importance for your survival on sea. The items are a spoon, a fork, a knife and a corkscrew. Suppose that you believe that the fork is most important for being rescued, the spoon is second most important, the knife is third most important, and the corkscrew is least important. You will then put the fork in the top row next to the number “1”, the spoon at position 2, the knife at position 3, and the corkscrew at position 4. You do this by clicking on an item and dragging it to the right place.

Please practice the task now by dragging the items in the list below to their correct position according to the above example.

If you have any questions about the task at this stage please raise your hand and an experimenter will come to you.

On the next page, we will explain how your answer is scored, and how your score translates to a payoff for this part.

	Your ranking
1	Fork
2	Spoon
3	Knife
4	Corkscrew

Please click the “Continue” button below to proceed.

EVALUATION OF YOUR SOLUTION

All participants start this task with 100 points. In order to calculate your score we will compare your answer with the answer given by the expert panel. If your proposed ranking correspond exactly to that of the experts your final score in the task you keep all 100 points. For each item that you ranked differently than the experts, you lose points. The number of points you lose for an item corresponds to the number of ranks between your proposed ranking and that of the experts. Thus, the further away from the experts' ranking your chosen ranking is for an item, the more points you lose. Your total number of lost points in the task is the sum of the points lost over all 10 items. The fewer points you lose (that is, the closer your ranking is to the expert ranking), the more you earn.

Consider the previous example in which you ranked a fork, a spoon, a knife and a corkscrew. Suppose that the experts' answer confirms that the fork is most important. However, the experts rank the knife as second most important, the corkscrew as third most important, and the spoon as least important. You can see the experts' ranking, and how it compares to your ranking, in the table below.

	Your ranking	Experts' Ranking
1	Fork	Fork
2	Spoon	Knife
3	Knife	Corkscrew
4	Corkscrew	Spoon

The amount of points you would lose in this hypothetical example would be calculated as follows:

- **Fork:** Both you and the experts ranked the fork the same. Thus, you lose 0 points for the fork ($1-1=0$).
- **Spoon:** You ranked the spoon as 2nd most important while the experts ranked it as 4th most important. Thus, you lose 2 points for the spoon ($4-2=2$).
- **Knife:** You ranked the knife as 3rd most important while the experts ranked it as 2nd most important. Thus, you lose 1 point for the knife ($3-2=1$).
- **Corkscrew:** You ranked the corkscrew as 4th most important while the experts ranked it as 3rd most important. Thus, you lose 1 point for the corkscrew ($4-3=1$).

Your total number of lost points would be the sum of lost points for all items. Thus, in total you would lose 4 points ($0+2+1+1$).

On the next page we will tell you how your score is translated into a payoff. Please click the "Continue" button.

YOUR PAYOFF IN TASK 1

In the actual task, where you are asked to rank the 10 items, the maximum number of points you can lose is 50. The maximum number of points you can get in this task is 100 (if you rank all 10 items correctly). Thus, the number of points you get will range from a maximum of 100 points to a minimum of 50 points.

Your payoff from Task 1 (if Task 1 is randomly chosen to count for your payment) is your final number of points divided by 2. So each point is worth CHF 0.5. In other words, your payoff in Swiss francs from this task is calculated according to the formula:

$$\text{Your payment from Task 1} = (100 - \text{Total number of points lost}) / 2$$

Before you make your decisions, we ask you to answer a few control questions to make sure the instructions are clear to everybody.

Please click the “Control Questions” button to proceed.

If you have any questions about the payment in Task 1 please raise your hand and an experimenter will come to you.

CONTROL QUESTIONS

Please enter your answers to the questions below. If you don't know the answers to these questions please raise your hand and the experimenter will come to you. When you are finished please click "Continue".

1. When you perform Task 1 you will give the item you believe is most important for survival the rank 1, the second most important item the rank 2, and so on until all 10 items in Task 1 are ranked. (True/False)
2. Suppose that you are asked to rank 3 items (A, B and C) according to their importance for survival. You rank item A as most important (1st) but the recommended answer is that item A is third most important (3rd). How many points do you lose for your ranking of item A?
3. Suppose that you are asked to rank 3 objects (D, E and F) according to their importance for survival. You lose 1 point for item D, 1 point for item E and 0 point for item F. How many points do you lose in total?
4. Your payment from Task 1 will be $(100 - \text{the points you lose})/2$. If you lose a total of 30 points, how many Swiss francs will you earn from Task 1?

TASK 1: YOUR ANSWER

Scenario

You have chartered a yacht with three friends, for the holiday trip of a lifetime across the Atlantic Ocean. Because none of you have any previous sailing experience, you have hired an experienced skipper and two-person crew.

In the middle of the Atlantic a fierce fire breaks out in the ships galley. The skipper and crew have been lost whilst trying to fight the blaze. Much of the yacht is destroyed and is slowly sinking.

Your location is unclear because vital navigational and radio equipment have been damaged in the fire. Your best estimate is that you are many hundreds of miles from the nearest landfall.

You and your friends have managed to save 10 items, undamaged and intact after the fire. In addition, you have salvaged a four man rubber life craft and a box of matches.

Your answer

The 10 items salvaged are listed below as soon as you click “Start”. The items will be listed in random order. That means the order has nothing to do with the correct answer and will be different for all participants. Please rank these 10 items in terms of their importance for being rescued. Drag the most important item to position one, the second most important item to position two and so forth until you ordered all ten items. As previously described, your suggested solution will be compared to the solution provided by an expert panel. The closer you are to the experts’ solution, the more money you will earn for this task.

You provide your answer by simply clicking on the relevant item and dragging it to the desired position in the ranking.

As soon as you start, you will have 8 minutes to solve this task. A timer in the upper right corner will show you how many minutes you have left. When 1 minute remains, the timer will turn red. Please make sure to enter the complete answer before the time is over.

Please click “Start” to begin the first task.

When you are done, please click the “Submit” button. If you have any questions about Task 1 please raise your hand and an experimenter will come to you.

	Your ranking
1 (most important)	A quantity of mosquito netting
2	A shaving mirror
3	A 25 liter container of water
4	A case of army rations
5	Maps of the Atlantic Ocean
6	A floating seat cushion
7	A 10 liter can of oil/petrol mixture
8	A small transistor radio
9	20 square feet of opaque plastic sheeting
10 (least important)	15 feet of nylon rope

[2. LOST AT SEA TASK: TEAM DISCUSSION]

TASK 2: GROUP DISCUSSION

In the second task you will work in groups and provide a joint group answer to the same task. All participants have therefore been randomly divided into 4-person groups.

Each group will go to a separate room to discuss the task. The group will have 10 minutes to discuss the task and to agree on a joint answer for the group.

We will provide pens and answer sheets for you to fill in the group answer. In the group room you will also find a paper with some brief, additional instructions. It is important to make sure that each group member has filled in the group's joint answer on their individual answer sheet. Each of you will need your answer sheet after you have returned to your seat.

Your payoff from this part depends only on the answer provided by your group, and does not depend on the answers provided by other groups. We will calculate the number of points for the group answer in the exact same way as in Task 1. If Task 2 is randomly chosen for payment, each group member will receive the same payoff based on the group answer. As before, your payoff will thus be calculated as:

$$\text{Your payment} = (100 - \text{Total number of points lost})/2$$

Please click the "Continue" button on your screen now. You will then be informed about which group you have been randomly assigned to, who you are grouped with, and which room you should go to. *When going to your assigned room, please don't speak with other participants until your group is in private in the designated room!*

ROOM ASSIGNMENT

Below you see your group number, as well as the names of the other three people in your group.

Groups 1 and 2 will stay in this room. Members of Group 1 please come to the front of the room, and members of Group 2 please go to the back of the room. The remaining participants in groups 3, 4 and 5, please wait for the experimenter to call your group out and guide you to your rooms. Once you are in the group room, please remain there until an experimenter comes and takes you back to the laboratory.

You have been assigned to GROUP NUMBER [X].

The other members of GROUP NUMBER [X] are:

-NN

-NN

-NN

GROUP ANSWER

All groups have now discussed the task and decided on an answer for their group. Please enter your group answer into the table below. It is important that you enter the ranking your group agreed on correctly! The computer program will compare your entered solution with the solutions entered by the other members of your group, and make sure that they are the same.

As before, you provide the answer by clicking on the relevant item and dragging it to the desired position.

When you are done, please click the “Submit” button.

Please note that the experiment will only continue once all group members have entered the same answer.

	Your group's ranking
1 (most important)	
2	
3	
4	
5	
6	
7	
8	
9	
10 (least important)	

[3. LOST AT SEA TASK: POSSIBILITY TO UPDATE FIRST ANSWER]

TASK 3: UPDATE YOUR INDIVIDUAL ANSWER

In Task 3, you once again work individually. Now you will get a chance to revise the individual answer you provided to the “*Lost at Sea*” task the first time, in Task 1. You can choose to update your previous answers, and change the ranking of the items in any way you want, or to provide the same answer once again. You are not bound to follow your old or your groups answer.

On the next screen, you will see one table with three columns. In column (1), you see the ranking that you submitted individually for Task 1. In column (2) you see the ranking provided by your group in Task 2. In column (3) you will be able to provide your individual ranking once again, by reordering the items as you prefer. As before, the items for you to rank will be presented in a random order.

If Task 3 is randomly chosen to count for payment, your final payment will be based on your revised individual answer in column (3). Based on your answer, your payoff will be calculated in exactly the same way as in Task 1 and 2:

$$\text{Your payment} = (100 - \text{Total number of points lost}) / 2.$$

Please click “Continue” to proceed.

TASK 3: UPDATE YOUR INDIVIDUAL ANSWER

Scenario

You have chartered a yacht with three friends, for the holiday trip of a lifetime across the Atlantic Ocean. Because none of you have any previous sailing experience, you have hired an experienced skipper and two-person crew.

Unfortunately in mid Atlantic a fierce fire breaks out in the ships galley and the skipper and crew have been lost whilst trying to fight the blaze. Much of the yacht is destroyed and is slowly sinking.

Your location is unclear because vital navigational and radio equipment have been damaged in the fire. Your best estimate is that you are many hundreds of miles from the nearest landfall.

You and your friends have managed to save 10 items, undamaged and intact after the fire. In addition, you have salvaged a four man rubber life craft and a box of matches.

Your Answer

The 10 saved items will be listed below as soon as you click “begin”. Please order the items in column (3) according their importance for survival in the “Lost at Sea” task.

You have 6 minutes to provide your answer. As before, a timer in the upper right corner will show you how many minutes you have left. When 1 minute remains, the timer will turn red. Please make sure to enter your answer completely before the time is running out.

When you are done please click the “Submit” button below to submit your answers. If you have any questions about Task 3 please raise your hand and an experimenter will come to you.

	(1) Your ranking	(2) Your group 's ranking	(3) Your updated ranking
1 (most important)	[X]	[X]	
2	[X]	[X]	
3	[X]	[X]	
4	[X]	[X]	
5	[X]	[X]	
6	[X]	[X]	
7	[X]	[X]	
8	[X]	[X]	
9	[X]	[X]	
10 (least important)	[X]	[X]	

[4. LOST AT SEA TASK: GUESS OF RELATIVE RANKING]

TASK 3: PERFORMANCE ESTIMATION

Now, we would like you to guess how well you did in Task 1 (the first individual answer you provided to the “*Lost at Sea*” task) compared to the other 3 members of your group. Please indicate your answer by clicking on one of the options below. If you think your answer was the best in your group you click on the first option. If you think your answer was the second best you click on the second option, and so on.

If your guess is correct, you earn CHF 2. If Task 3 is randomly chosen to count for payment, and you provide a correct guess, the CHF 2 will be added to your final payoff, and paid out to you at the end of the experiment.



My answer was the
best



My answer was the
second best



My answer was the
third best



My answer was the
fourth best

DESCRIPTION OF ELECTION AND THE LEADER'S ROLE

TASK 4

In Task 4 you will perform a task that is similar to the previous “*Lost at Sea*” task. It involves however another survival situation and you are asked to rank 10 different items based on their importance for survival.

You will be working in the same group as in Task 2, and we will ask you to submit one joint answer for the group. This joint answer will be evaluated in the same way as before according to the formula:

$$\text{Your payment} = (100 - \text{Total number of points lost})/2$$

If Task 4 is chosen for payment, all members of your group will be paid the same amount, based on the group's joint answer.

The task differs from Task 2 in how you decide on a group answer. This time, you will not have a group discussion in person. Instead, your group will elect a leader, who will interact with the other group members via the computer. The leader provides a final answer for the group. Below we first provide an overview of Task 4. Thereafter we explain each step in detail.

Overview of Task 4

1. Choosing two candidates: First, all group members indicate how much they want to become the group leader. The two group members who want to become group leader the most are then selected as candidates in the election.
2. Electing a leader: One of the two candidates is elected as the group leader. The group elects the leader through a group vote in which group members vote for their preferred leader. After the election all group members are informed who the candidates were and who got elected as leader. Who you voted for will stay anonymous.
3. Individual answers: Each group member, including the leader, solves Task 4 individually.
4. Group answer by leader: All group members' individual answers are displayed to the leader. The leader thereafter submits a final joint answer for the group.

Payoff: If Task 4 is randomly chosen to count for payment, all group members receive the same payment, which is based on the joint group answer submitted by the leader.

Please click “continue” to proceed.

HOW THE TWO CANDIDATES ARE CHOSEN

Before your group elects a leader, each group member will indicate, on a scale from 1-10, how much they want to become the group leader. “1” means that you very much want to become the leader, and “10” means that you do not want to become the leader at all.

Based on these choices, the two group members who want to be the leader the most will be chosen as candidates. If several group members have chosen the same number, the computer will randomly determine the order of these group members.

A few examples:

- **If the group members choose numbers 1, 3, 7, and 8:** The group members who chose 1 and 3 will become the candidates.
- **If the group members choose numbers 4,4,4, and 5:** The computer will randomly choose two candidates among the three group members who chose 4.
- **If the group members choose numbers 2,5,5, and 8:** The group member who chose 2 will become a candidate. The computer will randomly choose the other candidate among the group members who chose 5.

When all group members have indicated how much they want to become the group leader, the election of the leader takes place. We ask all group members to provide their vote before it is known who the two candidates are.

After the election the entire group will be informed who the candidates were and who got elected as leader. Your score how much you wanted to become the leader will not be disclosed to the group.

Before you proceed to the decisions we would like to ask you to answer a few control questions to make certain that the instructions were understandable.

Please click on “Control Questions” to get to the control questions.

CONTROL QUESTIONS ABOUT SELECTION OF CANDIDATES

Please work through the questions below. If you have any questions please raise your hand and an instructor will come to you. When you are done please click the “Continue” button.

1. If you want to become the group leader, should you pick the number 10 when asked how much you want to become the group leader? [TRUE/FALSE]
2. If you choose number 4, and the other group members choose numbers 1,2 and 10, will you become a candidate in the election? [YES/NO]
3. If you choose number 2, and the other group members choose numbers 2, 5 and 7, will you become a candidate in the election? [YES/NO]

DETAILED DESCRIPTION OF THE ELECTION

In order to determine the leader all group members will vote, individually and anonymously, for their preferred leader. Each group member votes by providing a ranking of the other three group members with the preferred leader at position 1, the second most preferred leader at position 2, and the third most preferred leader at position 3. You cannot vote for yourself.

After all votes are recorded, the computer checks which two group members are the candidates in the election. The votes from these two candidates will not be considered. The final vote from each of the other two group members, who are not candidates, goes to the highest ranked out of the two candidates.

To illustrate, imagine that you are in a group together with group members A, B, C, and D. Assume that B and C are the candidates. that you rank them in the following way:

- A ranked C as number 1, B as number 2, and D as number 3.
- D ranked B as number 1, A as number 2, and C as number 3.
- B got ranked once as number 1 and once as number 2. This results in an average ranking of $(1+2)/2=1.5$.
- C got ranked once as number 1 and once as number 3. This results in an average ranking of $(1+3)/2=2$.
- In this example **B becomes the leader**, as B received the better average ranking.

If both candidates tie, the decision will be made randomly.

Note that in this election you cannot vote strategically to increase your own chances of being elected as the leader. Your vote only counts if you yourself are not one of the two candidates in the election. Therefore, it is in the interest of all group members to provide their true, preferred, ranking of the other group members.

Before you proceed to the decisions we would like to ask you to answer a few control questions to make certain that the instructions were understandable.

Please click on “Control Questions” to get to the control questions.

CONTROL QUESTIONS ABOUT ELECTION AND PAYOFF

Please answer a few control questions to make sure the instructions were clear. If you have any questions don't hesitate to raise your hand. When you are done please click the "Continue" button.

1. When you vote, should you rank your preferred leader as number 1, 2, or 3?
2. If you are one of the two candidates in the election, your vote does not count for the outcome of the election. [TRUE/FALSE]
3. All group members get the same payoff from the task, and this payoff is based on the solution submitted by the leader. [TRUE/FALSE]

[5. INDICATE WILLINGNESS TO BECOME TEAM LEADER]

PLEASE INDICATE YOUR WILLINGNESS TO BECOME THE GROUP LEADER

We now ask you to indicate how much you want to become the group leader.

Please choose any number from 1 to 10, with 1 meaning that you would like to become leader very much, and 10 meaning not at all.

When you have made your choice, please click the “Continue” button.

How much would you like to become the group leader?

Very much

Not at all

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1	2	3	4	5	6	7	8	9	10

[6. ELECTION OF TEAM LEADER]

TASK 4: ELECTING A LEADER

In this part you submit your vote for who should become the group leader. Below you see a list of the other members of your group. You cast your vote by ranking the other group members according to who you would like to see lead your group.

As when you ranked the items in the other tasks, you can drag and drop the names of the other group members to the rank of your choice. Put the name of the person you prefer to be the group leader in position 1, the second most preferred person in position 2, and the person you prefer the least in position 3.

When you are done with your choice, please click the “Submit” button.

Remember that you cannot affect your own chances of being elected. If you are one of the two candidates in the election, your vote does not count for the outcome of the election. Thus, it is in your interest to rank all group members only according to who you would like to see lead the group.

1	
2	
3	

[7. GUESS OF VOTES RECEIVED]

TASK 4: ASSESSMENT OF ELECTION

All participants have voted and thereby ranked their group members on the first, second, and third rank. Basing on the votes of all group members, independently of who became a candidate, we can calculate the average rank of each group member. *We would like to ask you to estimate how high your average rank is in comparison to the rank of your group members (based on the votes of all 4 group members).*

Example: Imagine a group with the participants A, B, C, D:

- **A** got ranked as number 1 by all three group members. This results in the average rank $(1+1+1)/3=1$.
- **B** got ranked twice as number 2 and once as number 3. This results in the average rank $(2+2+3)/3=2.3$
- **C** got ranked once as number 1, once as number 2 and once as number 3. This results in the average rank $(1+2+3)=2$
- **D** got ranked once as number 2 and twice as number 3. This results in the average rank $(2+3+3)=2.7$

In this example, A got the best average ranking (1) and D the fourth best average ranking (2.7).

Please choose from the listed options. Choose the first option, if you are of the opinion that your average rank was the best in your group. If you are of the opinion that your average rank was the second best in the group, choose option 2, and so forth. In case two group members received the same average rank, the computer decides randomly who gets ranked better.

If your guess is correct, you earn CHF 2. If Task 4 is randomly chosen to count for payment, and you provide a correct guess, the CHF 2 will be added to your final payoff, and paid out to you at the end of the experiment.



Best average rank



Second best average rank



Third best average rank



Fourth best average rank

For the remainder of the study the instructor won't read out the instructions loudly. Please work through the rest of the experiment on your own. That means you may click continue without being asked to do so by the instructor. Please don't hesitate to raise a hand if you have any questions. An instructor will come to you.

[THE LEADER IS REVEALED]

All group members have now submitted their votes, and a leader has been elected. The candidates were [XXX] and [YYY].

The leader for your group is: [ZZZ].

Please click the “Continue” button when you have read the leader’s name. On the next screen you will then see the introduction to Task 4.

[8. DESERT SURVIVAL TASK: INDIVIDUAL WORK]

TASK 4: DESERT SURVIVAL

The fourth task is called “*Desert Survival*”. The complete scenario of the task is described on the next page.

Your task

Your task is to **rank 10 items in terms of their importance for your survival in the desert**. Put the most important item at rank 1, the second most important item at rank 2, and so on, until you have ranked all items. As before, the items will be presented in a random order to each participant.

Your answer will be **shown to the group leader** together with the answers from the other two group members. The group leader will then decide on, and submit, a joint group answer.

Payment

The closer the answer that the group leader submits is to the experts’ answer, the more money you will earn if Task 4 is randomly selected to count for your payment. Your payment is calculated as before according to the formula:

$$\text{Your payment} = (100 - \text{Total number of points lost})/2$$

As before, all group members will be paid the same amount.

Please click the “Continue” button below to get to the description of Task 4.

TASK 4: YOUR ANSWER

For Task 4, please imagine the following scenario:

It is approximately 10:00 am in mid-July and you, together with three of your friends, have just crash landed in the Atacama Desert in South America. Your light twin-engine plane including the pilot and co-pilot has completely burned out with only the frame remaining. None of you have been injured.

The pilot was unable to notify anyone of your position before the crash. However, he had indicated before impact that you were 50 miles from a mining camp, which is the nearest known settlement, and approximately 65 miles off the course that was filed in your Flight Plan. The immediate area is quite flat, except for occasional cacti, and appears to be rather barren.

The last weather report indicated that the temperature would reach 110 F today, which means that the temperature at ground level will be 130 F. You are dressed in lightweight clothing-short-sleeved shirts, pants, socks, and street shoes.

Everyone has a handkerchief and collectively, you have 3 packs of cigarettes and a ballpoint pen. Before your plane caught fire, your group was able to salvage 10 items.

Your Answer

The 10 items salvaged are listed below as soon as you click “Start”. Please rank these items in terms of their importance for survival (1 is most important and 10 is least important).

You have 8 minutes to solve this task as soon as you start. A timer in the upper right corner will show you how many minutes you have left. When 1 minute remains, the timer will turn red. Please make sure you entered your complete answer before time runs out.

Your answer will be shown to the leader together with the answers of the other group members when the leader decides about the group’s answer. To evaluate the leader’s answer, the answer will be compared to the experts solution as described before. All group members will receive the same payment basing on the group’s answer.

When you are done, please click the “Submit” button. If you have any questions about Task 4 please raise your hand and an experimenter will come to you.

	Your ranking
1 (most important)	A cosmetic mirror
2	An overcoat (for everyone)
3	2 liters of water per person
4	Torch with 4 battery-cells
5	A parachute (red & white)
6	A folding knife
7	A 45 caliber pistol (loaded)
8	A first-aid kit
9	A book entitled “Desert animals that can be eaten”
10 (least important)	A bottle of salt tablets

[9. DESERT SURVIVAL TASK: LEADER DECIDES ON TEAM ANSWER]

TASK 4: DECISION FOR THE GROUP [ONLY FOR THE LEADER]

You are the leader of your group. On the next page you will see the individual rankings provided by each group member, including your own. Your task is to submit a final ranking of the items on behalf of the group. The group answer you submit will be the basis for payment for all group members if Task 4 is randomly chosen for payment. The closer your answer is to the experts' answer, the more money you, and the members of your group, will get. It is up to you if you take the answers of your group members into consideration when submitting the group answer.

As before, the items for you to rank will be presented in a random order.

You have 6 minutes to submit an answer. When 1 minute remains, you will get a notification reminding you to finalize your answer.

If you have any questions about Task 4 please raise your hand and an experimenter will come to you.

Please click the "Continue" button to submit your answers when you are done.

	Your ranking	Ranking by [X]	Ranking by [Y]	Ranking by [Z]	Your proposed group answer
A cosmetic mirror	[n]	[n]	[n]	[n]	
An overcoat (for everyone)	[n]	[n]	[n]	[n]	
2 liters of water per person	[n]	[n]	[n]	[n]	
Torch with 4 battery-cells	[n]	[n]	[n]	[n]	
A parachute (red & white)	[n]	[n]	[n]	[n]	
A folding knife	[n]	[n]	[n]	[n]	
A 45 caliber pistol (loaded)	[n]	[n]	[n]	[n]	
A first-aid kit	[n]	[n]	[n]	[n]	
A book entitled "Desert animals that can be eaten"	[n]	[n]	[n]	[n]	
A bottle of salt tablets	[n]	[n]	[n]	[n]	

TASK 4: DECISION FOR THE GROUP [ONLY FOR NON-LEADERS]

You have now submitted your answer. All group members solutions are now shared with the leader, and the leader will decide on the final group answer.

On the next page you will see a large table with the solutions from all four group members. In the second table below we would like you to rank the items once while the leader works on the final group answer. It is up to you whether you want to take the answers of the other three group members into consideration when doing this or not. As before, the items for you to rank will be presented in a random order.

The answer you provide on the next page will not be relevant for your final payment, but it will be used as a benchmark later in the study.

Please click the “Continue” button to get to the next screen.

TASK 4: GROUP DECISION [ONLY FOR NON-LEADERS]

Below you see the individual rankings provided by each group member, including your own. Please submit your final ranking of the items.

You will have 6 minutes to complete this task.

	Your ranking	Ranking by [X]	Ranking by [Y]	Ranking by [Z]	Your proposed group answer
A cosmetic mirror	[n]	[n]	[n]	[n]	
An overcoat (for everyone)	[n]	[n]	[n]	[n]	
2 liters of water per person	[n]	[n]	[n]	[n]	
Torch with 4 battery-cells	[n]	[n]	[n]	[n]	
A parachute (red & white)	[n]	[n]	[n]	[n]	
A folding knife	[n]	[n]	[n]	[n]	
A 45 caliber pistol (loaded)	[n]	[n]	[n]	[n]	
A first-aid kit	[n]	[n]	[n]	[n]	
A book entitled "Desert animals that can be eaten"	[n]	[n]	[n]	[n]	
A bottle of salt tablets	[n]	[n]	[n]	[n]	

Please press the "Continue" button once you are done.

[10. EVALUATION OF LEADER]

TASK 4. EVALUATION OF LEADER [ONLY FOR NON-LEADERS]

Thank you for completing the “*Desert Survival*” task.

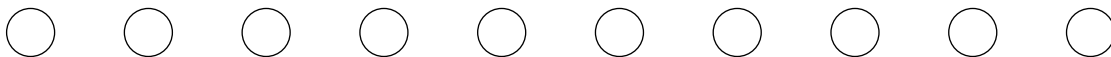
The group payment is made based on the answers chosen by the leader of the group. Before providing these answers, the leader was given information about the individual answers from all group members.

The leader submitted the following answer for the group:

	Group ranking
1 (most important)	[X]
2	[X]
3	[X]
4	[X]
5	[X]
6	[X]
7	[X]
8	[X]
9	[X]
10 (least important)	[X]

In total, the group answer selected by the leader resulted in X points, giving each group member a payoff of XX CHF.

How well do you think the leader did? Please evaluate your leader’s performance on the scale below from “1 (very good) to 10 (very bad).



Very good

Very bad

TASK 4. EVALUATION OF LEADER [ONLY FOR LEADERS]

Thank you for completing the “*Desert Survival*” task.

The group payment is made based on the answers you chose for the group. You submitted the following answer:

	Group ranking
1 (most important)	[X]
2	[X]
3	[X]
4	[X]
5	[X]
6	[X]
7	[X]
8	[X]
9	[X]
10 (least important)	[X]

In total, your answer resulted in X points lost, giving each group member a payoff of XX CHF.

How well do you think you did as a leader for your group in Task 4? Please evaluate your performance on the scale below from “1 (very good) to 10 (very bad).



Very good

Very bad

QUESTIONNAIRE

QUESTIONNAIRE

Nationality: _____

Which is the highest educational diplomas your mother has? (Many alternatives possible)

Elementary school
apprenticeship
high school
university of applied sciences
university
no answer

Which is the highest educational diplomas your father has? (Many alternatives possible)

Elementary school
apprenticeship
high school
university of applied sciences
university
no answer

How would you classify yourself politically on the left/right scale below? (0 to 10)

[HORIZONTAL LINE OF RADIO BUTTONS]

In general, how willing or unwilling you are to take risks on a scale from 0 to 10? 0 means you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any integer number between 0 and 10 to indicate where you fall on the scale. Possible choices are: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

[HORIZONTAL LINE OF RADIO BUTTONS]

Consider the survival tasks performed in this study. On average, do you think that men are better at such tasks, that men and women are equally good, or that women are better? Please provide an answer on the scale below, where 0 means that men are better, 5 that women and men are equal, 10 that women are better.

[HORIZONTAL LINE OF RADIO BUTTONS]

On average, do you think that men are better leaders, that men and women are equally good leaders, or that women are better leaders? Please provide an answer on the scale below, where 0 means that men are better, 5 that women and men are equal, 10 that women are better.

[HORIZONTAL LINE OF RADIO BUTTONS]

Do you have previous experience of leadership activities?

YES/NO

If yes, what?

Did you volunteer to become the leader of the group? [YES/NO]

If Yes: Why?

If No. Why not?

Do you think you had too much time, too little time, or just about the right amount of time to solve the tasks and provide your answers?

[AFTER THE QUESTIONNAIRE THE PARTICIPANTS TOOK PART IN THE IAT]