I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16625

Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams

Christopher F. Karpowitz
Stephen D. O'Connell
Jessica Preece
Olga Stoddard

NOVEMBER 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16625

Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams

Christopher F. Karpowitz
Brigham Young University

Jessica Preece
Brigham Young University

Stephen D. O'Connell
Emory University

Olga Stoddard
Brigham Young University and IZA

NOVEMBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams*

Policies that increase women's representation often intend to provide women with influence over processes and decisions of the organization in which they are implemented. This paper studies the effect of gender composition and leadership on women's influence in two field experiments. Our first study finds that male-majority teams accord disproportionately less influence to women and are less likely to choose women to represent the team externally. We then replicate this finding in a new context and with a larger sample. To investigate the relationship between formal leadership and women's influence and authority, the second study also varied the gender of an assigned team leader. We find that a female leader substantially increases women's influence, even in male-majority teams. With a model of discriminatory voting, we show that either increasing the share of women or assigning a female leader reduces the rate at which individual teammates discriminate against women by more than 50%. These conditions both increase the influence of women and improve women's experience in work teams by creating an institutional environment that reduces the expression of discriminatory behavior at the individual level.

JEL Classification: J16

Keywords: gender, field experiment

Corresponding author:
Olga Stoddard
Department of Economics
Brigham Young University
Provo, UT 84602
USA

E-mail: olga.stoddard@byu.edu

* This paper provides a substantial update and extension of IZA DP No. 13741, including an addition of Study 2 and corresponding analyses. Katelyn Cranney, Nephi Dummar, Mandi Eatough, Rachel Fisher-Sandhu, Marissa Gerber, Dani Hogan, Boston Kelley, Connor Kreutz, and Guido Romero provided excellent research assistance. Sandra Black, David Card, Zoe Cullen, Jeff Denning, Jennifer Doleac, Laura Gee, Lars Lefgren, Emily Leslie, Emily Nix, and Richard Patterson gave valuable feedback on earlier drafts of the paper. We thank audience members at the Harvard Kennedy School, University of Chicago Harris, Michigan State University, Notre Dame University, Texas A&M University, George Mason University, Claremont McKenna College, Luxembourg Institute of Socio-Economic Research, University of Chicago Experimental Seminar, Regensburg University, Brigham Young University, ASSA, Advances in Field Experiments Conference, European Conference of Politics and Gender, Science of Diversity and Inclusion Conference, IZA ECBE conference, Economics of Education Virtual Seminar, the 2019 Utah Experimental Conference, and LACEA 2023. Finally, we are grateful to Jack Cavanaugh at J-PAL, who led a successful independent computational replication of the analyses in this paper and provided feedback on the replication package. Research support was provided by the Russell Sage Foundation, Brigham Young University College of Family, Home, and Social Science, and Emory University's Program to Enhance Research and Scholarship. Study 2 was pre-registered under AEARCTR-0005363.

1 Introduction

In recent years, concerns about women's under-representation in many settings have led to a variety of public and private initiatives to diversify work teams, corporate boards, public commissions, academic panels, and other such groups. These initiatives vary dramatically in intensity. For example, in 2008, Norway implemented a gender quota of 40% women for corporate boards, and several other Western European countries followed suit (The Economist, 2018). By contrast, the only US state that has instituted a corporate board quota, California, set a much different standard: California's Women on Boards law requires all publicly held corporations whose principal offices are located in the state to have at least one female director on their Board of Directors. Early evidence suggests that many California corporations have chosen to comply with this law by adding a lone woman to their otherwise all-male boards (Padilla, 2020).

Presumably the purpose of these interventions is to bring women's voices, experiences, and expertise to the table so they can influence the discussions these bodies have and the decisions they make. Yet many critics argue that interventions placing a small number of women in a male-dominated setting are likely to fail in this respect because women's input is undervalued when they are significantly outnumbered by men. Are these programs effective? What are the conditions under which they are more or less likely to be effective? Does the proportion of women in a setting matter for their ability to influence deliberation and decision-making? Does it shape whether their influence is recognized and rewarded by others? What role does organizational and leadership structure play in exacerbating or ameliorating these dynamics? If these institutions matter, is it through changing beliefs or simply blunting the practical effects of these beliefs by changing behavior? Organizations in many settings grapple with these questions as they consider how to best take advantage of their human capital and create an inclusive environment for their employees and stakeholders.

We conduct two multi-year field experiments to test the extent to which group gender composition shapes women's influence in deliberation and decision-making and whether the identity of a team's leader matters for women's influence. We also explore the mechanisms through which treatment effects occur. Both studies involved undergraduates in the United States. Study 1 took place in an accounting program situated within a business school curriculum, and Study 2 took place in a general education course at a private university in the United States. Subjects in both studies participated in a group-based pedagogical approach specifically designed to emulate long-lived collaborative teams typical in modern work settings. In Study 1, we randomly assigned subjects first to a gender composition condition, and then within that condition to one of 75 mixed-gender majority-male or majority-female groups.[1] For Study 2, we randomly assigned enrollees into majority-male or majority-female mixed-gender composition conditions and then to one of 145 six-person study groups.[2]

Groups met weekly for four months to work on incentivized assignments, projects, and deliberative tasks that are common in many work environments. We recorded the audio of two team-building exercises (Studies 1 and 2) and weekly team meetings (Study 2) and collected individual and group-level measures of influence and performance. We collected survey-based evaluations from team members each month during the semester, accessed administrative data on program performance, and also observed group behavior in response to incentivized team-building exercises. Study 2 was pre-registered and purposefully provided for the measurement of the same set of outcomes of focus in Study 1 in order to serve as a replication test, and began collecting data after the conclusion of Study 1. Additionally, Study 2 added a crossed treatment condition in which we randomly selected a man or woman

[1] Of these groups, 73 were five-person groups and 2 were comprised of four members.

[2] Two groups included only five participants. Because of the gender composition of the sample populations, both studies also included a larger number of participants who were randomly assigned to either all-male (Study 1) or all-female (Study 2) conditions, but our primary focus here is on the mixed-gender conditions. Full details for both studies are reported below. In addition, 20 participants (2 from Study 1 and 18 from Study 2) declined to give consent for their individual-level data to be included in our study, so these individuals are dropped from individual-level analyses.

in each group to serve as team leader halfway through the semester, allowing us to test whether formal leadership structure shaped outcomes.

In both studies, we find that women in the numerical minority are disproportionately less likely to be rated as influential in team deliberations or to be chosen as a spokesperson for their team relative to women in female-majority teams. In other words, after accounting for the proportion of women in the group, group gender composition is causally related to who is perceived as influential and capable of representing the group. This disadvantage is specific to women — men in the numerical minority do not experience it. We then show that formal leadership structure also affects women's influence gap. Prior to the assignment of a team leader, patterns of influence in Study 2 largely mirror those of Study 1 — women in the minority are disadvantaged. However, after the random assignment of a team leader, patterns of influence are primarily shaped by whether the team was led by a man or a woman, with influence votes for women approaching parity in female-led groups, but not in male-led groups. Our research design also permits us to show that repeated interactions over time may sometimes ameliorate gendered influence deficits and that the behavior of both men and women contributes to women's lower influence in male-majority teams.

We then calibrate our empirical results to a simple voting model to detect the level of discrimination women face from teammates. This approach allows us to characterize the degree to which the changes in women's influence across conditions are the result of changed behavior at the individual level, or simply reflect compositional effects on the expression of otherwise stable levels of discrimination. The results of this exercise suggest that our experimental results stem from a large baseline penalty against women. For a woman to earn the vote of a teammate, she must outperform the highest-performing man by approximately .6 standard deviations in the male-majority groups, and .3 standard deviations in female-majority groups. As a result, changes in group gender composition reduce the average penalty experienced by a woman from her teammates by half. Increasing the share of women

meaningfully improves the experience and influence of women in these teams by reducing the intensity of discrimination expressed towards them. We then show that levels of sexist attitudes, collected via a standard survey module asked at the beginning and end of Study 2, do not change over the course of the study. This implies that the treatment conditions affected the *expression* of discriminatory beliefs but not the beliefs themselves, aligning with recommendations for approaches to address discrimination that focus on designing institutions that constrain the expression of discriminatory beliefs, rather than changing them (Bohnet, 2016; Beaman et al., 2009). Calibrating the empirical results from Study 2's leadership intervention reveals that female leaders also dramatically decrease the penalty against women, while male leaders appear to affect no change in the penalty experienced by women. Thus, changing the institutional structure of leadership can reduce the expression of discriminatory beliefs without requiring changes to gender composition.

Our findings contribute to a new and growing literature in economics and other social sciences on mixed-gender work teams.[3] Understanding how to design effective teams is a crucial task for employers. Teams can be an efficient way to coordinate production requiring a diversity of skills, talents, and information (Eckel and Grossman, 2005; Lindbeck and Snower, 2000) and have become increasingly common in the workplace (Nix, 2020; Lazear and Shaw, 2007; Boning et al., 2007; Hamilton et al., 2003; Lawler and Mohrman, 2003; Wuchty et al., 2007; Berg et al., 1996). Since the 1970s, sociologists have argued that token women are disadvantaged in groups (Kanter, 1977; Spangler et al., 1978; Ott, 1989; Jamieson et al., 1995; McDonald et al., 2004), though these studies struggled to isolate causal effects. More recently, lab experiments in political science have found that group gender composition shapes conversation dynamics, decision-making, and the distribution of authority (Karpowitz and Mendelberg, 2014; Mendelberg et al., 2014).

[3]Although we recognize the conceptual distinction between sex and gender, in this paper we use them interchangeably.

In economics, prior work finds that evaluations of competence are highly gendered, and the gender composition of an environment and how a task or domain is stereotyped matter a great deal for these perceptions (Bordalo et al., 2019; Karpowitz and Mendelberg, 2014; Preece, 2016). Challenges appear to be especially acute in settings where women have been historically underrepresented (Shan, 2023; Bordalo et al., 2016; Coffman, 2014). This leads to outcomes such as gender disparities in leadership because women correctly presume they will have less support from their team when men are in the majority (Radbruch and Schiprowski, 2023; Born et al., 2020); women taking and getting less credit for joint work with men (Coffman et al., 2021; Koffi, 2020; Isaksson, 2018; Sarsons, 2017); and women doing more of the "non-promotable" tasks in groups (Babcock et al., 2017).

Group gender composition also shapes a variety of other important outcomes. For example, it influences decisions about business strategies, with male and mixed-gender teams focusing on profit and research and development and all-female teams focusing more on sustainability (Apesteguia et al., 2012). Gender-diverse teams produce higher-impact science (Yang et al., 2022) on a more diverse range of topics (Truffa and Wong, 2021). Men on scientific selection committees support women candidates less when they have female colleagues (Bagues et al., 2017; Deschamps, 2021). Male-dominated educational settings may decrease women's performance (Bostwick and Weinberg, 2020; Hampone et al., 2022; Booth et al., 2018), though some studies find null effects (Anelli and Peri, 2019) or countervailing considerations (Zölitz and Feld, 2018; Brenoe and Zölitz, 2019). There is some evidence that women's performance drops when women work with men they do not choose, but not when working with men they choose (Aman-Rana et al., 2022; Calder-Wang and Gompers, 2021). These findings substantiate the idea that patterns of gender marginalization are not just normatively concerning, but also have implications for team effectiveness and firms' productivity (Yang et al., 2022; De Paola et al., 2022; Truffa and Wong, 2021).

Organizational structure and hierarchy also affect women in the workplace. Existing

research finds that leaders shape the behavior of subordinates, perhaps especially in smaller groups, and female leaders are good for female subordinates (Sahin et al., 2015; Andreoni, 2006; List and Rondeau, 2003; Kofoed et al., 2019; Calder-Wang and Gompers, 2021). Further, workplaces with more men and more male managers have higher levels of sexual harassment (Folke and Rickne, 2020; Dahl et al., 2020). Perhaps most surprisingly, there is field experimental evidence that women avoid organizations that have "flat" corporate structures (Hurst et al., 2022). Our results give some context for why women may avoid applying for jobs that promote an ostensibly more egalitarian environment — in the absence of a hierarchical structure, we find that norms in a majority-male setting dramatically favor men.

2 Research Design

2.1 Study 1

We partnered with a selective undergraduate accounting program in the US to randomize the gender composition of students' assigned work teams during the fall semester of their junior year. Students are competitively admitted to the program at the end of their sophomore year and they begin the rigorous program the following fall semester. The assigned five-person teams are a critical part of their academic experience: they attend classes together, work together on cases and other course assignments, and study together for exams. Program administrators designed these teams to replicate the work environment of many accounting firms, and faculty use these groups to train students on professional interactions. As is typical of teams in the business world, students are instructed to make decisions collectively and to respect each other's contributions. During the course of the semester, students complete both individual and group assignments and interact frequently with each other both in and

out of class.[4]

We enrolled 522 students (379 men and 146 women) across the fall semesters of 2016 and 2017. Women comprise only about 28% of the overall enrollment in the program, and historically the program responded to this imbalance by placing one or two female students per group to minimize the number of all-male teams. Program directors were, however, interested in understanding the effects of this method of team assignment on their female students' experiences. Hence, in this study, female students were randomly assigned to one of two conditions: 1 woman and 4 men (1F) or 3 women and 2 men (3F). Because there were significantly more men than women in the program, men were randomized into one of three conditions: all five men (0F), 1F or 3F.[5] Once individuals had been randomized into a condition, we randomized students into groups. The program stipulated that the average GPA of each group should be similar and that there should only be one international student per group. Hence, we block randomized on these variables.

Throughout the semester, we collected data from team-building exercises, surveys providing broader assessments of peers and group functioning, and administrative records to measure individual and group outcomes. Prior to group assignment, students took an extensive baseline survey. After group assignment, students completed monthly surveys during the fall semester that focused on their perceptions of group interactions and invited them to evaluate their team members' contributions.[6]

In addition to the survey and administrative data, teams participated in an incentivized team-building exercise at both the beginning and end of the semester. At the beginning of the semester, the task was the "Survival on the Moon" exercise, in which participants are given a list of 15 items and asked to rank them from most to least useful for survival on the moon

[4]This program is not unique in using assigned teams as a pedagogical and professionalization tool; many accounting programs and most MBA programs also assign students to work teams like these.

[5]In addition, 8 students were assigned to 2F groups, but these are dropped from the current analysis, given the lack of statistical power and the fact that these groups included fewer than 5 participants.

[6]Group members' names were always presented to survey respondents in random order.

(Hall and Watson, 1970). This task was originally conceived as a group-based pedagogical exercise, and expert answers were constructed by NASA. At the end of the semester, the task was the similar "Lost at Sea" task (PACE, 2016) developed by the U.S. Air Force.[7]

Upon arriving at the lab, students first completed the exercise on their own, giving us a measure of how well each individual independently performed on the task. Participants were then asked to deliberate with their group members to create a group ranking of the items. Because discussions were recorded, we are able to identify individual participation in discussions with great precision.[8] After they determined their group ranking, each individual privately completed an exit survey about their group deliberation experience in which they voted for a spokesperson to present their results to a panel of judges at a later date.[9] Performance was incentivized in several ways: in addition to the $5 show-up fee, students earned $50 for having the individual answer closest to the experts, $50 per group member for having the group answer closest to the experts, and another $50 for the whole group to share for the best presentation given by their chosen spokesperson.

Students were required to participate in the monthly surveys and team-building exercises for course credit and for the purpose of an internal program evaluation. However, to be included in the individual-level analysis in this paper, they had to consent to allow their individual data to be used for research purposes.[10] Participants were informed that both the

[7]These tasks are commonly used as leadership and team-building exercises in the corporate world; detailed task instructions are available in the supplemental appendix.

[8]Recording equipment was 6-channel audio recording, which yields a separate, high-quality recording for each member of the group as well as a track that records the group together.

[9]Students were told that five groups would be selected at random to have an opportunity to have their spokesperson present their group ranking to the judges in 1-2 weeks for a chance to earn additional $50 for the best presentation. The groups were thus incentivized to select the most capable group member as their spokesperson.

[10]In Study 1, 0.5% of the students assigned to mixed-gender groups (2 out of 373) did not consent to the use of their data, and their responses are not included in the individual-level results reported below. In Study 2, the corresponding share of students is 2.1% (18 out 868), and these are also excluded from individual-level analyses. Non-consent rates are uncorrelated with treatment status. In a clustered difference-of-proportions test, with standard errors clustered by group, consent does not differ by experimental condition ($z = 0.56$, $p = 0.57$).

internal program evaluation and the research study would examine team dynamics (without any mention of gender), and the experiment did not include any deception.

2.2 Study 2

Beginning in 2019, we partnered with a large general education course at a private university in the U.S. to implement another field experiment (hereafter, Study 2) with an intent to build upon the design of Study 1 and investigate a potential mechanism behind the original results. Study 2 was designed similarly to Study 1: it is a multi-year field experiment in which we randomly assigned 1,308 men and women into study groups, 868 of whom were assigned to mixed-gender groups. Groups met weekly for four months to work on incentivized assignments, projects, and deliberative tasks that are common in many work environments. We audio recorded their weekly interactions and collected a wide range of individual and group-level measures of influence and performance (including the same team-building tasks and outcomes as in Study 1). By measuring the same outcomes in Study 2 as in Study 1, we can test directly whether the results of Study 1 replicate in a different setting.

Different from Study 1, Study 2 took place in a liberal arts setting in which women comprised the overall majority. Groups included six members and consisted of 3 gender composition conditions — two women and four men in majority-male groups (2F), four women and two men in majority-female groups (4F), and all-women groups.[11] We fielded Study 2 over four semesters in 2019-2021 as part of a required general education course that covers basic principles of American history, politics, and economics.[12] Thus, Study 2's subject sample is drawn from a population more general than those pursuing a competitive application-only business degree. As such, Study 2 represents a much less male-dominated setting. While women comprised only 28% of the sample population in Study 1, they were

[11] Because exact numbers depended on enrollment in the study, a small number of groups had less than six members.

[12] Because of the pandemic, the 2020 and 2021 teams met online via Zoom.

about 65% of the sample in Study 2.

Study 2 also included a new treatment to investigate a potential mechanism for women's empowerment. Specifically, halfway through the semester, we randomly assigned half of the groups to have a male leader, and half to have a female leader. The leader was then chosen randomly from among available candidates in the condition to which the group was assigned. All group members were informed that a group leader was chosen at random and that the leaders possessed administrative responsibility to ensure that groups stayed organized and submitted their assignments on time.

While Study 2 shares the core design and measurement features of Study 1, it also adds some unique strengths. In particular, the symmetric design of 2F/4F groups in Study 2 also allows for direct comparisons of men and women when they constitute an equivalent numerical minority. This allows us to investigate whether men face the stark disadvantages women do when they find themselves in the numerical minority in directly analogous group contexts. Finally, the inclusion of Study 2 permits us to pursue additional analysis to investigate men's and women's behaviors across conditions. For example, we can test whether women's higher propensity to vote for other women in female-majority groups is a potential mechanism behind our findings. Because the minority-women condition in Study 2 includes two women (as opposed to one woman in Study 1), we also test whether women's likelihood to vote for other women systematically varies across 2F and 4F conditions. We also were not required to block on GPA or other features while assigning students to groups in Study 2. Table 1 provides a side-by-side comparison of the main features of Studies 1 and 2.

2.3 Sample and Balance Tests

In Study 1, admission to the program is highly competitive. Matriculated students typically have high GPAs and considerable leadership experience. As reported in Appendix Table

Table 1: Features of Study 1 and Study 2

Core Features	**Study 1**	**Study 2**
Subjects' field of study	Accounting	Liberal arts
University type	Private	Private
Share female in study population	28%	65%
Group size	5	6
Number of groups	75	145
Individual subjects	371	850
Majority-male condition	1 woman, 4 men	2 women, 4 men
Majority-female condition	3 women, 2 men	4 women, 2 men
Crossed intervention	None	Gender of group leader

Note: Because they are the primary focus of this study, groups and individuals are restricted to mixed-gender groups only. Individual subjects are further restricted to only those who consented for their individual-level data to be included in our study. In Study 1, 2 participants did not consent, and in Study 2, 18 participants did not consent. Patterns of consent are not correlated with the experimental conditions. In addition, two majority-male groups in Study 1 included only 4 members (1 woman and 3 men), and two groups in Study 2 included 5 members.

A1, the men and women in our sample both came to the program with impressive academic credentials, and there are no statistically significant differences either within genders and across conditions or across genders in academic performance or demographic characteristics. If anything, women came to the study *more* qualified to exercise leadership — defined as having held a position in high school student government — than the men who participated. Across a wide variety of background characteristics collected prior to group assignment via administrative data and a baseline survey (see Appendix Tables A2 and A3 for a description of all variables and their summary statistics), we find no within-gender differences between the men and women assigned to the various treatment conditions.

Compared to Study 1, our Study 2 sample consists of students who were considerably younger and less likely to be international students or to be married. This is consistent with Study 2 being fielded in a required general education course, where the vast majority of students are college freshmen. As in Study 1, our randomization was successful, as we observe few differences across conditions. Only one measure, political affiliation, is statistically

significantly different for the men across conditions, with men in majority male groups being slightly less conservative than men in majority female groups.[13] Appendix Table A1 also reports balance tests for our combined sample of 1,221 students and confirms no within-gender differences between the men and women assigned to the various treatment conditions. However, we also report covariate-adjusted results in appendix tables as a demonstration of robustness.

In terms of statistical power, Study 2 collected a sample sufficient to detect effect sizes smaller than those observed in Study 1. This required a longer-term data collection effort and resulted in Study 2 containing nearly twice the number of groups (145) that were in Study 1 (75). Appendix Table A5 shows the approximate sample size needed to power the replication of effect sizes found in Study 1 for our primary outcomes. For all outcomes, the sample size of Study 2 was more than sufficient to power the detection of effect sizes of the magnitudes found in Study 1.[14]

3 Results

3.1 Women's Empowerment and Group Gender Composition

The first outcome measure we explore relates to perceptions of influence. Perceptions matter in a variety of contexts, including workplace evaluations, which often carry salary and career implications. Hence, we explore whether group gender composition matters for perceptions of who is most and least influential in a group. At the conclusion of both the "Lost on the Moon" and "Lost at Sea" team-building exercises, we conducted an exit survey that asked

[13] If anything, this difference would work to mute the differences between the conditions, at least to the extent there is a strong correlation between conservative political affiliation and more traditional gender norms.

[14] While the full Study 2 sample includes 145 groups, we were unable to collect data on the team-building task for 22 groups during the first semester of Study 2. Because three of our four main outcomes are measured during team building tasks, we exclude these groups when comparing key results, resulting in our effective Study 2 sample consisting of 123 groups in most of the analyses.

participants to identify the team member who was the most influential and the team member who was least influential in the decision-making deliberations they had just experienced. We also asked team members to make similar judgments about who was most "supportive" as the group members completed the task. We expect women's disempowerment to be most pronounced in judgments that involve authority and influence, where prejudices about women's leadership will be most relevant. Thus, we expect that group gender composition will matter for judgments about women's influence, but not necessarily for supportiveness, which involves characteristics like serving and helping the group, but not leading it.

Beyond reporting these perceptions, group members also made an incentivized choice for whom to select as spokesperson for their group. The spokesperson would have a chance to earn another $50 for the team by doing a short presentation defending their group's decision. If the influence votes were perceived as a low-stakes evaluation of group members, the selection of spokesperson was not. With an additional $50 at stake, each group member was incentivized to vote for the most competent member of the group. In the spokesperson decision, participants voted individually and without notice or prior deliberation as part of the task exit survey. Every group member was listed on the "ballot," so there was no nomination stage in which people could express that they were not interested in being spokesperson.

Results from the team-building exercise are helpful because group members' perceptions were recorded immediately after their interactions, arguably providing greater measurement precision. However, team members also completed monthly surveys about their group interactions, including an assessment of who had been the most and least influential "during group discussion and collaboration." These are not reports of perceptions about a specific task, but rather are meant to capture group dynamics more broadly. Conceptually, we view these more "global" assessments of group leadership akin to the kind of peer review one might get in a "360-degree review" or when colleagues make decisions about who they would like to

work with in the future.[15]

Table 2 reports the experimental differences between the majority male and majority female conditions for four variables — most and least influential votes for the team-building task, incentivized spokesperson votes for the team-building task, and most influential votes for the monthly global assessment surveys. Specifically, we estimate the following equation using OLS:

$$Y_g = \alpha + Maj_M\beta + \epsilon_g \tag{1}$$

The analysis is conducted at the group level, and the dependent variable is the proportion of votes for women divided by the proportion of women in group g. This quantity yields a measure of "votes per woman," and by normalizing in this way, we can directly compare across the two experimental conditions. Maj_M captures the gender composition of the group — specifically, the effect of being in a majority male group. The coefficient of interest is β, which represents the difference between the point estimate for the majority and minority male groups. In Study 1, the dependent variable is averaged across both team-building tasks — the "Lost on the Moon" task conducted at the beginning of the semester and the "Lost at Sea" task completed at the end.[16] Study 2 results include only the early semester task, given that a second experimental treatment designed to boost women's authority in some groups occurred prior to the second task, which we discuss further below. For the pooled sample analysis, we combine the two group-level datasets.

Table 2 provides powerful evidence of a large and statistically significant treatment effect. Women in majority-male teams were dramatically less likely to be seen as the most influential member of their team, even after accounting for their low numbers in the group. For example, in Study 1, the effect on votes per woman of being in a majority male group is approximately

[15]In the analysis reported below, we use the first two months only because this choice allows for the most direct comparison between the studies. The formal leadership treatment in Study 2 complicates comparisons in the later months. Nonetheless, we will also detail how responses changed after the leadership treatment in Study 2, and we will detail change over time in Study 1 using all four months below.

[16]Study 1 results are essentially identical for each task separately.

Table 2: Effects of Group Gender Composition on Women's Influence

	Study 1 (N=75)	Study 2 (N=123)	Studies 1 + 2 (N=198)
Team-Building Most Influential Votes	-0.437***	-0.292***	-0.346***
	(0.133)	(0.109)	(0.085)
Mean of outcome	0.67	0.58	0.61
St. dev. of outcome	0.62	0.62	0.62
Team-Building Incentivized Spokesperson Votes	-0.343*	-0.232**	-0.273***
	(0.176)	(0.098)	(0.092)
Mean of outcome	0.77	0.55	0.63
St. dev. of outcome	0.79	0.56	0.66
Global Assessment Most Influential Votes	-0.316*	-0.207*	-0.249***
	(0.159)	(0.108)	(0.090)
Mean of outcome	0.69	0.75	0.73
St. dev. of outcome	0.71	0.61	0.65
Team-Building Least Influential Votes	0.675**	0.044	0.283**
	(0.256)	(0.144)	(0.134)
Mean of outcome	1.35	1.21	1.26
St. dev. of outcome	1.18	0.81	0.97

Note: Table presents group-level analysis of the effects of majority male groups on perceptions of influence. In each case, the dependent variable is votes for women divided by the proportion of women in the group. All estimates are robust to the inclusion of a vector of controls (see Appendix Table A6). Study 2 analysis excludes 22 groups from the semester when the team-building task was not conducted. Global assessment votes also exclude those groups for purposes of comparison, but the substantive pattern does not change if the groups are included. In both studies, results for global assessment votes are restricted to the first two months of assessments, which are the assessments prior to the leader intervention for Study 2. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels are indicated by $*<.1$, $**<.05$, $***<.01$.

70 percent of the standard deviation of the outcome variable. The effect is somewhat smaller in Study 2, as might be expected given that the inclusion of two women in 6-person groups was designed to be a harder test, but still statistically robust and nearly half of the standard deviation of the dependent variable. Women in majority male groups were also substantially less likely to be chosen as spokesperson, a choice where the stakes were higher because group members had a substantial incentive to choose the best group member. These results are marginally significant in Study 1, but $p < 0.05$ in Study 2, and $p < 0.01$ when the two studies are combined. The fact that women were also less likely to be chosen when the choice was

consequential and incentivized suggests our attitudinal findings reflect real authority deficits. Insofar as this kind of selection process mirrors opportunities for visibility and promotion in the workplace — for example, presenting findings to a boss or client or being chosen as the "lead" on a project — this result is especially troubling.[17]

The effects of group gender composition are not limited to the discrete team-building tasks but can also be found in broader evaluations of group functioning. In the first two months of global assessments of influence in the group, women in majority male teams were less likely than women in majority female teams to be seen as the group's most influential member. These treatment effects are marginally significant in each study separately, but with the increased power available when combining both studies, $p < 0.01$. These patterns represent women's disempowerment not only in an evaluation of a specific task, but also in more general impressions of women's leadership within the group, judgments that may be analogous to a variety of workplace evaluations and performance reviews.

For both team-building tasks and global assessments, the patterns shown in Table 2 are robust to a variety of analytical choices and model specifications; for example, it holds with the inclusion of controls for group-level characteristics (Appendix Table A6), with weights designed to replicate levels of sexism in a representative sample of the United States (Appendix Table A7), and with corrections for multiple hypothesis testing (Appendix Table

[17]Female-majority groups perform slightly worse in the team-building task, but the magnitude of the difference is substantively small. Moreover, this difference is not replicated in a graded group project that occurred in Study 2. At the end of the semester, Study 2 teams completed a writing assignment that was evaluated by two independent graders who were trained by a professional consultant/educator of advanced writing and English composition. The consultant and the graders were unaware of the nature of the study. Scores were assigned both to the individual parts of the assignment that were submitted by each group member, as well as the overall group grade. Students first divided responsibilities for the project to complete independently, and then the individual inputs were used to create a single final submission for the group. We therefore have measures of individual student performance as well as group performance on this incentivized task. In addition, we have course grades for participants in both studies, and that measure also yields no evidence that majority male groups out-performed majority female groups. Appendix Table A9 thus shows that majority-women groups are not consistently lower performing across all types of tasks, and group performance may or may not be affected by the gender composition of the team, depending on the nature of the task.

A8).[18]

As expected, the treatment works in reverse for the least influential votes in Study 1. This pattern indicates that women's disempowerment does not occur among group members at the "top" of the distribution only, but is spread throughout the entire team. In Study 2, the treatment effect on least influential votes does not replicate ($p = 0.77$), though importantly, this is not because women in majority male groups are somehow empowered. Instead, women's predicted share of least influential votes meets or exceeds expectations in *both* conditions (see Appendix Figure A2 for details).

The profound nature of women's disempowerment can also be represented visually. Figure 1 reports the team-building task influence, spokesperson, and global assessment results for both studies, separately and aggregated together.[19] The horizontal dashed line in Figure 1 represents how many votes women in that treatment condition would receive if they were selected in proportion to their presence in the group. For the five-person groups in Study 1, this would be 20% for the female minority condition and 60% for the female majority condition. In Study 2's 6-person groups, the dashed line represents 33% for the male majority condition and 66% for the female majority condition. Normalizing in this way allows for easier comparison across treatment conditions and across studies. The variable of interest is equivalent to the number of votes per woman in the group — a number that would equal one if influence were distributed randomly. In these figures, both the comparisons across treatment conditions (which represent the causally identified experimental effects) and the

[18] In Appendix Table A6, we present the results of the following equation:

$$Y_g = \alpha + Maj_M\beta + X_g\gamma + \epsilon_g, \tag{2}$$

where X is a vector of controls, including average GPA, average age, race (proportion Nonwhite), marital status (number of married team members), and the presence of an international student. The inclusion of controls tends to sharpen the precision of the estimates and sometimes increases effect sizes. Effects are always in the same direction as models without controls. Details of the weighting decisions (Table A7) and multiple hypothesis testing (Table A8) can be found in the appendix.

[19] These estimates are generated from the results in Table 2. The companion figure for task-based least influential votes can be found in the Appendix (see Appendix Figure A2).

comparison to the "line of equality" where women receive influence proportional to their numbers are important ways of evaluating women's empowerment.[20]

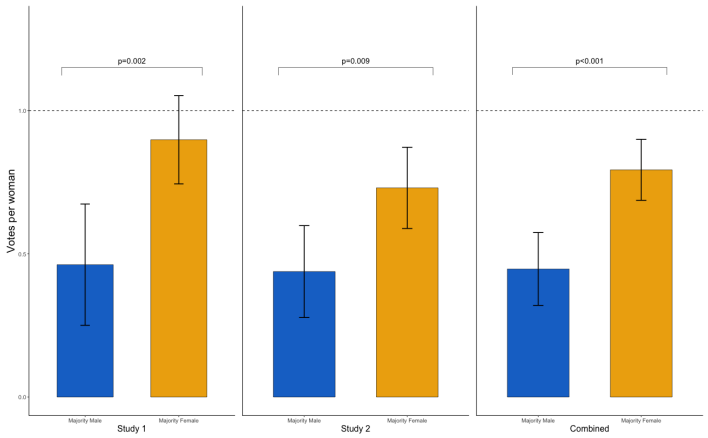
Figure 1 shows that the pattern of results is very similar across dependent variables and replicates across studies. Women's influence in male-majority teams was substantially lower than in female-majority teams and substantially lower than the line of equality. For example, as can be seen in Panel B, the lone women in Study 1's majority male teams received about .46 fewer spokesperson votes per woman than women in the majority female groups. This rate of votes received is only about 55 percent of the baseline expectation. By comparison, members of majority women teams chose women as most influential about as often as expected given their proportion in the group. Similarly substantial influence gaps are present for all three dependent variables.

Women's disempowerment in groups where they were outnumbered replicates in Study 2. Again we find that team members evaluated women in the majority male condition as far less influential than their numbers in the group would lead us to expect. Women in majority female teams were comparatively more influential — though, importantly, in all three measures, neither the point estimate nor the confidence interval for majority female groups crosses the dashed line, meaning that women in those groups were still perceived as less influential than if influence were distributed randomly. When both studies are combined, treatment effects remain robust. Results across both studies yield powerful evidence, then, that group gender composition affects women's influence in discussion and decision-making.

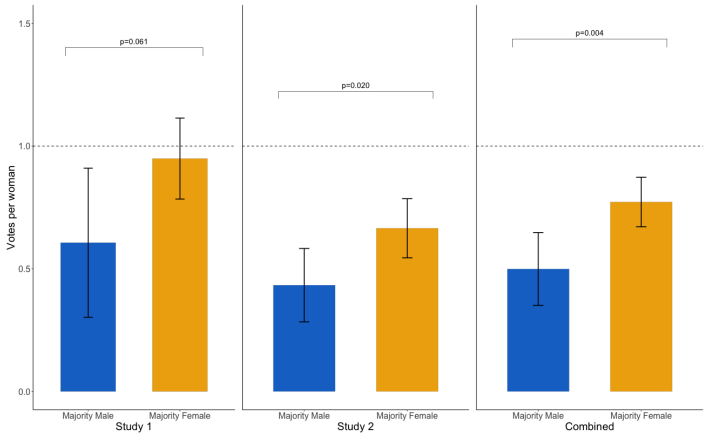
Notably, we do not find the same kind of group gender composition effects when we ask participants to identify the most *supportive* members of their group. As Panel C of Appendix Figure A2 shows, women are perceived as supportive at the rates we would expect in both conditions, and in Study 2, they were regarded as *more* supportive than expected in the

[20]Arguably, the "line of equality" is a minimal standard: merely votes proportional to numerical presence in the group.

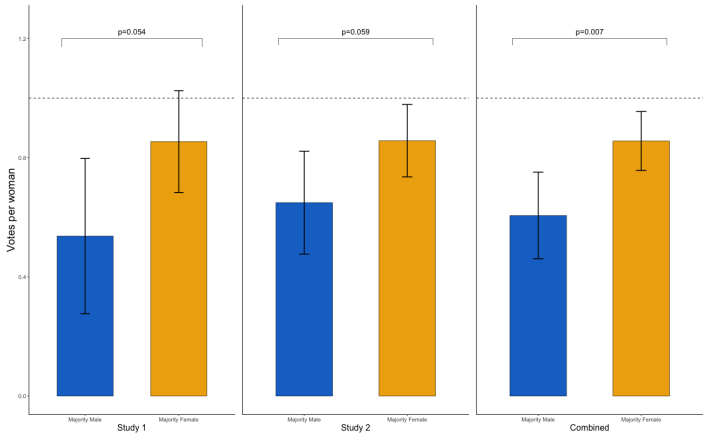
Figure 1: Women's Influence and Authority by Group Gender Composition



(a) Team-Building Task: Most Influential Votes



(b) Team-Building Task: Incentivized Spokesperson Votes



(c) Global Assessment: Most Influential Votes

Note: 95% confidence intervals. Bars represent the average proportion of votes for women divided by the proportion of women in the group. (See Table 2 for full results.)

minority female condition.[21] Thus, women's disadvantage occurs where we expect it — for assessments about traditionally "masculine" leadership traits of influence and authority — and evaporates for judgments about service and supporting roles in the group.[22]

Finally, the structure of our team-building task also allows us to measure empirical influence on group decisions by identifying who was able to pull the group outcome closest to their initial ranking. We operationalize empirical influence by identifying which individual in the group had the smallest total difference between their pre-deliberation ranking of items on the "Lost on the Moon" or "Lost at Sea" task and the final group outcome.[23] This is a proxy for having the most actual influence in deliberation and allows us to construct a group-level measure of whether a woman exercised the most influence during the group discussion.

We find that women in majority-male groups are significantly less likely than men to pull the group's decision in their direction. Specifically, in Study 2, where the gender composition conditions are parallel, the likelihood that a woman in a majority-male group is the most empirically influential group member (including ties) is .39, while it is .80 for a man in the same condition. The difference is large and statistically significant ($p < 0.001$, two-tailed difference-of-means test). In contrast, when women are in majority-female groups, they exert similar levels of empirical influence, relative to men (.64 vs. .56, $p = 0.27$).[24] The difference-in-differences in empirical influence across the two conditions also favors men ($p = 0.002$), meaning that women's influence deficit is greater than men's in conditions of similar numerical disadvantage. Hence, across multiple measures of both perceived and empirical influence, we document a considerable deficit in women's influence and authority

[21]The fact that the dynamics for supportiveness are quite different also provides evidence that women's disadvantage in perceived influence is not merely a "mechanical" function of their numerical disadvantage, but rather a substantive judgment.

[22]This is consistent with prior studies which have found a close cultural tie between authority and masculinity and supportiveness and femininity (Jamieson et al., 1995; Manne, 2017).

[23]In the case of ties, more than one group member could be most influential.

[24]We show in Appendix Table A10 that this result cannot be explained by women in male majority groups being either more incorrect (Panel A), or closer to the implied group average from initial rankings (Panel B).

when they are a minority in the group.

3.2 Gender vs. Numerical Disadvantage

One important question emerging from the analysis we have presented to this point is whether women's influence deficit in teams where they are outnumbered by men is merely the disadvantage that *any* numerical minority would face or whether this disadvantage is unique to women. In other words, is this disadvantage driven simply by minority status and not related to gender *per se*? To address this question, we turn first to individual-level analysis of Study 1's two team-building exercises to analyze how often a woman was chosen as the group's most influential member.[25] We find that in groups with one woman, women received the most influential votes 9.7 percent of the time across the two team-building exercises, and they received the spokesperson votes only 11 percent of the time. These results clearly fall well below the 20 percent we would expect if influence and spokesperson votes were distributed at random, confirming the stark disadvantage we have already demonstrated.

But is this disempowerment similar to, or different from, what any outnumbered individual in the group would experience? Unfortunately, because of the gender composition of students enrolled in our partner program in Study 1, it was impossible to assign perfectly symmetric conditions. Thus, we conduct a placebo analysis using the 27 all-male teams, randomly choosing one of the men to be a "placebo" team member and point of comparison for each task.[26] Given the composition of the teams, our placebo team member worked with four men, just as the women in Study 1's majority male teams did. We then determined whether the placebo team member received the most influence votes in the group, repeating this process of randomization (with replacement) 1,500 times. The distribution of results from

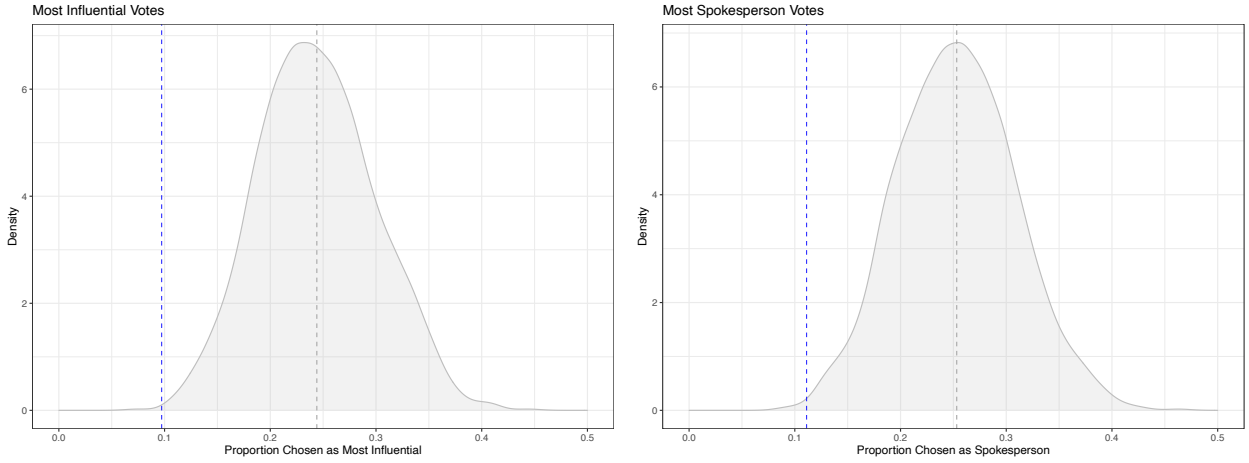
[25]For this analysis, we stacked the two team-building exercises in a single dataset. The dependent variable is a dichotomous indicator of whether the group member received the "most influential" vote. In the case of ties, multiple group members could be "most influential."

[26]This analysis includes only the 5-member all-male teams in which all group members consented to the use of their individual-level data.

these iterations can be seen in Figure 2, where Panel A shows results for influence votes and Panel B presents spokesperson votes. The mean proportion for the distribution can be seen in the gray dashed line. For influence votes, 24.4 percent of randomly chosen placebo men received the most influence votes, and 25.3 percent received the most spokesperson votes, on average.[27]

The key piece of evidence in this analysis is the relationship between the dashed blue line, which is the proportion of women chosen as most influential or spokesperson, and the distribution for our placebo men. In both cases, our estimate for women working with four men falls in the extreme tails of the distribution — below the cutoff for the first percentile for both measures. These estimates are more than 2.5 standard deviations away from the means for the distribution of men. Put simply, the results suggest that it would be extremely rare for a random man working with four other men to experience anything close to the disadvantage women face in groups with four men.

Figure 2: Placebo Test of Women's Influence and Authority



(a) Team-Building Task Influence (b) Incentivized Spokesperson Votes

Note: The dashed blue line represents the proportion of women chosen as most influential or spokesperson in Study 1's majority male groups. The dashed gray line represents the mean proportion of randomly chosen placebo group members chosen as most influential or as spokesperson. The gray distribution represents the results of 1500 iterations of the placebo test.

[27]These numbers are slightly above 20 percent because ties are possible.

In Study 2, the research design allows for a direct comparison of men's and women's influence in identical conditions of numerical disadvantage, and we can assess whether men face a disadvantage when outnumbered by women. Again, the results are clear: we find substantial evidence of women's disproportionately low influence relative to men, but no similar pattern for men outnumbered by women. Only 7 percent of women in majority male teams received the most influence votes, compared to 29 percent of men in majority female teams. In the incentivized spokesperson choice, only 9 percent of women in majority male groups were chosen as spokesperson, compared to 37 percent of men in majority female groups. In both cases, the differences between the genders are statistically significant ($p < 0.0001$) and substantial.[28] Even when we compare men and women in the exact same conditions of numerical disadvantage, women experience dramatically lower levels of perceived authority, while men face little to no influence deficit.

Hence, across multiple measures, analytic approaches, and studies, results offer strong support for the idea that women in majority male groups were seen as much less influential in group deliberation than 1) women in majority female groups, 2) than men working with other men, or 3) than men working in teams where women predominate. Team members in groups where men outnumbered women only rarely judged women as the most influential member of the group and tended to see them instead as the group's least influential member, even after we adjust for women's share in the group. For example, out of all the women in majority male teams in Study 1, *none* were chosen as most influential in both of their group's team-building exercises, compared to 22 out of 154 men in that condition.[29] By contrast, team members perceived women in majority female groups as comparatively much more

[28]The p-values are computed from a two-tailed difference-of-means test, using individual-level data from the first team-building task in Study 2.

[29]If influence votes were allocated at random, we would expect approximately 36 percent of participants in majority male groups to choose a woman as most influential at least once across the two team-building exercises in Study 1. (The expected value is $1 - (0.8 * 0.8) = 36\%$.) Instead, only 18% of both men and women actually voted for a woman at least once — only half the expected value.

influential (though in Study 2, such increased influence did not always reach the standard of equality).

3.3 Effects of Formal Leadership Assignment

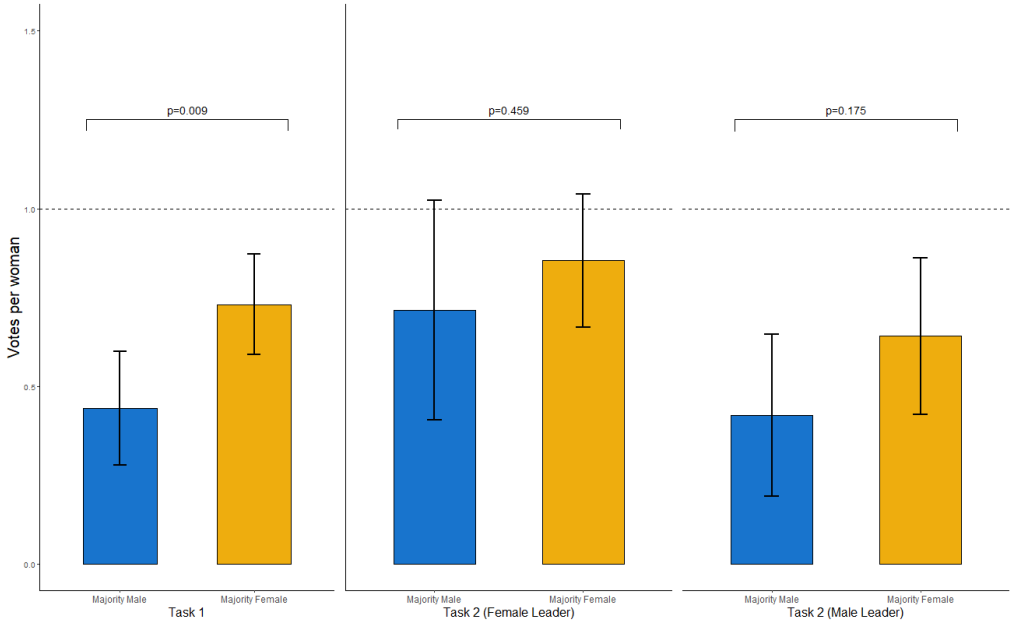
Existing literature suggests that organizational hierarchy and the gender of team leaders might matter for women's experiences (Sahin et al., 2015; List and Rondeau, 2003; Kofoed et al., 2019; Calder-Wang and Gompers, 2021). One of the virtues of Study 2 is that it offers a second experimental intervention designed to investigate how formal designations of authority affect women's perceived influence in group deliberations. Recall that halfway through the semester, we randomly assigned each group to receive either a male or female leader. Within each randomly chosen leader condition, we then randomly chose one man or woman in the group to serve as the leader. Responsibilities were not onerous and largely administrative — leaders ensured that the group stayed organized and that assignments were submitted properly — although being a group leader was nonetheless a grant of formal authority within the team by course instructors.

Figure 3 highlights the effects of the leader assignment on both perceived influence (Panel A) and incentivized spokesperson votes (Panel B) during Study 2's team-building tasks. The leftmost figure in each panel illustrates votes per woman by experimental condition in the first team-building task, which occurred prior to the leader assignment.[30] As we have shown, women in majority male groups experienced a dramatic influence deficit, both relative to the line of equality and relative to women in majority female groups.

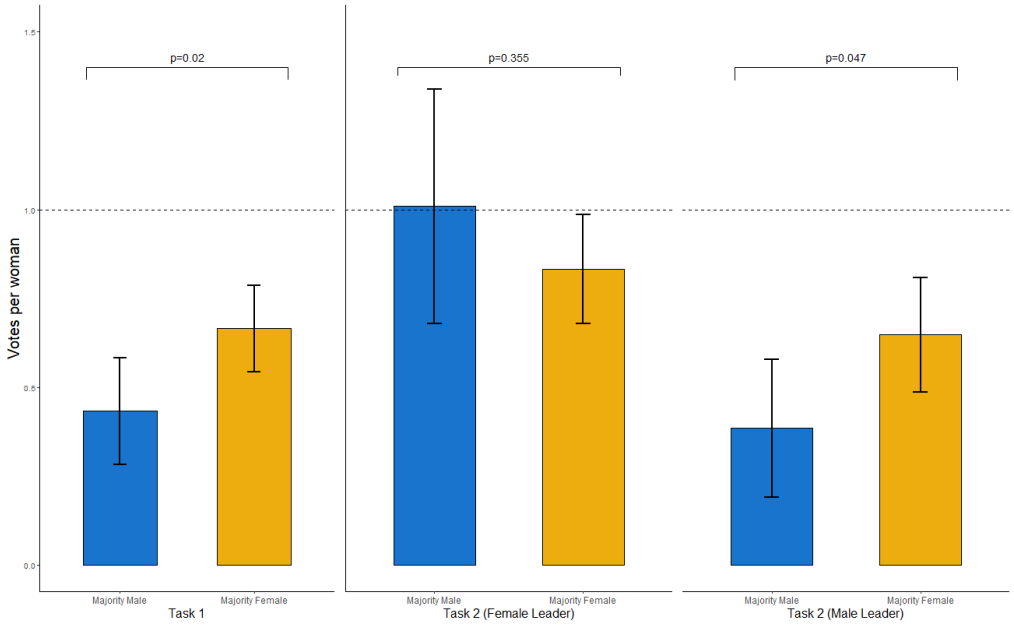
The bars to the right of Task 1 in Figure 3 show what occurred in Task 2 after groups were assigned a female or male leader. We highlight three striking patterns. First, in groups with a female leader, the difference in votes per woman between the majority male and majority female condition narrows dramatically and is no longer statistically significant

[30]These results are identical to what we showed for Study 2 in Figure 1.

Figure 3: The Effects of Formal Leadership on Women's Influence



(a) Perceived Team-Building Influence



(b) Team-Building Task Spokesperson Votes

Note: 95% confidence intervals. Bars represent the average proportion of votes for women divided by the proportion of women in the group. (See Appendix Table A11 for regression results.)

($p = 0.459$ for perceived influence and $p = 0.355$ for spokesperson votes). The presence of a female leader moderated and essentially eliminated the effect of group gender composition. In regression analysis of both perceived influence and spokesperson votes, the effect of gender composition is large and statistically significant in Task 1, when there was no leader, but disappears in Task 2 when the effect of a female determines differences in outcomes (see Appendix Table A11 for details). Again, these results are robust to multiple hypothesis testing corrections (Appendix Table A8).

Second, this change occurred because of a large and statistically significant increase in women's influence in majority-male groups with a female leader. The difference between votes per woman in Task 1 majority male groups with no leader and votes in Task 2 majority male groups with a female leader is significant at $p < 0.05$ for both perceived influence and spokesperson votes.[31] In majority-male teams with a female leader, the confidence interval for perceived influence crosses the line of equality, while for spokesperson votes, the point estimate exceeds the estimate for majority-female groups and also achieves equality. Granting formal authority to a female team member therefore starkly increased women's influence.

Finally, the Task 2 results for groups assigned a male leader are indistinguishable from the Task 1 results. None of the differences from Task 1 to Task 2 groups with a male leader are statistically significant.[32] These dynamics are consistent with the notion that male leadership is the default norm in the absence of a formal grant of authority to women in the group and are in line with the previous experimental findings that women avoid applying for jobs at firms that tout their "flat" organizational hierarchy (Hurst et al., 2022).

[31]Women in majority female groups also appeared to experience an increase in votes per women. Because women's influence in majority-female groups was already higher prior to leader assignment, this increase was smaller and not statistically significant in the case of perceived influence. For spokesperson votes, the increase is significant at p = 0.055. The analysis in Figure 3 uses all Task 1 groups as a point of comparison, but if we compare Task 2 groups with themselves prior to the assignment of a leader, all comparisons are in the same direction, though statistical power is more limited. Even so, the paired difference-of-means between Task 1 and Task 2 in spokesperson votes per woman is still significant at $p = 0.009$ for majority-male teams.

[32]In Study 1, with no leader assigned, we find no change between Task 1 and Task 2. See Appendix Figure A3.

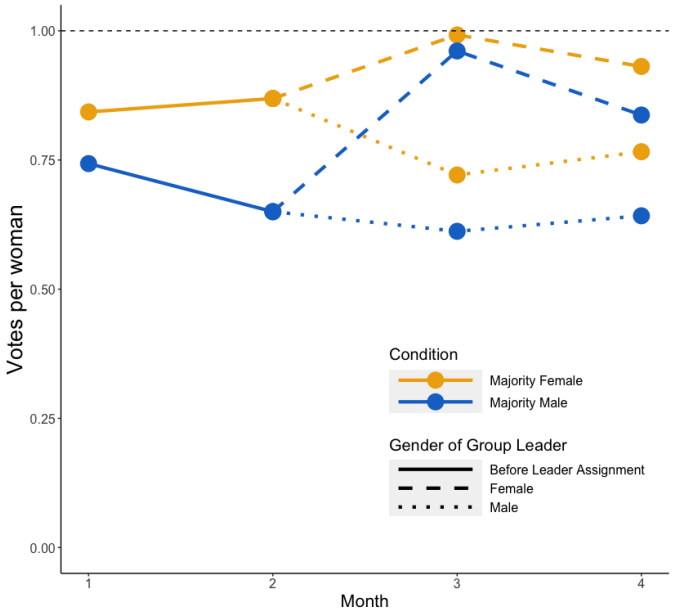
Whose perceived leadership was enhanced in groups where a woman was randomly chosen to be the formal leader? Here we find a difference between group judgments about their spokesperson and votes for the "most influential" member in the team-building task. As summarized in Appendix Table A12, in the choice of group spokesperson, more than three-quarters of the votes for women went to the group's formal leader (84% in majority male groups and 76% in majority female groups). By contrast, for "most influential" judgments, a little less than half of women's votes went to the chosen leader (48% in majority male groups and 44% in majority female groups), while women other than the chosen leader received a bare majority of the votes. In other words, the formal leader herself was empowered in the votes for spokesperson, but with respect to judgments about influence in the team-building task, the formal leader was not the only female beneficiary of women's leadership. This evidence is consistent with positive spillover effects in judgments of influence to other women on the team.

3.4 Over Time Global Assessments

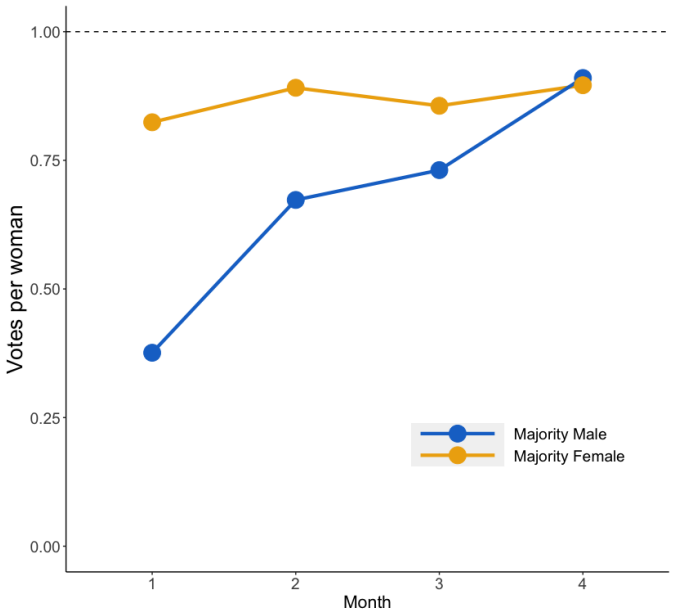
Another unique feature of our research design is the ability to track outcomes over time. Real-world deliberative settings like workplaces, schools, and political bodies generally include people who work together repeatedly on a variety of tasks and get to know each other's strengths and weaknesses. It is possible that the deficits of authority and influence women in majority male groups experience are largely the result of early assessments based on stereotypes before individuals have other information on which to judge their teammates (Bohren et al., 2019; Crowder-Meyer et al., 2020). On the other hand, there are some reasons to believe that time may not result in improvements and may even exacerbate the problem. For example, social psychologists find that familiarity really can breed contempt among those who are dissimilar (Norton et al., 2013), and sociologists show that network dynamics tend to reinforce homophily and social inequality over time (DiMaggio and Garip, 2012; Kossinets

and Watts, 2009).

Figure 4: Global Assessments Over Time



(a) Study 2: Time and Formal Leadership



(b) Study 1: Trends Over Time

Note: Points represent the average votes per woman in each month and experimental condition. (See Appendix Tables A13 and A14 for regression results.)

In this section, we disaggregate the global assessment results over time to examine how they changed each month. The top panel of Figure 4 displays the results for Study 2, and the bottom panel presents the results for Study 1.[33] The results are mixed. In Study 2, we find no improvement in women's influence between the first and second months of global assessments. In both months, there is a statistically significant difference in votes per woman between majority female and majority male groups, and if anything, the gender composition gap increases in the second month. In the third and fourth months, patterns were primarily driven by the leader intervention, which again sharply moderated the effects of gender composition. Regardless of the team's gender composition, groups assigned a female leader essentially met the standard of equality.[34] Women's influence in majority-male teams that were assigned a male leader remained at the same level as it was prior to the leader assignment, significantly below the line of equality ($p = 0.004$). Women's influence in majority-female teams with a male leader also dipped well below equality ($p = 0.008$). In a regression analysis, female leadership strongly predicts women's influence ($b = 0.31, p = 0.01$), while gender composition does not ($b = -0.07, p = 0.57$).[35] In the last month of the study, these patterns attenuated somewhat, but formal leadership still appears to matter more than gender composition.[36]

In Study 1, where no leader assignment interrupted the time series, we do see marked improvement for women in majority male groups. In the first month, women in majority male groups receive substantially fewer votes per woman (0.38) than women in majority female groups (0.82, $p < 0.001$). This disparity narrows each month, and by December, there

[33]For ease of interpretation, we omit confidence intervals from these figures, but discuss patterns of statistical significance in the text.

[34]When testing the hypothesis that the observed mean equals one in the first month after the leader assignment, $p = 0.92$ for majority-female teams and $p = 0.82$ for majority-male teams. In the second month after leader assignment, the pattern persists with $p = 0.28$ and $p = 0.25$, respectively.

[35]See Appendix Table A13 for details.

[36]In the final global assessment, women in teams with a female leader still received more votes per woman than women in teams with a male leader, regardless of the team's gender composition. In regression analysis for the last month (Appendix Table A13), the effect of female leadership remains positive and marginally significant ($b = 0.18, p = 0.08$), while gender composition has no statistically significant effect, though it is in the expected direction ($b = -0.11, p = 0.29$).

is no difference in women's influence between the conditions ($p = 0.95$).[37] In these final rankings, the confidence intervals for both majority male and majority female groups cross the gender equality mark, though the point estimates are just below it. Notably, we do not find evidence of this same improvement in discrete judgments about influence in the second team-building task.[38] More research is necessary to understand the potential divergence between discrete task judgments and global assessments, as well as the differences between studies. One possible theory for why Study 1 global assessments yield a different pattern over time than Study 2 is that the women in Study 1 teams were uniformly extremely highly qualified for admission to the highly competitive accounting program, and their high levels of achievement became difficult for group members to ignore, at least in global assessments. But it is also possible that Study 1's trend in global assessments is merely an anomaly, given that we find no evidence of change in other measures or across studies.

3.5 Sources of Women's Influence Deficit

Our results thus far provide strong, replicable experimental evidence that in the absence of a female team leader, women in the numerical minority experience a significant disadvantage in terms of influence and authority in group settings relative to men in similar circumstances, to women in majority-female groups, and to the standard of equal influence. But what is the source of women's influence deficit in groups without a formal leader? To answer this question, we decompose the voting patterns in our data to examine how men and women distribute their votes to themselves and to others.

Figure 5 presents team building task "most influential" votes separately for men and women across conditions for both studies combined.[39] Unlike the group-level models presented

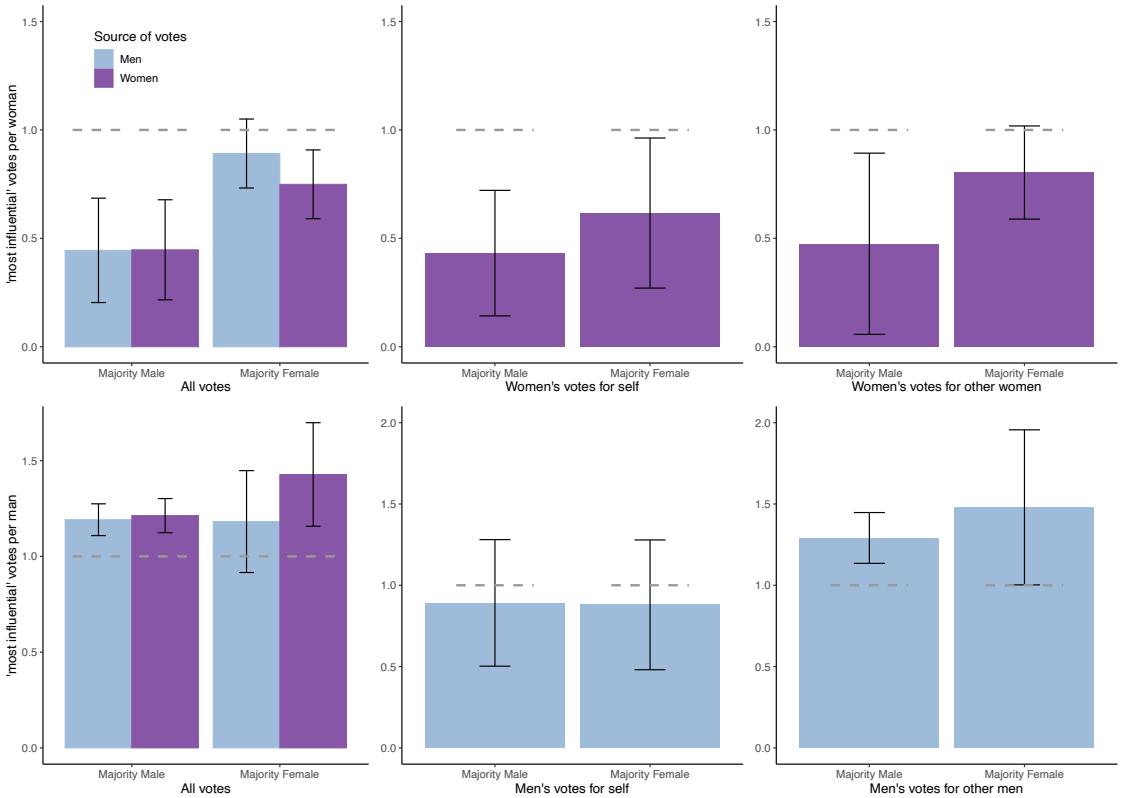
[37] See Appendix Table A14 for details.

[38] See Appendix Figure A3.

[39] We present these analyses disaggregated by studies in Appendix Figure A4 which shows highly consistent patterns. See also Appendix Figure A5 for this analysis using non-normalized fraction of votes as the outcome measure.

in Table 2, this analysis proceeds at the individual level, with a stacked dataset that includes each team-building task from Study 1 separately and the first team-building task for Study 2.[40] The top panel presents the effect of the conditions on influence votes per woman, while the bottom panels show influence votes per man. The left panel in each row highlights the sources of total votes per man or woman, while the other two panels disaggregate the voting patterns to show how each gender's totals are related to self-votes and to votes from fellow women (or men). The dashed line in each figure again reflects the standard of equality.

Figure 5: Sources of undervoting for women: Votes per man and woman, by sex of vote grantor and across conditions



Note: 95% confidence intervals. In the top panel, bars represent the average proportion of votes for women divided by the proportion of women in the group. In the bottom panel, bars represent the average proportion of votes for men divided by the proportion of men in the group. Because the lone women in the majority male groups in Study 1 had no opportunity to vote for other women, the relevant bar in the top right panel is generated only from Study 2, where there were two women in the group.

[40] We omit the second team-building task in Study 2 to avoid conflation with the leader intervention. Note that confidence intervals are constructed using standard errors clustered by group.

The top left panel shows that women's disadvantage in majority male groups can be traced to the behavior of both men and women — both sexes voted for women at essentially the identical rate, one far lower than what equal influence would require. Women's corresponding increase in influence in majority female groups is also a function of both men's and women's votes, with a slightly larger relative boost from men. In fact, the confidence interval for votes from men overlaps the line of equality in majority female groups.

The next two panels in the top row of Figure 5 show that women voted for themselves at rates lower than what equal influence would require,[41] and they were also less likely than expected to allocate their votes to other women.[42] While women's propensity to vote for themselves was largely unaffected by the gender composition of the group, women tended to perceive other women as more influential when they were in the majority, and the confidence interval for that estimate crosses the line of equality. In other words, our evidence is consistent with the possibility that women do not disadvantage other women in majority female groups.

The corresponding panels in the bottom row of Figure 5 demonstrate that men's behavior toward themselves and other men was quite different from women's. Men voted for themselves at essentially the expected rate in both experimental conditions, in contrast to women's low levels of self-voting. In addition, men allocated their votes to other men at rates significantly higher than expected, regardless of experimental condition. In other words, men both perceived themselves as influential and tended to *over-reward* other men with influence, even when they were the numerical minority in the group. To the extent that homophily contributes to patterns of influence, it appears that men's homophily is a key factor. Overall, the analysis of vote allocation provides additional direct and novel evidence that men never experience the same influence gap as women, even when men are in the numerical minority.

[41]This result is consistent with previous work. See, for example, the Exley and Kessler (2022) study documenting gender differences in self-promotion.

[42]Because the lone women in the majority male groups in Study 1 had no opportunity to vote for other women, the relevant bar in the top right panel is generated only from Study 2, where there were two women in the group.

Men's influence advantage does not come from men alone, however. Strikingly, as the bottom left panel of the figure shows, both men *and* women contributed: regardless of whether men were in the minority or in the majority in their group, they were voted as most influential at a *higher* rate than what one would expect if the votes were cast at random. Both the point estimates and confidence intervals for women's votes for men exceed the line of equality, including in groups where women were in the majority. Men's relatively high rate of influence in majority women teams can be traced, in part, to the fact that women disproportionately allocate their votes to men instead of to themselves or other women.

4 Separating Individual and Group Behavior: Model and Simulations

We next characterize our experimental results with the calibration of a simple model of voting in the presence of discrimination. Individuals form groups, which undertake a task in which each member of the group realizes a measure of performance $X_i \sim N(0,1)$. Performance is assumed to be perfectly observed by all members of the group without noise or uncertainty. Consistent with our finding about women's under-voting of themselves, we allow women to first assess their own performance with constant self-promotion penalty, d_{self}. Also following our results in the previous section, both women and men assess women's, but not men's, performance with penalty d. Voting then occurs based on each individual's penalized assessment of their teammates, which depends on one's gender and the relative performance of peers of different genders. For example, a woman will vote for herself only if her performance margin over a female second-best performer is greater than $d_{self} - d$; otherwise, she will vote for the female second-best performer. She will vote for herself over a male second-best performer if her performance margin over him is greater than d_{self}. Men will vote for a woman only if her performance margin is d or greater than the highest-performing man.

The two measures we construct to calibrate our empirical results are women's rate of self-voting and influence votes per woman. Using the number of groups per gender composition in our pooled experimental sample, we first recover the expected rate of self-voting among women under varying levels of d_{self}. The series in Appendix Figure A6 plots the quadratic fit from this simulation.[43] Since we observe self-voting directly in the data, we place the observed means of women's self-votes (from Figure 5) at their indicated heights, and calculate the implied rate of d_{self} in both conditions. The downward arrows indicate the implied d_{self} of 0.68 standard deviations in the male-majority condition, and 0.44 standard deviations in the female-majority condition. The implied self-penalty imposed by women is large — on average, greater than 0.5 standard deviations, and the self-penalty in either condition rejects self-voting at parity (see Figure 5). Across conditions, there is suggestive evidence of a reduction of approximately 28 percent of the self-promotion gap and 35 percent of the implied self-penalty, although these are not statistically distinguishable from each other (see A6).

In the subsequent step, we use these implied self-penalty rates as fixed parameters in the simulation of group-level influence votes per woman. Figure 6a plots the quadratic fit from the simulations results in series for each composition condition, which are hereafter referred to as "voting curves." Similar to the above, we plot the observed experimental mean values of votes per woman in each condition (y-axis) to recover the implied intensity of discrimination in the assessment of performance of women by men in each condition (x-axis). Figure 6a places the experimental means on the series across conditions and measures of team-building task influence, global influence, and incentivized spokesperson votes. The first finding is that the average level of discrimination applied to women in any measure or condition is quite

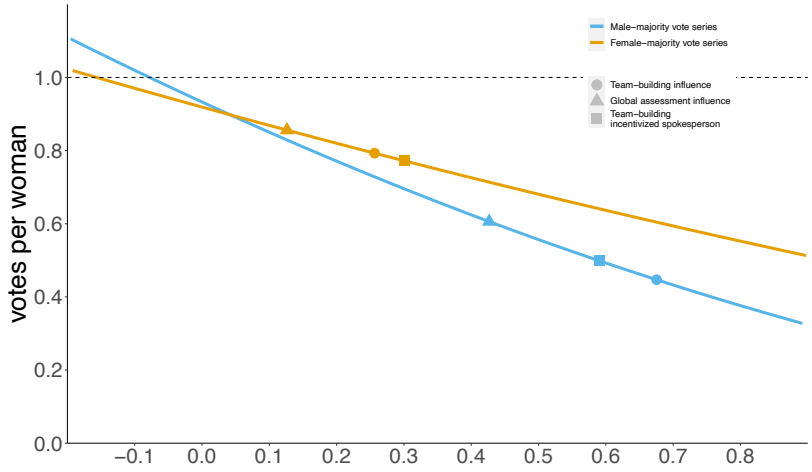
[43] For each value of d_{self}, we estimate 1,000 draws of performance for each individual and calculate the rate at which individuals vote for themselves among their group members after considering d_{self}. For self-voting, we normalize the share of votes to self in each draw by dividing by 1/N, whereby proportionate voting is indicated by a value of 1.

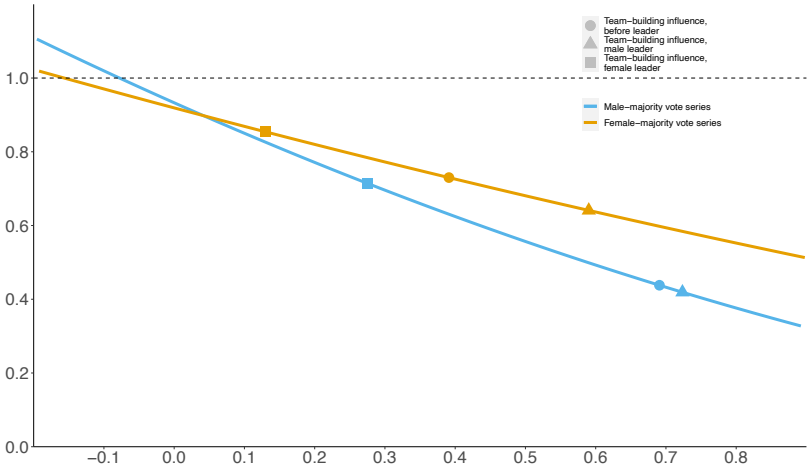
high: the average penalty on women in the team-building influence measure is 0.67 standard deviations in the male-majority condition, and 0.25 standard deviations in the female-majority condition. In practical terms, this means that a woman must outperform the best man in her group by two-thirds of a standard deviation in the male-majority condition, and one quarter of a standard deviation in the female-majority condition, to earn a teammates' vote.

Figure 6: Simulated voting rates, by condition and intensity of discrimination

(a) Perceived influence votes per woman

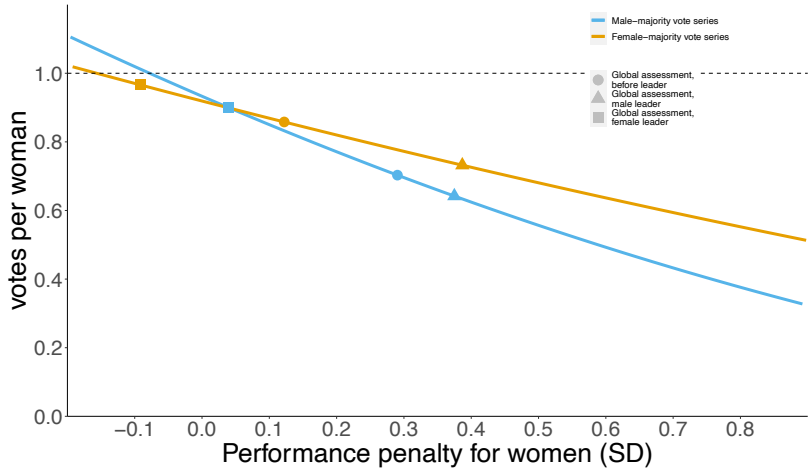
(b) Team-building influence votes per woman, by leader

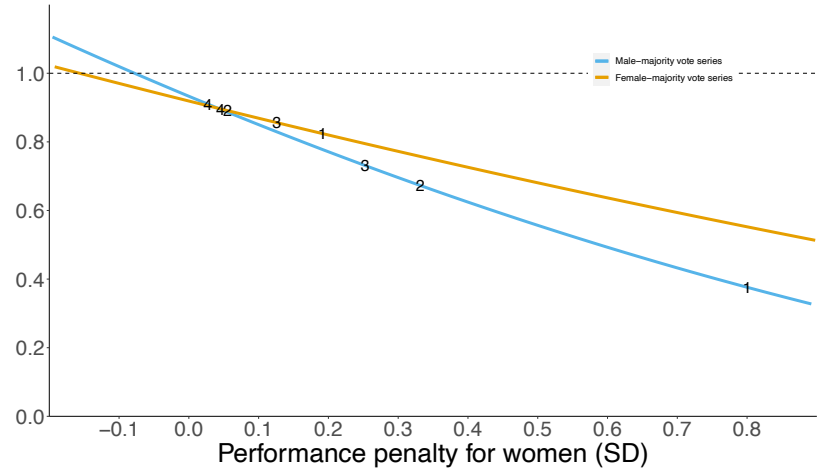




(c) Global influence votes per woman, by leader

(d) Study 1: Global influence votes per woman by month





Note: Chart presents results of the voting model calibration. The series are the fitted voting curves from 1,000 iterations of a simulation of the voting model described in Section 4 at each hundredth value of the indicated performance penalty. The indicated points placed on the curves reflect the same y-axis values as the corresponding quantities in Figures 1 and 3.

Level differences, recovered by analyses such as Table 2, reflect the average experience of women as they move across the voting curves, but do not characterize changes in the behavior of teammates. Thus, while the measure of votes per woman could increase when moving from majority-male to majority-female groups without any change in the underlying the intensity of the discrimination penalty faced by women (a *compositional* effect), the increases in votes per women seen in our study are sufficiently large to suggest a *behavioral* change in the intensity of discrimination expressed in the female-majority groups, compared to male-majority groups.

The next finding from this analysis relates to measurement: Figure 6a places three measures of influence along the voting curves: the team-building task influence, global influence, and the incentivized spokesperson votes. These measures exhibit different levels of discrimination even within condition, despite being recovered over the same time period, from the same groups, and in the same study context. All measures also exhibit a net reduction in the implied penalty across conditions, and these reductions are of relatively similar magnitude regardless of whether the measure is incentivized or specific to the discrete team-building task.[44]

The effects of our crossed leadership intervention on voting rates in Study 2 are plotted in Panels 6b and 6c. In these figures, we plot the voting rates from data collected both before and after the assignment of a formal leader. Because groups within each composition condition were assigned to male and female leaders, we can characterize the changes associated with the imposition of a leader as shifts along each voting curve instead of across voting curves. Focusing first on team-building task influence in Panel 6b, the male-majority condition exhibits a minimal change in the implied discrimination rate when assigned a male leader; however, assigning a female leader in a male-majority condition substantially reduces the level

[44]Appendix Table A15 summarizes the results of this exercise, noting specifically which parameters were extracted from the data and which were results of the model calibration.

of discrimination in the assessment of women's influence, from approximately 0.7 standard deviations to 0.3 standard deviations. Comparing this to patterns in the female-majority condition, the imposition of a female leader has similar effects in reducing the level of discrimination (from 0.4 standard deviations to 0.1 standard deviations), while the imposition of a male leader appears to increase the rate of the gender penalty from 0.4 standard deviations to 0.6 standard deviations. A similar pattern of effects holds in the global influence assessment in Panel 6c. Starting from a lower initial level of discrimination in both conditions, female leaders reduce the expression of discrimination to less than zero (i.e., a slight positive discrimination in favor of women), while male leaders either slightly or substantially increase gender discrimination in male- and female-majority conditions, respectively.

Finally, our results on influence assessments over time in Study 1 — which was not subject to the leadership treatment — suggest that groups in both conditions reduce their expressed level of gender discrimination over time. Figure 6d shows how male-majority groups that begin at nearly 0.8 standard deviations penalty and end the study with a penalty close to zero. We find no evidence that average levels of sexism among men — solicited via a standard survey-based questionnaire module (Glick and Fiske, 1996) — change due to either the gender composition of the group or the gender of the leader.[45] In other words, the experimental conditions affected the *expression* of discriminatory attitudes, not the attitudes themselves. While we caution that the underlying experimental data is not consistent across both studies, the simulations suggest, then, that in addition to changing the formal structure of group leadership, one potential pathway to reducing discrimination in some male-dominated settings may be through exposure to female peers over time.

[45]See Appendix Table A16. Men and women in Study 2 report slightly lower levels of sexism at the end of the experiment than at the beginning, though both pre-treatment and at the end of the semester, men have significantly higher levels of sexism than women. We do not have a similar post-treatment measure of sexism in Study 1, unfortunately. Women's decline in sexism is greater in groups with a male leader than with a female leader, though the magnitude of the difference-in-differences is small.

5 Alternative Mechanisms

Our rich data allows us to explore a number of additional mechanisms behind the lower influence of women in majority male groups, including women's own behavior such as talk time and task performance. It is important to note that even though we are able to compare across conditions in these analyses, they rely on post-treatment outcomes. Hence, we do not interpret these models as causal. Nevertheless, we find these correlations important and believe they provide valuable insights into observed patterns and mark useful directions for direct testing in future research.

First, we explore whether women can overcome the influence deficit they experience in male majority groups by simply "leaning in" and becoming more active participants in group deliberations and speaking up more.[46] Leveraging our audio recordings of group discussions during team building tasks, we compare women's average number of speaking turns across conditions.[47] As reported in Appendix Table A17, we do not find that women in majority male groups are significantly less likely to speak up during group deliberations. In other words, in both conditions, the women in our sample appear to be "leaning in" by participating actively in discussions. Moreover, while we find a strong positive correlation between speaking turns and being seen as influential in both conditions, in male majority groups that relationship is driven almost entirely by men (see Appendix Table A18). The large and negative interaction between participant's gender and speaking turns means that the ability to turn conversational participation into influence evaporates for women, especially in majority-male groups.

Next, we ask whether better performance on the task allows women to become more influential in male-majority groups. Our findings, reported in Appendix Table A19, suggest

[46]Prior studies have shown a strong effect of group gender composition on women's participation in deliberation (Karpowitz and Mendelberg, 2014).

[47]While we have transcribed all of the audio data from Study 1, in Study 2, we were only able to use audio data from two out of the four semesters due to Covid-19 interruptions.

that the answer is no. While the men in male majority groups are able to convert their better task performance to more influence in the group, women receive none of these positive returns for excellent task performance (or penalties for poor performance). The positive interaction term between female and task performance indicates that women who are better at the task are not able to be more influential in the group, especially in Study 1. If anything, better task performance for women in these groups might even be penalized. By contrast, there is no correlation between task performance and influence votes in female-majority groups for either men or women. When considering the most efficient use of human capital, this may not be ideal. However, the efficient conversion of task performance into influence for men that we see in the male-majority groups must be weighed against the opposite pattern for women on these teams. Hence, these findings have policy implications for organizations that care about the egalitarian treatment of employees. When these results are combined with the findings about talk time, they cast significant doubt on the idea that women in the minority can control the factors that lead to being seen as influential. Instead, those factors appear to be largely structural in our data.

6 External Validity and Generalizability

The findings from our study may be broadly applicable to a variety of settings as many social, academic, and professional contexts involve mixed-gender collaborative teams that may or may not have formal leaders. Nonetheless, field experiments face tradeoffs between internal and external validity, and the potential generalizability of our findings can be affected by 1) selection, 2) attrition, 3) naturalness, and 4) scale (List, 2020). In terms of selection, our sample is younger, whiter, and more highly educated than the population of the United States. Nonetheless, mean levels of sexism in our sample, as measured by a widely used indicator of ambivalent sexism (Glick and Fiske, 1996), are highly comparable to the degree

of sexism in a representative sample of Americans who participated in the 2018 Cooperative Congressional Election Study (CCES), a high-quality national survey (Appendix Figure A1).[48] The difference in means between our sample and the CCES sample (weighted to reflect the general population of the United States) is substantively small (0.02 on a scale of zero to one) and not statistically significant ($t = 0.71$, $p = 0.48$). Most importantly, when we use the CCES to re-weight our sample to match the joint distribution of gender and sexism in the US population, our findings are unchanged (Appendix Table A7). To the extent that sexism contributes to the gender inequalities we document, we view our results as reflective of the patterns we would find across the United States as a whole.

Our studies enjoyed minimal attrition and high compliance: survey completion rates were above 98 percent for every monthly survey in Study 1 and above 93 percent for Study 2 (Appendix Table A4). In neither study were survey non-completion, study withdrawal rates, or non-consent related to treatment condition.[49] Both studies were natural field experiments in which subjects were engaged in common tasks, and neither control nor treatment groups were placed on artificial margins when making decisions (Harrison and List, 2004). Given these design features, we believe our study setting can potentially generalize our findings to common situations and conditions, although the true scope of generalizability will only be known with replications in these varied settings.

The most feasible way to scale our gender composition treatment would be for organizations to ensure that when women are assigned to teams in organizational settings, they are never in the minority. In numerically gender-imbalanced settings, one effect of such intervention might be an increased share of gender-segregated teams, and costs of such an

[48] See Appendix Section 1.1 for details.

[49] See footnote 10 above for details of the (lack of) relationship between consent and treatment condition. In addition, withdrawal rates were low and did not vary across conditions either in the pooled sample or by gender. Appendix Table A4 shows that across the five monthly surveys, the difference in survey completion rates by experimental condition was always substantively small and never statistically significant at the 95 percent confidence level.

approach would need to be carefully weighed against the benefits to women's experiences. The leadership intervention lends itself to a modest degree of scale in settings like workplace teams, corporate boards, legislative committees, etc. Like the composition treatment, extreme scaling of this policy — allowing leadership positions to be held only by women — might incur backlash among male subordinates whose opportunities become limited by design. Future work should investigate the relationship and mechanisms behind the intensity of organizational leadership quotas and counterproductive behavioral reactions.

7 Discussion

Across two distinct multi-year field experiments and multiple indicators, our results show a consistent pattern of devaluing women's participation and expertise in work teams, especially when they are in the minority. It is not difficult to imagine the negative effects this devaluation might have on women's opportunities to advance in their workplaces. For example, if women have difficulty shaping group decisions on a project and are perceived by their peers as less influential, it will be harder for them to advocate for a promotion for their contributions to that project. If women are less likely to be chosen as a team spokesperson, they will not get the kind of face time with bosses and clients that leads to new opportunities and a reputation as a rising star. While we find some limited evidence that peers gradually seem to acknowledge women's influence over a period of time, things like salary negotiations or promotion decisions often hinge on an employee being able to point to specific contributions to team success. Women in the minority face unique challenges in exercising team leadership that may compound across settings and throughout their career.

Notably, the influence deficit that we document for women in the minority does not extend to men in majority female groups. By comparing men's and women's outcomes in conditions where they comprise an equivalent numerical minority, we find no evidence of

disempowerment for men outnumbered by women. While women experience dramatically lower levels of both perceived and actual influence and authority, men face no such influence deficit. This is consistent with prior studies which find no disadvantage for men when they are in the minority across a wide range of professional and educational settings (Radbruch and Schiprowski, 2023; Shan, 2023; Delfino, 2021).

What is the source of this influence deficit for women in minority? We decompose the voting patterns in our data and discover that women vote for themselves and other women at rates considerably lower than expected given their presence in the group. While women's self-votes are unaffected by the gender composition of the group, their propensity to vote for the other women in the group is significantly higher in female-majority groups relative to male-majority groups. In contrast, men vote for themselves at rates equal to what is expected given their proportion in the group, and they perceive the other men in the group to be disproportionately more influential regardless of the condition. Thus we find that men's higher perceived influence in groups can be explained in part by men's tendency to favor themselves and other men and in part by the fact that women also allocate a disproportionate share of their votes to men instead of to themselves or to other women.

Importantly, we do not find the same kind of group gender composition treatment effects when we ask participants to identify the most and least supportive members of their group. In fact, rather than being seen as influential, women in male-majority groups are seen as highly supportive. Given the close cultural ties between authority and masculinity and supportiveness and femininity (Jamieson et al., 1995; Manne, 2017), one wonders the extent to which sexism plays a role in these findings. Our pre-treatment measures of sexism suggest that men in our sample (as in the US as a whole) are considerably more sexist than women (Appendix Table A1), so a majority-male environment is simply more deeply saturated in sexist beliefs than a majority-female environment. However, it is important to remember that on average the men in the majority-female condition express similar baseline levels of

sexism as the men in majority-male groups.[50] Despite this, their behavior is significantly more egalitarian in female-majority groups. In other words, random assignment ensures that the differences across the conditions are not merely a result of differences in the attitudes men and women bring to their teams.

To identify the underlying cause of our results, we simulate a simple model of voting in the presence of discrimination to which to calibrate our experimental results. This framework allows us to estimate the discrimination parameter in each of our experimental conditions. The implied level of discrimination applied to women is approximately 0.67 standard deviations in the male-majority condition, and reduces to 0.25 standard deviations in the female-majority condition. This fundamentally changes women's experience in these groups. As discussed above, relative to the women in male-majority groups, women in female-majority groups are significantly more likely to be evaluated by their peers as most influential and chosen as a spokesperson to represent their group. Thus, changing the group-as-institution strongly affects the degree of sexism experienced by women via the expression of discrimination in those groups. This result supports the argument that altering institutional structures is likely to be more effective in reducing sexist behavior than attempting to alter beliefs (Bohnet, 2016; Beaman et al., 2009).

A unique feature of Study 2 is that it allowed us to investigate the effect of an alternative institutional intervention — formal leadership — on women's influence and authority in groups. Halfway through the semester, we assigned group leaders randomly varying their gender across groups. We find striking effects of formal leadership on the influence of women. Specifically, the presence of a female leader almost entirely eliminated the difference in votes per woman in male-majority and female-majority conditions. Importantly, it did so by

[50] If anything, in Study 2, men in the majority female condition have slightly *higher* levels of baseline sexism than those in the majority male condition (see Appendix Table A16), but this difference is not statistically significant ($p = 0.18$). Men's baseline sexism is not statistically distinguishable across gender composition conditions in Study 1 either.

increasing women's influence in majority-male groups, providing strong evidence that formal authority empowered women. In contrast, we observe small but statistically insignificant decreases in women's influence in groups with male leaders, suggesting that male leadership is likely the default norm in the absence of formal authority.

Our analysis of alternative mechanisms, including women's speaking time and task performance, points to significant limits to women's ability to control their perceived and actual influence in teams. Instead, the structural factors tied to group-level features dominate evaluations of women's influence. In short, our findings suggest there are strong reasons to further investigate how the gender composition of a workplace and its formal leadership structure contribute to persistent gender gaps in participation, pay, and advancement, especially in historically male-dominated sectors.

References

Aman-Rana, Shan, Clement Minaudier, Brais Alvarez Pereira, and Shamyla Chaudry (2022), "Gender and choice over co-workers: Experimental evidence." *SSRN Electronic Journal*, URL https://dx.doi.org/10.2139/ssrn.3958386.

Andreoni, James (2006), "Leadership giving in charitable fund-raising." *Journal of Public Economic Theory*, 8, 1–22, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9779.2006.00250.x.

Anelli, Massimo and Giovanni Peri (2019), "The effects of high school peers' gender on college major, college performance and income." *The Economic Journal*, 129, 553–602.

Apesteguia, Jose, Ghazala Azmat, and Nagore Iriberri (2012), "The impact of gender composition on team performance and decision making: Evidence from the field." *Management Science*, 58, 78–93.

Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart (2017), "Gender differences in accepting and receiving requests for tasks with low promotability." *American Economic Review*, 107, 714–747.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva (2017), "Does the gender composition of scientific committees matter?" *American Economic Review*, 107, 1207–38, URL https://www.aeaweb.org/articles?id=10.1257/aer.20151211.

Beaman, Lori, Raghabendra Chattopadhyay, Ester Duflo, Rohini Pande, and Anna Topalova (2009), "Powerful women: Does exposure reduce bias?" *The Quarterly Journal of Economics*.

Berg, Peter, Eileen Appelbaum, Thomas Bailey, and Arne L Kalleberg (1996), "The per-

formance effects of modular production in the apparel industry." *Industrial Relations: A Journal of Economy and Society*, 35, 356–373.

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

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg (2019), "The dynamics of discrimination: Theory and evidence." *American Economic Review*, 109, 3395–3436.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw (2007), "Opportunity counts: Teams and the effectiveness of production incentives." *Journal of Labor Economics*, 25, 613–650.

Booth, Alison L., Lina Cardona-Sosa, and Patrick Nolen (2018), "Do single-sex classes affect academic achievement? an experiment in a coeducational university." *Journal of Public Economics*, 168, 109–126.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016), "Stereotypes." *The Quarterly Journal of Economics*, 131, 1753–1794.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2019), "Beliefs about gender." *American Economic Review*, 109, 739–73.

Born, Andreas, Eva Ranehill, and Anna Sandberg (2020), "Gender and willingness to lead: Does the gender composition of teams matter?" *Review of Economics and Statistics*.

Bostwick, Valerie and Bruce Weinberg (2020), "Nevertheless she persisted? gender peer effects in doctoral stem programs." *NBER Working Paper No. 25028*, 107.

Brenoe, Anne Ardila and Ulf Zölitz (2019), "Exposure to more female peers widens the gender gap in stem participation."

Calder-Wang, Sophie and Paul A. Gompers (2021), "And the children shall lead: Gender diversity and performance in venture capital." *Journal of Financial Economics*, 142, 1–22, URL https://www.sciencedirect.com/science/article/pii/S0304405X21001483.

Coffman, Katherine B., Clio Bryant Flikkema, and Olga Shurchkov (2021), "Gender stereotypes in deliberation and team decisions." *Games and Economic Behavior*, 28, 61–79.

Coffman, Katherine Baldiga (2014), "Evidence on self-stereotyping and the contribution of ideas." *The Quarterly Journal of Economics*, 129, 1625–1660.

Crowder-Meyer, Melody, Shana Kushner Gadarian, Jessica Trounstine, and Kau Vue (2020), "A different kind of disadvantage: Candidate race, cognitive complexity, and voter choice." *Political Behavior*, 42, 509–530.

Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth (2020), "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams*." *The Quarterly Journal of Economics*, 136, 987–1030.

De Paola, Maria, Francesca Gioia, and Vincenzo Scoppa (2022), "Female leadership: Effectiveness and perception." *Journal of Economic Behavior & Organization*, 201, 134–162, URL https://ideas.repec.org/a/eee/jeborg/v201y2022icp134-162.html.

Delfino, Alexia (2021), "Breaking gender barriers: Experimental evidence on men in pink-collar jobs." *IZA Discussion Paper 14083*.

Deschamps, Pierre (2021), "Gender quotas in hiring committees: a boon or a bane for women?"

DiMaggio, Paul and Filiz Garip (2012), "Network effects and social inequality." *Annual Review of Sociology*, 38, 93–118.

Eckel, Catherine C. and Philip J. Grossman (2005), "Managing diversity by creating team identity." *Journal of Economic Behavior & Organization*, 58, 371–392.

Exley, Christine L and Judd B Kessler (2022), "The Gender Gap in Self-Promotion*." *The Quarterly Journal of Economics*, 137, 1345–1381, URL https://doi.org/10.1093/qje/qjac003.

Folke, Olle and Johanna Karin Rickne (2020), "Sexual harassment and gender inequality in the labor market."

Glick, Peter and Susan T. Fiske (1996), "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism." *Journal of Personality and Social Psychology*, 70, 491–512.

Hall, J. and W.H. Watson (1970), "The effects of a normative intervention on group decision-making performance." *Human Relations*, 23, 299–317, URL https://www.psychologicalscience.org/observer/nasa-exercise.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan (2003), "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation." *Journal of Political Economy*, 111, 465–497.

Hampone, Menaka, Francesca Truffa, and Ashley Wong (2022), "Peer effects and the gender gap in corporate leadership: Evidence from MBA students."

Harrison, Glenn W and John A List (2004), "Field experiments." *Journal of Economic Literature*, 42, 1009–1055.

Hurst, William Reuben, Saerom Lee, and Justin Frake (2022), "The hidden cost of flat hierarchies for applicant pool diversity: Evidence from experiments." *Academy of Management Proceedings*, 2022.

Isaksson, Siri (2018), "It takes two: Gender differences in group work.", URL https://www.dropbox.com/s/hafhh70p3li9tij/1120jmp.pdf?dl=0. [Online; accessed 2019-03-31].

Jamieson, Kathleen Hall, Kathleen Hall, et al. (1995), *Beyond the double bind: Women and leadership*. Oxford University Press on Demand.

Kanter, Rosabeth Moss (1977), "Some effects of proportions on group life: Skewed sex ratios and responses to token women." *American Journal of Sociology*, 82, 965–990.

Karpowitz, Christopher F and Tali Mendelberg (2014), *The silent sex: Gender, deliberation, and institutions*. Princeton University Press.

Koffi, Marlene (2020), "Innovative ideas and gender inequality." *Working Paper*.

Kofoed, Michael S et al. (2019), "The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at west point." *Journal of Human Resources*, 54, 430–467.

Kossinets, Gueorgi and Duncan J Watts (2009), "Origins of homophily in an evolving social network." *American Journal of Sociology*, 115, 405–450.

Lawler, EE and Susan A Mohrman (2003), "Pay practices in fortune 1000 corporations." *WorldatWork Journal*, 12, 45–54.

Lazear, Edward P. and Kathryn L. Shaw (2007), "Personnel economics: The economist's view of human resources." *Journal of Economic Perspectives*, 21, 91–114.

Lindbeck, Assar and Dennis J. Snower (2000), "Multitask learning and the reorganization of work: From tayloristic to holistic organization." *Journal of Labor Economics*, 18, 353–376, URL https://doi.org/10.1086/209962.

List, John and Daniel Rondeau (2003), "The impact of challenge gifts on charitable giving: an experimental investigation." *Economics Letters*, 79, 153–159, URL https://EconPapers.repec.org/RePEc:eee:ecolet:v:79:y:2003:i:2:p:153-159.

List, John A (2020), "Non est disputandum de generalizability? a glimpse into the external validity trial." *National Bureau of Economic Research*, NBER Working Paper 27535, URL http://dx.doi.org/10.3386/w27535.

List, John A, Azeem M Shaikh, and Yang Xu (2019), "Multiple hypothesis testing in experimental economics." *Experimental Economics*, 22.

Manne, Kate (2017), *Down girl: The logic of misogyny*. Oxford University Press.

McDonald, Theodore W, Loren L Toussaint, and Jennifer A Schweiger (2004), "The influence of social status on token women leaders' expectations about leading male-dominated groups." *Sex roles*, 50, 401–409.

Mendelberg, Tali, Christopher Karpowitz, and Nicholas Goedert (2014), "Does Descriptive Representation Facilitate Women's Distinctive Voice? How Gender Composition and Decision Rules Affect Deliberation." *The American Journal of Political Science*, 58, 291–306.

Nix, Emily (2020), "Learning spillovers in the firm." *Working paper.*

Norton, Michael I, Jeana H Frost, and Dan Ariely (2013), "Less is often more, but not always: Additional evidence that familiarity breeds contempt and a call for future research."

Ott, E Marlies (1989), "Effects of the male-female ratio at work: Policewomen and male nurses." *Psychology of Women Quarterly*, 13, 41–57.

PACE (2016), "Teambuilder: Lost At Sea." URL https://www.airman.af.mil/Portals/17/002%20All%20Products/003%20PACEsetters/Teambuilder_LOST_AT_SEA.pdf?ver=2016-11-02-111101-433. [Online; accessed 13-April-2023].

Padilla (2020), "Women on boards: March 2020 report." URL https://bpd.cdn.sos.ca.gov/women-on-boards/WOB-Report-04.pdf. [Online; accessed 13-April-2023].

Preece, Jessica Robinson (2016), "Mind the gender gap: An experiment on the influence of self-efficacy on political interest." *Politics Gender*, 12, 198–217.

Radbruch, Jonas and Amelie Schiprowski (2023), "Gender differences in influence: Evidence from hiring committees." *Mimeo*.

Sahin, Selhan, Catherine Eckel, and Mana Komai (2015), "An experimental study of leadership institutions in collective action games." *Journal of the Economic Science Association*, 1.

Sarsons, Heather (2017), "Recognition for group work: Gender differences in academia." *American Economic Review*, 107, 141–145.

Shan, Xiaoyue (2023), "The minority trap: Minority status drives women out of male-dominated fields." *Working paper*.

Spangler, Eve, Marsha A Gordon, and Ronald M Pipkin (1978), "Token women: An empirical test of kanter's hypothesis." *American Journal of Sociology*, 84, 160–170.

The Economist (2018), "Ten years on from norway's quota for women on corporate boards." URL https://www.economist.com/business/2018/02/17/ten-years-on-from-norways-quota-for-women-on-corporate-boards. [Online; accessed 13-April-2023].

Truffa, Francesca and Ashley Wong (2021), "Undergraduate gender diversity and the direction of scientific research." *Working paper*.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi (2007), "The increasing dominance of teams in production of knowledge." *Science*, 316, 1036–1039.

Yang, Yang, Tanya Y. Tian, Teresa K. Woodruff, Benjamin F. Jones, and Brian Uzzi (2022), "Gender-diverse teams produce more novel and higher-impact scientific ideas." *Proceedings*

of the National Academy of Sciences, 119, e2200841119, URL https://www.pnas.org/doi/abs/10.1073/pnas.2200841119.

Zölitz, Ulf and Jan Feld (2018), "The effect of peer gender on major choice." *University of Zurich, Department of Economics, Working Paper*.

Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams

Supporting Information Appendix

1 Sample Characteristics and Key Variables

Appendix Table A1: Randomization Balance by Gender and Condition

Study 1	Female		Male	
	Maj. Male	Maj. Female	Maj. Male	Maj. Female
Age	22.53	22.69	23.88	23.81
White	0.74	0.87	0.95	0.91
Married	0.32	0.23	0.32	0.37
GPA	3.80	3.78	3.78	3.79
Leadership experience	0.84	0.77	0.74	0.76
International student	0.21	0.10	0.06	0.10
Parental income category	3.76	3.85	4.33	4.13
Political affiliation	5.16	5.47	5.84	5.84
Ambivalent sexism index	0.45	0.41	0.58	0.54
Observations	38	108	154	71
Study 2	Female		Male	
	Maj. Male	Maj. Female	Maj. Male	Maj. Female
Age	19.21	19.44	20.28	20.31
White	0.95	0.93	0.93	0.95
Married	0.05	0.03	0.05	0.05
GPA	3.80	3.77	3.75	3.75
Leadership experience	0.84	0.81	0.89	0.89
International student	0.02	0.03	0.04	0.04
Parental income category	3.21	3.30	3.20	3.20
Political affiliation	4.89	5.08	5.39	5.72*
Ambivalent sexism index	0.39	0.41	0.53	0.56
Observations	143	278	292	137
Studies 1 + 2	Female		Male	
	Maj. Male	Maj. Female	Maj. Male	Maj. Female
Age	19.91	20.35	21.52	21.49
White	0.91	0.92	0.94	0.94
Married	0.10	0.09	0.15	0.16
GPA	3.80	3.77	3.76	3.76
Leadership experience	0.84	0.80	0.84	0.84
International student	0.06	0.05	0.04	0.06
Parental income category	3.33	3.45	3.59	3.52
Political affiliation	4.95	5.19	5.55	5.76
Ambivalent sexism index	0.40	0.41	0.55	0.55
Observations	181	386	446	208

Note: Significance levels are indicated by $* < .05$, $** < .01$.

Appendix Table A2: Description of Dependent Variables

General	Variable Description	Data Source
Treatment Indicator for Male Majority	Indicator variable: 1 if the group has majority men, 0 otherwise	Study condition
Treatment Indicator for Female Leader	Indicator variable: 1 if the group has female leader, 0 otherwise (Study 2 only)	Study condition
Group-Level Dependent Variables		
Women Most Influential Votes (Team-Building Task)	Average number of votes for a woman as the most influential in the team-building task divided by the proportion of women in the group	Team-building task
Women Least Influential Votes (Team-Building Task)	Average number of votes for a woman as the least influential in the team-building task divided by the proportion of women in the group	Team-building task
Women Spokesperson Votes	Average number of votes for a woman to be the group spokesperson in the team-building task divided by the proportion of women in the group	Team-building task
Women's Empirical Influence	Indicator variable: 1 if a woman had the lowest difference from the final group-level ranking in either team-building task session	Team-building task
Women Most Influential Votes (Global Assessment)	Average number of votes for a woman as the most influential in a global assessment divided by the proportion of women in the group	Self-reported survey
Women Least Influential Votes (Global Assessment)	Average number of votes for a woman as the least influential in a global assessment divided by the proportion of women in the group	Self-reported survey
Task Errors	Total difference between individual and expert rankings in team-building task	Team-building task
Individual-Level Dependent Variables		
Average Speaking Turns	The number of times an individual spoke in the team-building task divided by the group conversation length (see below)	Transcript audio data
Group Conversation Length	The number of times anyone in the group spoke in the team-building task	Transcript audio data
Fall Semester GPA	GPA on a four-point scale in junior core classes during fall semester (Study 1 only)	Administrative data
School Year GPA	GPA on a four-point scale in junior core classes during both fall and winter semesters (Study 1 only)	Administrative data
Course GPA	Final grade on a four-point scale in general education course (Study 2 only)	Administrative data
Individual Score in Group Project	The average group project individual contribution score out of 15 points across two different graders (Study 2 only)	Administrative data
Most Influential Self Vote	Indicator variable: 1 if individual voted for self as the most influential member of their group in the team-building task, 0 otherwise	Team-building task
Group-Level Controls		
Age	Average age of group members	Administrative data
White	Proportion of group that is white	Administrative data
Online	Indicator variable: 1 if semester was held online (over Zoom), 0 if held in person	Administrative data
Average GPA	Average prior GPA of group members on a four-point scale	Administrative data
International	Number of international students in group	Administrative data
Married	Number of married students in group	Self-reported survey

Appendix Table A3: Summary Statistics

General	Mean	Standard Deviation	N
Treatment Indicator for Male Majority (Studies 1 + 2)	0.514	0.501	220
Treatment Indicator for Female leader (Study 2 only)	0.51	0.502	145
Group Level Variables			
Women Most Influential Votes (Team-Building Task)			
Study 1	0.671	0.621	75
Study 2 (Task 1)	0.58	0.622	123
Study 2 (Task 2)	0.656	0.696	122
Studies 1 + 2	0.615	0.622	198
Women Least Influential Votes (Team-Building Task)			
Study 1	1.346	1.185	75
Study 2 (Task 1)	1.207	0.81	123
Studies 1 + 2	1.259	0.969	198
Women Spokesperson Votes			
Study 2 (Task 2)	0.722	0.673	122
Studies 1 + 2	0.631	0.664	198
Women's Empirical Influence			
Study 1	0.48	0.503	75
Study 2 (Task 1)	0.423	0.496	123
Studies 1 + 2	0.444	0.498	198
Women Most Influential Votes (Global Assessment)			
Study 1 (Months 1 + 2)	0.689	0.712	75
Study 2 (Months 1 + 2)	0.779	0.608	145
Study 2 (Months 3 + 4)	0.812	0.610	145
Studies 1 + 2 (Months 1 + 2)	0.748	0.645	220
Women Least Influential Votes (Global Assessment)			
Study 1 (Months 1 + 2)	1.098	1.084	75
Study 2 (Months 1 + 2)	0.964	0.714	145
Studies 1 + 2 (Months 1 + 2)	1.01	0.858	220
Individual Level Variables			
Task Errors			
Study 1	52.174	8.581	371
Study 2 (Task 1)	45.19	12.089	678
Average Speaking Turns			
Study 1	0.201	0.051	371
Study 2 (Task 1)	0.161	0.143	348
Group Conversation Length			
Study 1	315.458	45.39	371
Study 2 (Task 1)	503.641	223.663	348
Fall Semester GPA (Study 1 only)	3.562	0.322	368
School Year GPA (Study 1 only)	3.566	0.308	368
Course GPA (Study 2 only)	2.993	0.901	824
Individual Score in Group Project (Study 2 only)	11.265	1.824	812
Most Influential Self Vote			
Study 1	0.259	0.439	371
Study 2 (Task 1)	0.169	0.375	850
Group Level Controls			
Age (Studies 1 + 2)	21.158	1.906	220
White (Studies 1 + 2)	0.923	0.107	220
Online (Studies 1 + 2)	0.359	0.481	220
GPA Average (Studies 1 + 2)	3.771	0.103	220
International (Studies 1 + 2)	0.286	0.482	220
Married (Studies 1 + 2)	0.695	0.913	220

Appendix Table A4: Global Assessment Survey Response Rates and Effect of Conditions on Non-Response

	Month 1	Month 2	Month 3	Month 4	Month 5
Study 1					
Majority Male Groups	0.006	-0.010	0.000	0.017*	0.001
	(0.006)	(0.013)	(0.008)	(0.009)	(0.013)
Constant	0.994***	0.989***	0.994***	0.983***	0.983***
	(0.006)	(0.008)	(0.006)	(0.009)	(0.009)
Response Rate	99.73	98.38	99.46	99.19	98.38
Num.Obs.	371	371	371	371	371
R2 Adj.	0.000	-0.001	-0.003	0.006	-0.003
Study 2					
Majority Male Groups	-0.006	0.018	0.019	0.029*	-0.009
	(0.007)	(0.011)	(0.015)	(0.016)	(0.017)
Constant	0.990***	0.966***	0.940***	0.930***	0.935***
	(0.005)	(0.009)	(0.012)	(0.012)	(0.012)
Response Rate	98.71	97.53	94.94	94.47	93.06
Num.Obs.	850	850	850	850	850
R2 Adj.	0.000	0.002	0.001	0.003	-0.001

Note: Table presents individual-level OLS analysis of the effects of group composition on survey non-response. The dependent variable is a dichotomous indicator of whether or not the study participant completed at least some portion of the monthly survey. Standard errors are clustered by group and reported in parentheses. Significance levels are indicated by $* < .1$, ** $< .05$, *** $< .01$.

Appendix Table A5: Study 2 Power Calculations

Outcome	Study 1 effect size (in SD)	Sample size per condition to power replication of same effect size	Sample size, Female-majority condition	Sample size, Male-majority condition
Team-building influence, most influential	0.652	29	60	63
Team-building influence, least influential	0.572	38	60	63
Team-building task spokesperson votes	0.445	62	60	63
Global influence votes, most influential	0.459	58	71	74

Note: Table presents sample sizes needed to power a two-sample, one-sided test of means at 0.05 significance level and .80 power.

1.1 Sexism

To evaluate pre-treatment levels of sexism, the baseline surveys of both studies included questions drawn from a commonly used and well-validated indicator, ambivalent sexism (Glick and Fiske, 1996). This measure includes two subscales, hostile sexism and benevolent sexism, and is strongly predictive of stereotypes about women. Chosen because they loaded highly on the ambivalent sexism factor in exploratory factor analysis across five studies analyzed by Glick and Fiske (1996), the following items were included on the baseline questionnaires (randomly ordered):

- Women are too easily offended.*
- Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."*
- Women exaggerate problems they have at work.
- Men should be willing to sacrifice their own well-being in order to provide financially for the women in their lives.
- Many women have a quality of purity that few men possess.*
- No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman.

Response options ranged from strongly agree to strongly disagree. The first three items function as indicators of hostile sexism, and the second three items tap benevolent sexism. The six items scale together well (Cronbach's $\alpha = 0.66$ in Study 1 and 0.77 in Study 2). As shown in Table A1, as expected, men in both studies scored higher on this measure of sexism than did women.

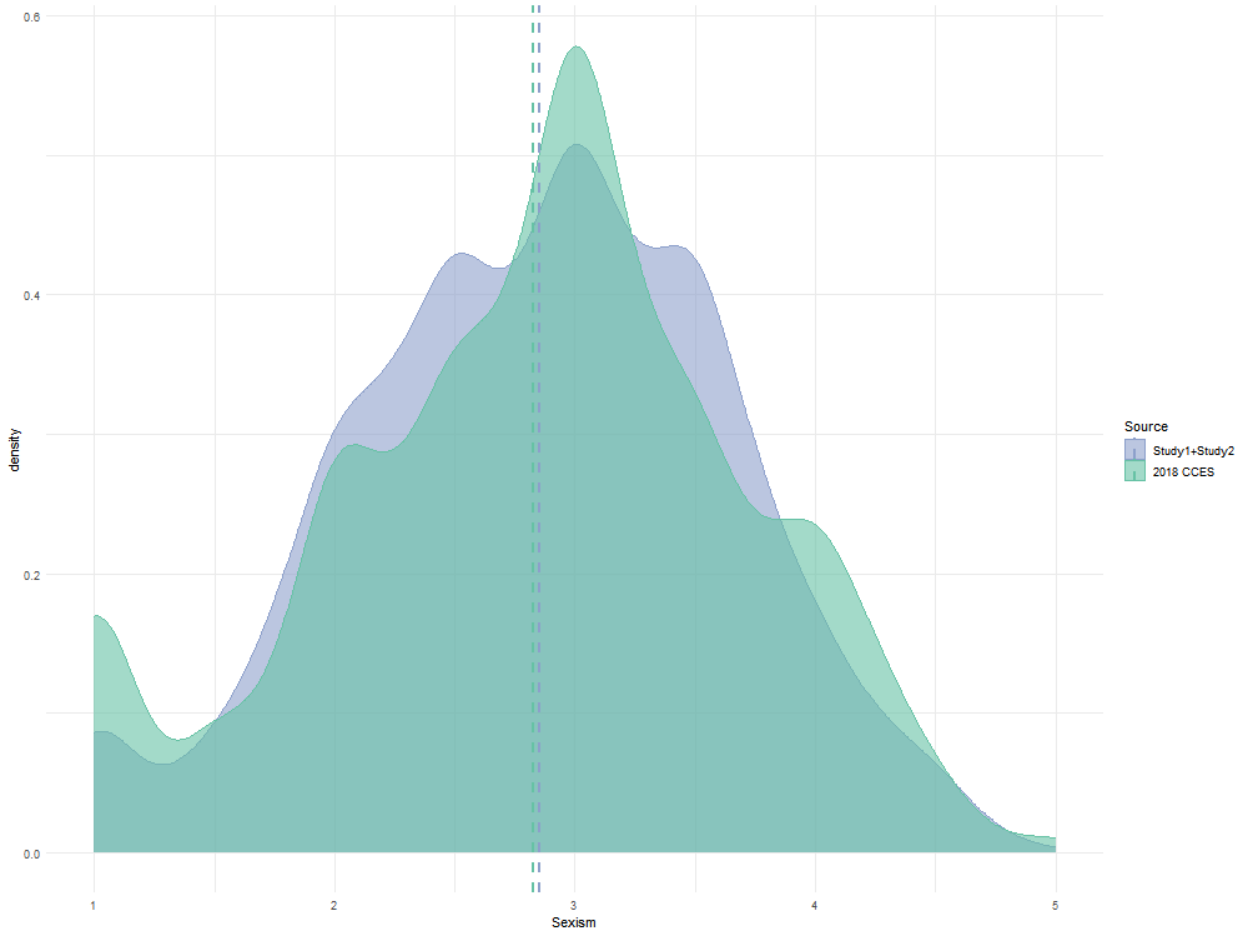
To compare levels of ambivalent sexism in our sample with a nationally representative sample, we use the UMass-Amherst module of the 2018 Cooperative Congressional Election Study (CCES), a high-quality national survey of 1,000 respondents fielded between September 27 and November 5, 2018.[1] Three of the ambivalent sexism items listed above were included on both Study 1 and Study 2 as well as the 2018 CCES and are indicated with an asterisk. Though abbreviated, these items enjoy psychometric properties similar to the larger index (Cronbach's $\alpha = 0.54$ in Study 1 and Study 2 combined and 0.53 in the CCES). We thus constructed a measure of ambivalent sexism by first averaging the two hostile sexism measures, then averaging the hostile sexism indicator and the benevolent sexism question to produce an ambivlanet sexism index that runs from 1-5.[2]

[1] Details of the 2018 study can be found at https://doi.org/10.7910/DVN/ZSBZ7K/WZWCZ1. The CCES is an internet survey that uses proximity matching methods and sample weighting techniques to produce a sample that mirrors the characteristics of the general population of the United States.

[2] We construct the measure in this way for purposes of comparing samples because we do not have equal numbers of hostile and benevolent sexism questions. Of the three questions our studies and the CCES have in common, two are indicators of hostile sexism and one measures benevolent sexism.

Figure A1 shows the distribution of the ambivalent sexism index in Study 1 and Study 2 combined and in the survey weighted 2018 CCES. The sample averages are indicated with the dashed lines. They are nearly identical (2.85 in Study 1 and Study 2 and 2.83 in the 2018 CCES). The difference between the samples is substantively small and not statistically distinguishable from zero in a difference-of-means test ($t = 0.71$, $p = 0.48$). We take this as strong evidence that our sample is similar in its sexism to the United States as a whole.

Appendix Figure A1: Distribution of Ambivalent Sexism by Source



2 Supporting Tables and Figures

Appendix Table A6: Effects of Group Gender Composition on Women's Influence (Analysis with Controls)

	Study 1 (N=75)	Study 2 (N=123)	Studies 1 + 2 (N=198)
Team-Building Most Influential Votes	-0.498***	-0.316***	-0.372***
	(0.132)	(0.111)	(0.083)
Mean of outcome	0.67	0.58	0.61
St. dev. of outcome	0.62	0.62	0.62
Team-Building Incentivized Spokesperson Votes	-0.455***	-0.236**	-0.297***
	(0.161)	(0.104)	(0.089)
Mean of outcome	0.77	0.55	0.63
St. dev. of outcome	0.79	0.56	0.66
Global Assessment Most Influential Votes	-0.367**	-0.149	-0.243**
	(0.164)	(0.117)	(0.094)
Mean of outcome	0.69	0.75	0.73
St. dev. of outcome	0.71	0.61	0.65
Team-Building Least Influential Votes	0.670**	0.014	0.259**
	(0.266)	(0.147)	(0.131)
Mean of outcome	1.35	1.21	1.26
St. dev. of outcome	1.18	0.81	0.97

Note: Table presents group-level analysis of the effects of majority male groups on perceptions of influence with controls. In each case, the dependent variable is votes for women divided by the proportion of women in the group. Controls include group-level average GPA, age, race (proportion Nonwhite), marital status (number of married team members), and the presence of an international student. Study 2 analysis excludes 22 groups from the semester when the team-building task was not conducted. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Appendix Table A7: Effects of Group Gender Composition on Women's Influence (Analysis with weights from CCES 2018)

	Study 1	Study 2	Studies 1 + 2
Panel A: Team-Building Most Influential Votes			
Majority male group	-0.495***	-0.305***	-0.351***
	(0.157)	(0.109)	(0.085)
Mean of outcome	0.79	0.57	0.62
St. dev. of outcome	0.62	0.62	0.61
Panel B: Team-Building Incentivized Spokesperson Votes			
Majority male group	-0.442**	-0.254***	-0.296***
	(0.174)	(0.096)	(0.089)
Mean of outcome	0.84	0.54	0.62
St. dev. of outcome	0.73	0.55	0.64
Panel C: Global Assessment Most Influential Votes			
Majority male group	-0.384**	-0.202*	-0.242***
	(0.168)	(0.108)	(0.089)
Mean of outcome	0.79	0.74	0.74
St. dev. of outcome	0.68	0.60	0.63
Panel D: Team-Building Least Influential Votes			
Majority male group	0.663**	0.093	0.293**
	(0.290)	(0.150)	(0.135)
Mean of outcome	1.28	1.23	1.26
St. dev. of outcome	1.10	0.82	0.93

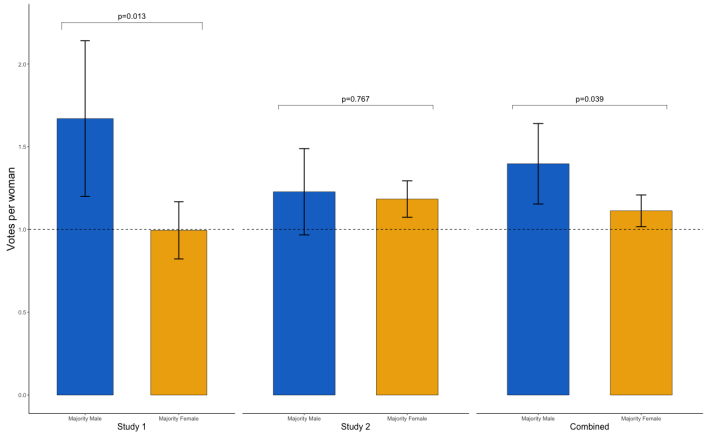
Note: Table presents group-level analysis of the effects of majority male groups on perceptions of influence, weighted by gender and sexism to match a representative sample of the population of the United States. In each case, the dependent variable is votes for women divided by the proportion of women in the group. Study 2 analysis excludes 22 groups from the semester when the team-building task was not conducted. Global assessment votes also exclude those groups for purposes of comparison. In both studies, results for global assessment votes are restricted to the first two months of assessments, which are the assessments prior to the leader intervention for Study 2. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Appendix Table A8: Main estimates corrected for multiple hypothesis testing

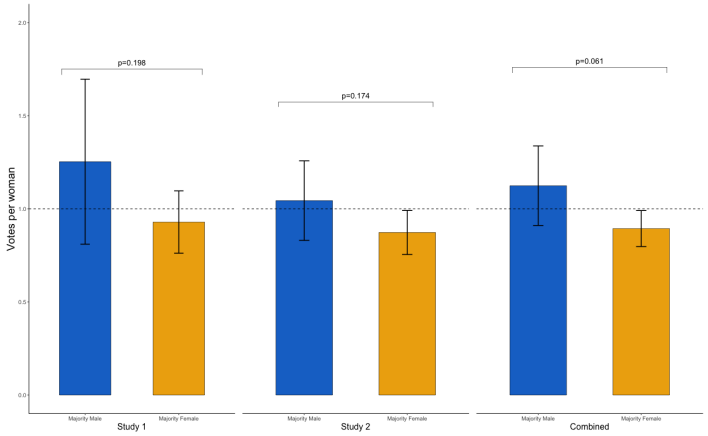
		p-value		
	Effect size	Remark 3.1	Theorem 3.1	Remark 3.7
Panel A: Composition effects, Study 1				
Team-Building Most Influential Votes	0.436	0.003	0.011	0.0113
Team-Building Incentivized Spokesperson Votes	0.343	0.067	0.067	0.0676
Global Assessment Most Influential Votes	0.316	0.064	0.123	0.1233
Team-Building Least Influential Votes	0.674	0.010	0.03	0.03
Panel B: Composition effects, Study 2				
Team-Building Most Influential Votes	0.291	0.007	0.028	0.028
Team-Building Incentivized Spokesperson Votes	0.232	0.024	0.068	0.068
Global Assessment Most Influential Votes	0.207	0.065	0.122	0.1223
Team-Building Least Influential Votes	0.043	0.765	0.765	0.7656
Panel C: Composition effects, Studies 1 + 2				
Team-Building Most Influential Votes	0.345	0.000	0.000	0.0003
Team-Building Incentivized Spokesperson Votes	0.272	0.006	0.017	0.0173
Global Assessment Most Influential Votes	0.249	0.007	0.015	0.015
Team-Building Least Influential Votes	0.283	0.034	0.034	0.0343
Panel D: Leadership effects, Study 2				
Team-Building Most Influential Votes	0.255	0.037	0.102	0.1026
Team-Building Incentivized Spokesperson Votes	0.411	0.000	0.000	0.0003
Global Assessment Most Influential Votes	0.210	0.044	0.086	0.086
Team-Building Least Influential Votes	0.094	0.495	0.495	0.4953

Note: Column (1) displays treatment effect estimates we report in Table 2 and Figure 3. Columns (2)-(4) display multiplicity-adjusted p-values computed using Remark 3.1, Theorem 3.1, and Remark 3.7 procedures from List et al. (2019). Inference remains unchanged after any of the corrections available in the implementation routines available from List et al. (2019).

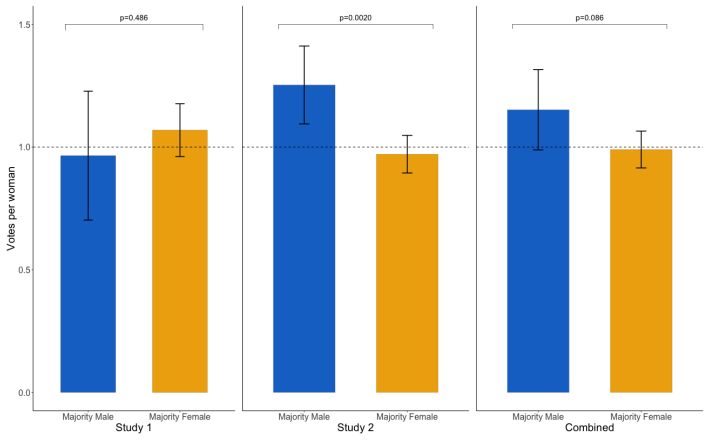
Appendix Figure A2: Women's Influence and Authority by Group Gender Composition



(a) Team-Building Task: Least Influential Votes



(b) Global Assessment: Least Influential Votes



(c) Team-Building Task: Most Supportive Votes

Note: 95% confidence intervals. Bars represent the average proportion of votes for women divided by the proportion of women in the group.

Appendix Table A9:
Individual and Group Performance on Team-building Task, Final Project, and Course Grades

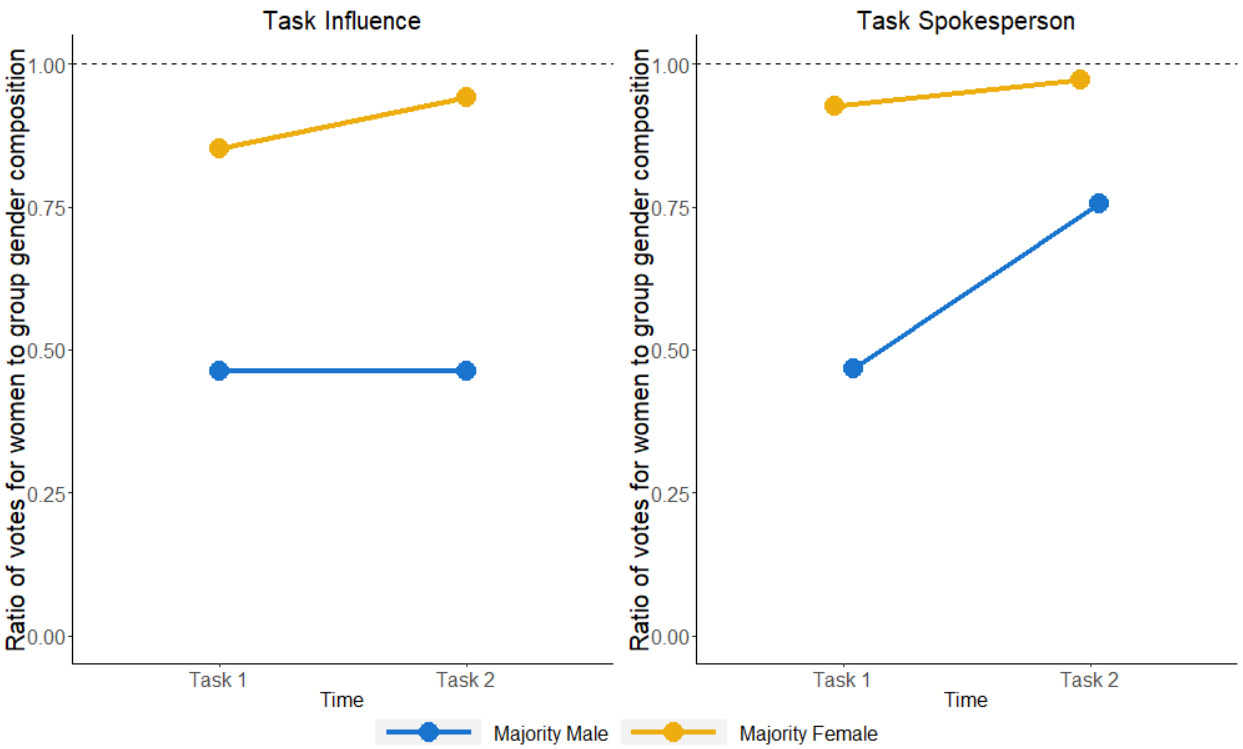
	(1) Individual Task Errors Studies 1 + 2	(2) Group Task Errors Studies 1+2	(3) Individual Project Score Study 2	(4) Group Project Score Study 2	(5) Group Average Grade Study 1	(6) Group Average Grade Study 2
Majority male group	-4.543***	-3.519**	-0.106	-0.116	-0.019	0.078
	(1.092)	(1.563)	(0.154)	(0.185)	(0.033)	(0.068)
N	198	198	812	144	75	145
Mean of outcome	42.22	46.84	11.27	12.72	3.56	2.99
St. dev. of outcome	8.01	11.15	1.82	1.11	0.14	0.41

Note: Heteroskedasticity-robust standard errors are reported in parenthees. In individual-level analysis, standard errors are clustered by group. Significance levels are indicated by $*<.1$, $**<.05$, $***<.01$.

Appendix Table A10: Women's Individual Pre-deliberation Task Performance Relative to Expert Scores and Implied Initial Performance of Group

	Study 1			Study 2	Studies 1 + 2
	Task 1	Task 2	Task 1 + 2	Task 1	All Combined
Panel A: Women's Average Initial Task Error Rate					
Majority male group	3.905*	-2.100	0.902	-0.829	-0.978
	(2.111)	(2.197)	(1.420)	(1.324)	(1.085)
Mean of outcome	46.05	63.41	54.73	46.59	49.08
St. dev. of outcome	11.39	11.76	7.59	11.49	11.10
Panel B: Women's Average Distance from Initial Group Average					
Majority male group	4.063**	1.358	2.710*	-1.070	-0.166
	(1.998)	(1.818)	(1.401)	(1.374)	(1.057)
Mean of outcome	30.04	39.16	34.60	32.81	33.36
St. dev. of outcome	10.81	9.72	7.57	11.92	10.80
N (both panels)	147	147	147	334	481

Appendix Figure A3: Study 1 Team-Building Task Measures Over Time



Appendix Table A11: Effect of Formal Leadership on Women's Influence, Study 2

	Pre-leader	Post-leader	Post-leader
Panel A: Team-Building Most Influential Votes			
Majority male group	-0.292***		-0.180
	(0.109)		(0.123)
Female Leader		0.255**	0.255**
		(0.124)	(0.123)
Mean of outcome	0.58	0.66	0.66
St. dev. of outcome	0.62	0.70	0.70
Panel B: Team-Building Incentivized Spokesperson Votes			
Majority male group	-0.232**		-0.039
	(0.098)		(0.115)
Female Leader		0.411***	0.411***
		(0.116)	(0.116)
Mean of outcome	0.55	0.72	0.72
St. dev. of outcome	0.56	0.67	0.67
Panel C: Global Assessment Most Influential Votes			
Majority male group	-0.207*		-0.088
	(0.108)		(0.106)
Female Leader		0.211**	0.210*
		(0.106)	(0.106)
Mean of outcome	0.75	0.79	0.79
St. dev. of outcome	0.61	0.60	0.60
Panel D: Team-Building Least Influential Votes			
Majority male group	0.044		-0.073
	(0.144)		(0.132)
Female Leader		-0.094	-0.094
		(0.134)	(0.135)
Mean of outcome	1.21	1.14	1.14
St. dev. of outcome	0.81	0.74	0.74

Note: Table presents group-level analysis of the effects of majority male groups and women leaders on perceptions of influence. The dependent variable is influence or spokesperson votes for women divided by the proportion of women in the group. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Appendix Table A12: Who Received Influence Votes in Groups with a Female Leader?

	Proportion of Votes Received by	
	Leader	Other Women
Most Influential Votes		
Majority Male	0.48	0.52
Majority Female	0.44	0.56
Spokesperson Votes		
Majority Male	0.84	0.16
Majority Female	0.76	0.24

Note: Cell entries represent the proportion of influence or spokesperson votes received by the woman randomly chosen as the group's leader and by other women in the group. Study 2 only.

Appendix Table A13: Effects of Formal Leadership and Gender Composition, Study 2

	(1)	(2)	(3)	(4)
	Month 1	Month 2	Month 3	Month 4
Female Leader	—	—	0.310**	0.181*
			(0.123)	(0.103)
Majority Male Group	-0.100	-0.219*	-0.069	-0.109
	(0.112)	(0.121)	(0.122)	(0.102)
Constant	0.843***	0.869***	0.701***	0.758***
	(0.061)	(0.066)	(0.093)	(0.081)
Observations	145	145	145	145
R^2	0.005	0.022	0.044	0.028

Note: Table presents group-level analysis of global assessments for each month. The dependent variable is the perceived influence of women divided by the proportion of women in the group. Heteroskedasticity-robust standard errors are reported in parentheses. The 22 groups that did not complete the lab tasks are included in this analysis. Significance levels are indicated by $*<.1$, $**<.05$, $***<.01$.

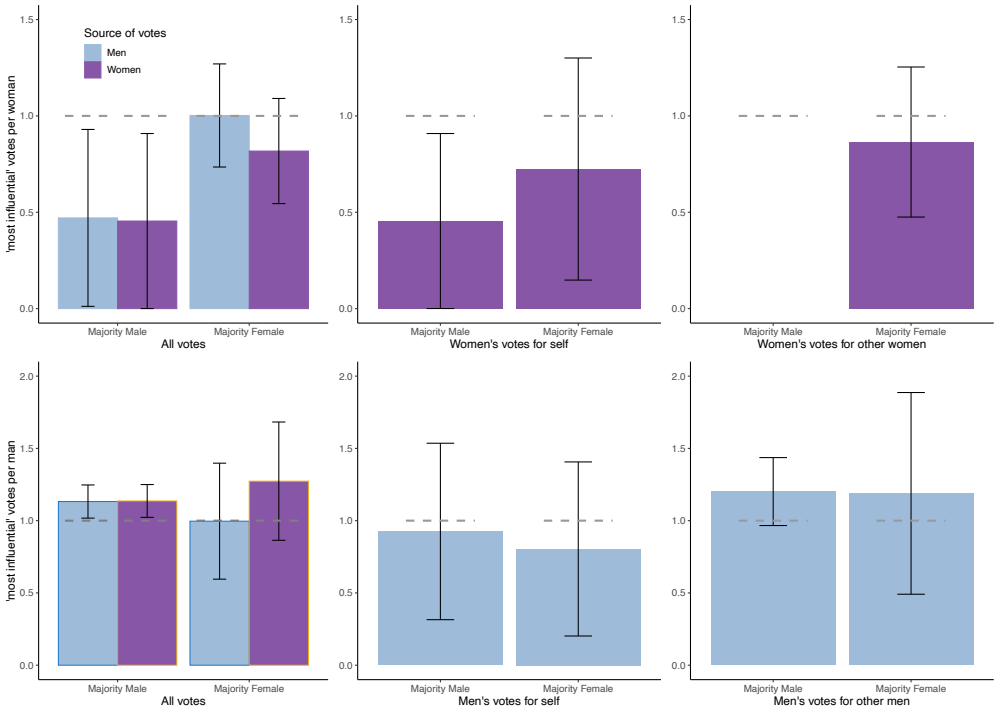
Appendix Table A14: Effects of Gender Composition Over Time, Study 1

	(1) Month 1	(2) Month 2	(3) Month 3	(4) Month 4
Majority Male Group	-0.448***	-0.218	-0.126	0.0144
	(0.154)	(0.207)	(0.198)	(0.194)
Constant	0.824***	0.891***	0.856***	0.896***
	(0.093)	(0.099)	(0.093)	(0.087)
Observations	75	75	75	75
R^2	0.102	0.014	0.005	0.000

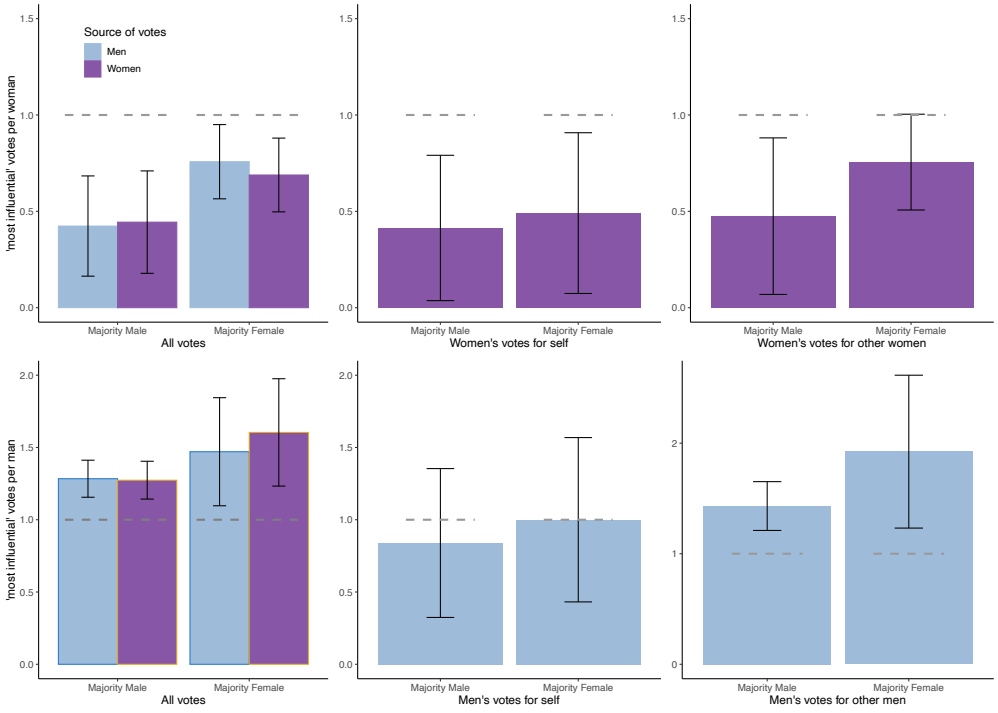
Note: Table presents group-level analysis of global assessments for each month. The dependent variable is the perceived influence of women divided by the proportion of women in the group. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels are indicated by $* < .1$, ** $< .05$, *** $< .01$.

Appendix Figure A4: Tracing the source of disempowerment: Votes per man/woman, by sources and across condition, separately by study

(a) Study 1

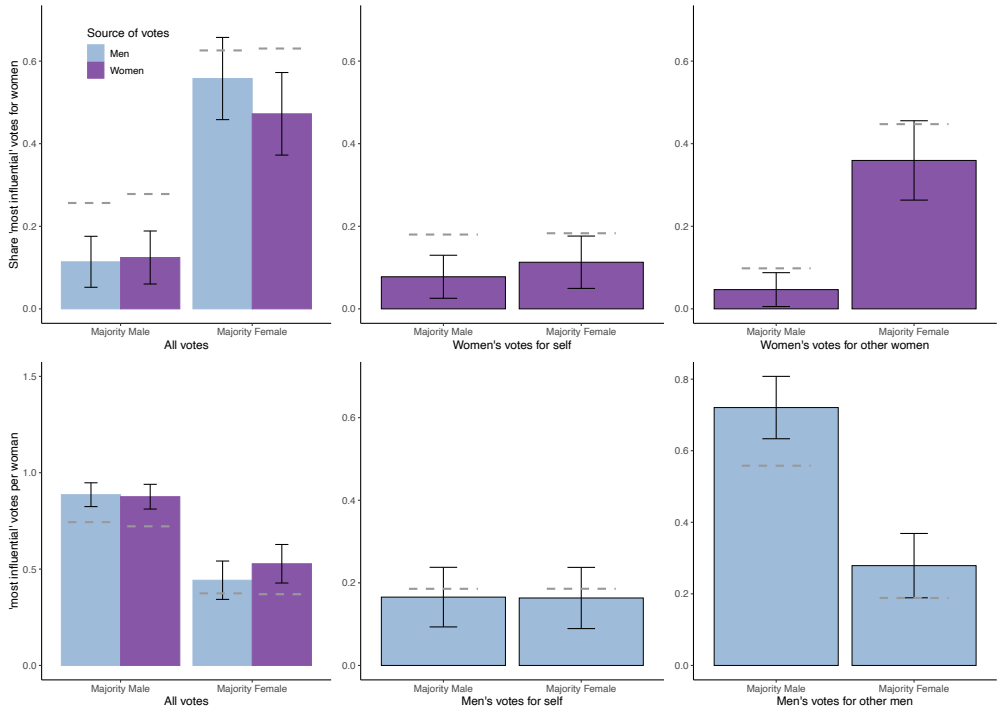


(b) Study 2



Note: 95% confidence intervals. In the top panel, bars represent the average proportion of votes for women divided by the proportion of women in the group. In the bottom panel, bars represent the average proportion of votes for men divided by the proportion of men in the group.

Appendix Figure A5: Tracing the source of disempowerment: Fraction of total votes for men and women, by source and across conditions



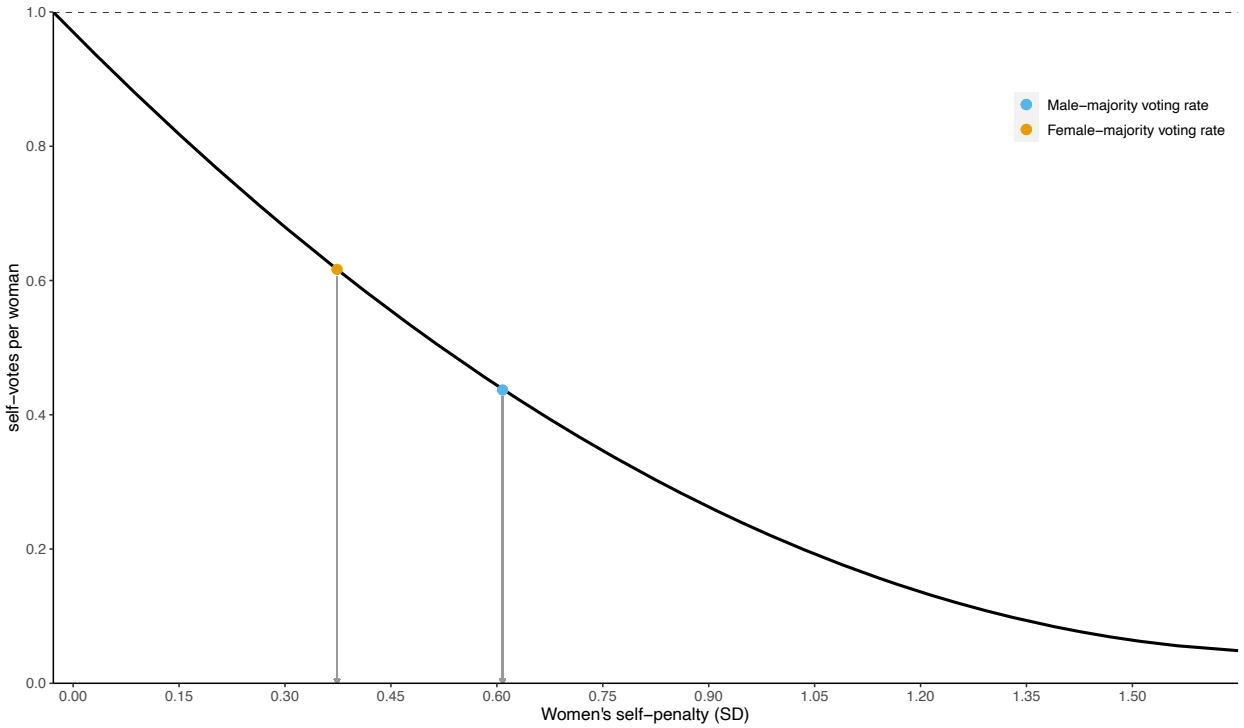
Note: 95% confidence intervals. In the top panel, bars represent the share of total votes for women. In the bottom panel, bars represent the share of total votes for men.

Appendix Table A15: Input Parameters and Simulation Results

Measure	Male-majority groups	Female-majority groups	source
Panel A: Simulation inputs			
Number of groups	102	96	Data
Women's rate of self-voting (normalized)	0.44	0.62	Data
Women's implied self-penalty (in SD)	0.61	0.37	Model
Panel B: Discrimination across conditions			
Votes per woman, task influence	0.45	0.79	Data
Implied discrimination penalty (in SD)	0.68	0.26	Model
Votes per woman, global influence	0.61	0.86	Data
Implied discrimination penalty (in SD)	0.43	0.13	Model
Votes per woman, incentivized spokesperson election	0.50	0.77	Data
Implied discrimination penalty (in SD)	0.59	0.30	Model

Notes: Table contains inputs parameters to and results of model simulations.

Appendix Figure A6: Simulated self-voting rates, by intensity of self-penalty



Note: Chart presents results of model calibration. The series is the fitted self-voting curve from 1,000 iterations of a simulation of the voting model described in Section 4 at each hundredth value of the indicated performance self-penalty among women. The indicated points placed on the curves reflect the same y-axis values as the corresponding quantities in Figure 5.

Appendix Table A16: Change in Sexism Over Time

	Pre-Treatment	End-of-Semester	Difference	p value
Men				
Female Leader	0.542	0.514	0.028	0.002
Male Leader	0.539	0.511	0.028	0.008
Difference-in-Differences			-0.0002	0.99
Majority Female	0.557	0.530	0.027	0.02
Majority Male	0.533	0.505	0.028	0.001
Difference-in-Differences			-0.0007	0.96
Women				
Female Leader	0.383	0.356	0.027	0.005
Male Leader	0.421	0.365	0.056	< 0.001
Difference-in-Differences			-0.029	0.038
Majority Female	0.408	0.367	0.041	< 0.001
Majority Male	0.390	0.346	0.044	0.007
Differences-in-Differences			-0.003	0.83

Note: Our measure of sexism is a standard measure of ambivalent sexism developed by Glick and Fiske (1996). Pre-treatment measure occurred prior to assignment to groups. End-of-semester measure occurred as part of the last global assessment survey. Difference is indicated by Pre-treatment minus End-of-semester. Standard errors are clustered by group.

Appendix Table A17: Relationship Between Experimental Conditions and Speaking Turns

	Study 1	Study 2	Studies 1 + 2
Majority male group	0.013**	-0.009	0.002
	(0.006)	(0.024)	(0.012)
Female	0.014	-0.040	-0.011
	(0.009)	(0.028)	(0.015)
Majority male group × Female	-0.027*	0.016	-0.003
	(0.014)	(0.043)	(0.024)
N	371	348	719
R^2	0.01	0.06	0.04
Mean of outcome	0.20	0.23	0.21
St. dev. of outcome	0.05	0.16	0.12

Note: Table presents individual-level analysis. The dependent variable is the individual's average proportion of speaking turns. Controls include the length of the group conversation and indicators for all-male and all-female conditions. The excluded category is majority women groups. Included controls are GPA, age, whether an individual is an international student, whether an individual is white, whether an individual is married, and an indicator for whether the individual's group met online. Heteroskedasticity-robust group-clustered standard errors are reported in parentheses. Significance levels are indicated by $*<.1$, $**<.05$, $***<.01$.

Appendix Table A18: Converting Speaking Turns to Influence

	Study 1		Study 2		Studies 1 + 2	
	(1) Majority Male Group	(2) Majority Female Group	(3) Majority Male Group	(4) Majority Female Group	(5) Majority Male Group	(6) Majority Female Group
Speaking turns	6.163***	8.212***	6.417***	7.117***	6.250***	6.906***
	(1.840)	(2.844)	(0.721)	(2.133)	(0.702)	(1.931)
Female	0.395	0.348	0.235	0.070	0.154	-0.071
	(0.450)	(0.837)	(0.206)	(0.456)	(0.208)	(0.408)
Speaking turns × Female	-5.065**	-3.639	-3.820***	-1.155	-3.744***	-1.086
	(2.213)	(4.481)	(1.208)	(2.385)	(1.141)	(2.188)
N	192	179	175	173	367	352
R^2	0.10	0.09	0.40	0.40	0.28	0.28
Mean of outcome	1.00	1.00	0.90	0.96	0.95	0.98
St. dev. of outcome	1.14	1.20	1.41	1.67	1.28	1.45

Note: Table presents individual-level analysis. The dependent variable is the individual's average number of influence votes. Heteroskedasticity-robust group-clustered standard errors are reported in parentheses. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Appendix Table A19: Converting Task Performance to Influence

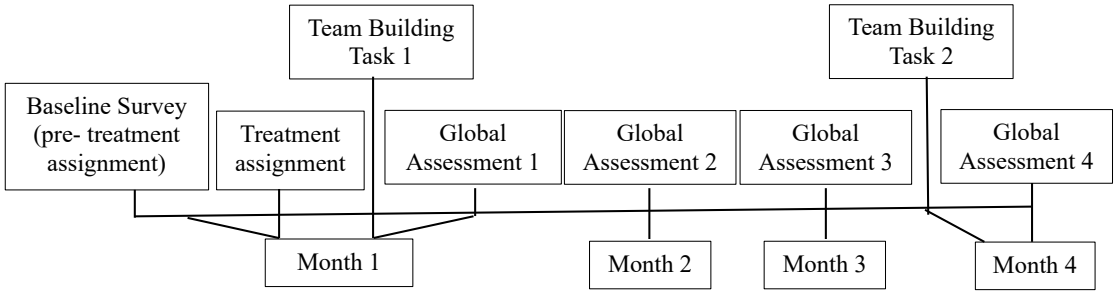
	Study 1		Study 2		Studies 1 + 2	
	(1) Majority Male Group	(2) Majority Female Group	(3) Majority Male Group	(4) Majority Female Group	(5) Majority Male Group	(6) Majority Female Group
Task Errors	-0.029**	-0.004	-0.021***	-0.001	-0.022***	-0.005
	(0.012)	(0.013)	(0.007)	(0.017)	(0.006)	(0.012)
Female	-2.638***	1.136	-1.440***	-0.559	-1.640***	-0.506
	(0.859)	(1.228)	(0.476)	(0.847)	(0.428)	(0.678)
Female Task Errors	0.039**	-0.026	0.014	-0.006	0.018**	-0.002
	(0.017)	(0.021)	(0.010)	(0.017)	(0.008)	(0.013)
N	192	179	351	327	543	506
R^2	0.09	0.04	0.11	0.06	0.10	0.04
Mean of outcome	1.00	1.00	0.96	0.97	0.97	0.98
St. dev. of outcome	1.14	1.20	1.45	1.59	1.35	1.46

Note: Table presents individual-level analysis. The dependent variable is the individual's average number of influence votes. Heteroskedasticity-robust group-clustered standard errors are reported in parentheses. Twenty-two students did not consent to the use of their individual survey data (two from Study 1 and twenty from Study 2), and their responses are not included in the results above. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

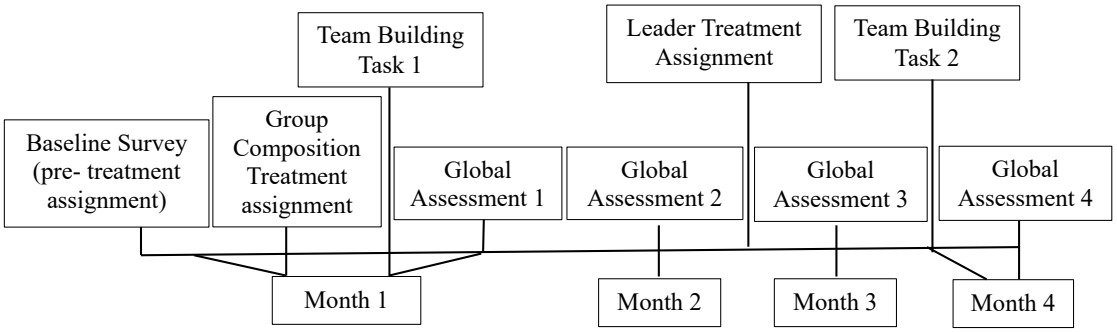
3 Experimental Protocol

3.1 Overview of Research Design

Appendix Figure A7: Research Timelines



(a) Study 1 Timeline



(b) Study 2 Timeline

3.2 Instructions for Team-Building Tasks

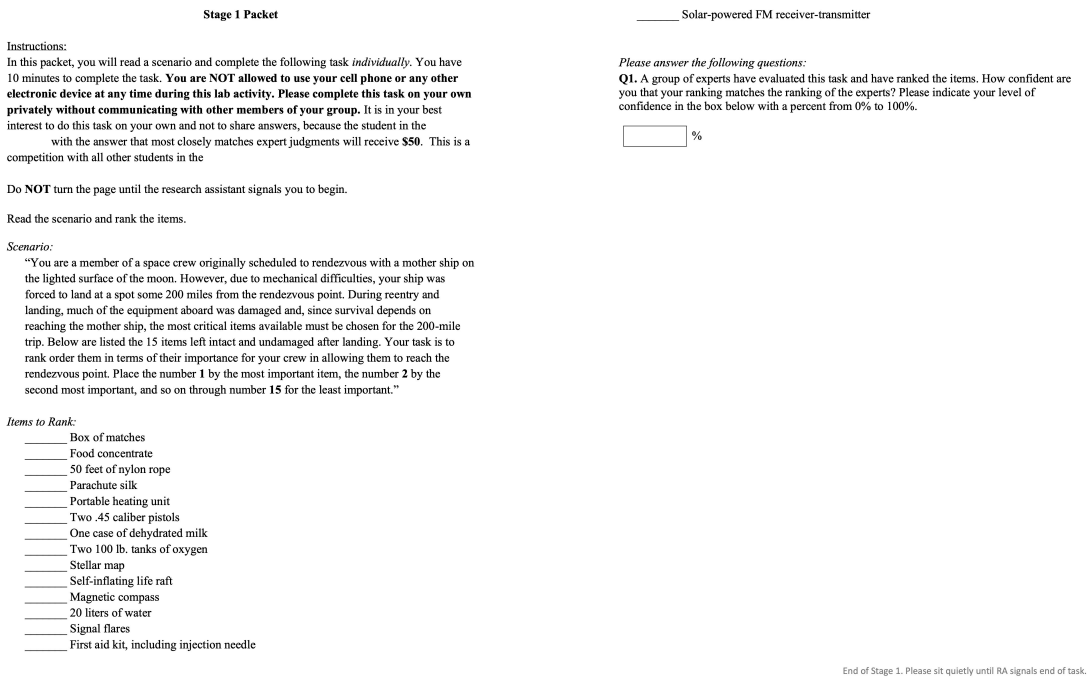
In both studies, the groups met twice a semester to complete a team-building exercise task (once in September, shortly after groups were assigned, and the second time in December, just prior to the end of the semester).

Each session consisted of three stages and subjects were given instructions for each stage separately. In Stage 1, subjects had 10 minutes to complete the ranking activity individually. In September subjects were given the "Survival on the Moon" ranking activity, and in December they were given the "Lost at Sea" activity. The research assistant remained in the room to prevent discussion, collaboration or cheating. Subjects were told that the person with the highest number of correct answers would be paid 50 dollars after the session. See Figures A8 and A9 below for specific subject instructions for the two tasks in Study 1. Study 2 instructions were identical, though some groups submitted their responses online.

In Stage 2, the group had 15 minutes to complete the same ranking task, but as a group. Groups were instructed to spend time discussing and completing the task together in collaboration with each other. Subjects were told that the group ranking that had the highest number of correct answers would receive 50 dollars per group member. See Figure A10 for specific subject instructions to this stage.

After this task was completed, each group member was asked to silently complete the exit questionnaire, which included questions about their group interaction during Stage 2 and a spokesperson vote. Specifically, students were asked to select one spokesperson per group. The spokesperson would present the results of the group ranking to a panel of judges at a later date and would have an opportunity to earn an additional 50 dollars for the group (see Figure A11).

Appendix Figure A8: Stage 1 - Individual Task Instructions for September Team-Building Exercise



Stage 1 Packet

Instructions:
In this packet, you will read a scenario and complete the following task *individually*. You have 10 minutes to complete the task. **You are NOT allowed to use your cell phone or any other electronic device at any time during this lab activity. Please complete this task on your own privately without communicating with other members of your group.** It is in your best interest to do this task on your own and not to share answers, because the student in the with the answer that most closely matches expert judgments will receive **$50**. This is a competition with all other students in the

Do **NOT** turn the page until the research assistant signals you to begin.

Read the scenario and rank the items.

Scenario:

"You are a member of a space crew originally scheduled to rendezvous with a mother ship on the lighted surface of the moon. However, due to mechanical difficulties, your ship was forced to land at a spot some 200 miles from the rendezvous point. During reentry and landing, much of the equipment aboard was damaged and, since survival depends on reaching the mother ship, the most critical items available must be chosen for the 200-mile trip. Below are listed the 15 items left intact and undamaged after landing. Your task is to rank order them in terms of their importance for your crew in allowing them to reach the rendezvous point. Place the number **1** by the most important item, the number **2** by the second most important, and so on through number **15** for the least important."

Items to Rank:

_____ Box of matches
_____ Food concentrate
_____ 50 feet of nylon rope
_____ Parachute silk
_____ Portable heating unit
_____ Two .45 caliber pistols
_____ One case of dehydrated milk
_____ Two 100 lb. tanks of oxygen
_____ Stellar map
_____ Self-inflating life raft
_____ Magnetic compass
_____ 20 liters of water
_____ Signal flares
_____ First aid kit, including injection needle
_____ Solar-powered FM receiver-transmitter

Please answer the following questions:

Q1. A group of experts have evaluated this task and have ranked the items. How confident are you that your ranking matches the ranking of the experts? Please indicate your level of confidence in the box below with a percent from 0% to 100%.

[ ] %

End of Stage 1. Please sit quietly until RA signals end of task.

Note: Packet also contained extra pages and space for notes.

Appendix Figure A9: Stage 1 - Individual Task Instructions for December Team-Building Exercise

Stage 1 Packet

Instructions:
In this packet, you will read a scenario and complete the following task *individually*. You have 10 minutes to complete the task. **You are NOT allowed to use your cell phone or any other electronic device at any time during this lab activity. Please complete this task on your own privately without communicating with other members of your group.** It is in your best interest to do this task on your own and not to share answers, because the student in the with the answer that most closely matches expert judgments will receive **$50**. This is a competition with all other students in the Program.

Read the scenario and rank the items.

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 15 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 is to rank the 15 items in terms of their importance for you, as you wait to be rescued. Place the number **1** by the most important item, the number **2** by the second most important and so forth until you have ranked all **15** items."

Items to Rank:

_____A sextant
_____A mirror
_____A quantity of mosquito netting
_____A 25-liter container of water
_____A case of ready-to-eat meals
_____Maps of the Atlantic Ocean
_____A floating seat cushion
_____A 10 liter can of oil/petrol mixture
_____ A small transistor radio
_____20 square feet of opaque plastic sheeting
_____A can of shark repellent
_____One bottle of 160 proof rum
_____15 feet of nylon rope
_____2 boxes of chocolate bars
_____An ocean fishing kit & pole

Please answer the following questions:

Q1. A group of experts have evaluated this task and have ranked the items. How confident are you that your ranking matches the ranking of the experts? Please indicate your level of confidence in the box below with a percent from 0% to 100%. Please use whole numbers.

[ ] %

Note: Packet also contained extra pages and space for notes.

Appendix Figure A10: Stage 2 - Group Task Instructions for September Team-Building Exercise

Stage 2 Packet

Instructions:
In this packet you will read the same scenario as before and complete the following task *as a group*. You will have **15 minutes** to discuss and complete the task. Your discussion will be audio-recorded. A research assistant will notify you when you have 5 minutes and 1 minute left to complete the task. When the 15 minutes have finished, you will turn in the packet to the research assistant. It is in your best interest to complete this task to the best of your group's ability, because the group with the best answer will receive **$50 per group member**. It is also important that you **not** share information about this activity with other individuals or groups so that your chance of winning does not decrease. This is a competition with other groups in the

You may begin the task when the research assistant leaves the room.

If you choose, you may use the scratch paper and pens provided. There are five pieces of scratch paper in the packet. Please remain in your seats and do not use the white board.

Discuss the scenario with your group members and rank the items accordingly.

Scenario:

> "You are a member of a space crew originally scheduled to rendezvous with a mother ship on the lighted surface of the moon. However, due to mechanical difficulties, your ship was forced to land at a spot some 200 miles from the rendezvous point. During reentry and landing, much of the equipment aboard was damaged and, since survival depends on reaching the mother ship, the most critical items available must be chosen for the 200-mile trip. Below are listed the 15 items left intact and undamaged after landing. Your task is to rank order them in terms of their importance for your crew in allowing them to reach the rendezvous point. Place the number **1** by the most important item, the number **2** by the second most important, and so on through number **15** for the least important."

Items to Rank:

_____ Box of matches
_____ Food concentrate
_____ 50 feet of nylon rope
_____ Parachute silk
_____ Portable heating unit
_____ Two .45 caliber pistols
_____ One case of dehydrated milk
_____ Two 100 lb. tanks of oxygen
_____ Stellar map
_____ Self-inflating life raft
_____ Magnetic compass
_____ 20 liters of water
_____ Signal flares
_____ First aid kit, including injection needle
_____ Solar-powered FM receiver-transmitter

Note: Packet also contained extra pages and space for notes.

Appendix Figure A11: Stage 3 - Post-Task Survey for Team-Building Exercise

Stage 3 Packet

Name: _____

Net ID: _____

(This information will be used to connect your responses to your group for purposes of analysis. No identifying information will ever be shared or published, and your responses will be kept completely confidential at all times.)

Instructions:
In this packet you will answer a series of confidential feedback questions about your group discussion and decisions. **Please complete this task on your own privately without communicating with other members of your group.**

For each of these questions, please answer as honestly as possible in regards to how you personally felt about the group interaction.

The answers you provide here will in no way affect your payment or any other part of your experience today. We simply want to know your honest impressions about the conversation you just had with the other members of your group. Your answers will never be shared with other members of your group.

You have 10 minutes to complete the questionnaire. When you have finished, please exit the room with your belongings and Stage 3 Packet. Leave the clipboard and pen on your chair. DO NOT TAKE ANY LAB MATERIALS WITH YOU. Please exit quietly so as not to disturb others while they finish the final stage. The research assistant will collect the packet from you and give you your base payment for your participation in the activity. You will be notified of the competition results once every group has had an opportunity to participate in the lab activity.

Please sign below to state that you will NOT share information about this lab activity with other students in the program until after all groups have had the opportunity to participate in the activity and the presentations have been given. This is in your best interest, because sharing information will decrease your chance and your group's chance of winning the cash prizes.

Signature: _____ Date: _____

Q1. A group of experts have evaluated this task and have ranked the items. How confident are you that your ***group's*** ranking from Stage 2 matches the ranking of the experts? Please indicate your level of confidence in the box below with a percent from 0% to 100%. Please use whole numbers.

☐ %

Q2. Have you ever done the task – the scenario and item ranking – found in Stages 1 and 2 before in another setting?

☐ Yes ☐ No

Q3. Below you will find a series of statements about the group discussion that you may or may not agree with. Please indicate the extent to which you agree or disagree with each of the following statements by clicking on the appropriate button. Remember, there are no right or wrong answers to any of these questions. We are only interested in your own personal opinions.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
My opinions were influential in shaping the group discussion and final decision.	❍	❍	❍	❍	❍
I feel like my voice was heard during the group discussion.	❍	❍	❍	❍	❍
I would describe myself as a leader in the group discussion.	❍	❍	❍	❍	❍
The group discussion helped me better understand the different ranking possibilities.	❍	❍	❍	❍	❍
Group work made everything slower and harder to accomplish in Stage 2 compared to Stage 1.	❍	❍	❍	❍	❍
Disagreement among group members made our discussion difficult.	❍	❍	❍	❍	❍
A few people dominated the discussion.	❍	❍	❍	❍	❍
All different perspectives were welcome in our discussion.	❍	❍	❍	❍	❍
Members of my group treated each other with respect and courtesy.	❍	❍	❍	❍	❍
Members of my group were too quick to agree with each other.	❍	❍	❍	❍	❍

Q4. Of the group members present, select the member who was the ***most*** influential member of your group during the group discussion. This can include yourself. (Indicate using the letter on the nameplate in front of the group members or refer to the seating chart sheet attached to this packet. Even though you may feel that several members of the group were influential, please mark only *one* option. Choose the person you feel was ***most*** influential. If you are in a group of four members, do not mark option E.)

☐ A ☐ D
☐ B ☐ E
☐ C

Q5. Of the group members present, select the member who was the ***least*** influential member of your group during the group discussion. This can include yourself. (Indicate using the letter on the nameplate in front of the group members or refer to the seating chart sheet attached to this packet. Even though you may feel that several members of the group were not influential, please mark only *one* option. Choose the person you feel was ***least*** influential. If you are in a group of four members, do not mark option E.)

☐ A ☐ D
☐ B ☐ E
☐ C

Q6. Of the group members present, select the member who was the ***most*** supportive member of your group during the group discussion. This can include yourself. (Indicate using the letter on the nameplate in front of the group members or refer to the seating chart sheet attached to this packet. Even though you may feel that several members of the group were supportive, please mark only *one* option. Choose the person you feel was ***most*** supportive. If you are in a group of four members, do not mark option E.)

☐ A ☐ D
☐ B ☐ E
☐ C

Q7. Of the group members present, select the member who was the ***least*** supportive member of your group during the group discussion. This can include yourself. (Indicate using the letter on the nameplate in front of the group members or refer to the seating chart sheet attached to this packet. Even though you may feel that several members of the group were not supportive, please mark only *one* option. Choose the person you feel was ***least*** supportive. If you are in a group of four members, do not mark option E.)

☐ A ☐ D
☐ B ☐ E
☐ C

Q8. How satisfied or dissatisfied are you with your group discussion?

☐ Very Satisfied
☐ Somewhat Satisfied
☐ Neither Satisfied nor Dissatisfied
☐ Somewhat Dissatisfied
☐ Very Dissatisfied

Q9. How satisfied or dissatisfied are you with your group's final ranking?

☐ Very Satisfied
☐ Somewhat Satisfied
☐ Neither Satisfied nor Dissatisfied
☐ Somewhat Dissatisfied
☐ Very Dissatisfied

Q10. Which of the following best captures how your group made a decision about the ranking?

☐ One or two group members made the decision for the group, but in the end, most others weren't sure or would have preferred something different.
☐ Three or four group members made the decision, but in the end, there were one or two who weren't sure or would have preferred something different.
☐ All members of the group made the decision together, and in the end, no members of the group weren't sure or would have preferred something different.

Q11. Is there anything else you would like to tell us about your group's discussion or decision? Please enter your comments below.

Q12. Finally, we would like you to select one team member who you would like to be the spokesperson for your group. This can include yourself. After all groups have participated in this lab activity, 5 groups will be selected at random for an opportunity to present and defend their group's choice for rankings. Each group must select one member to be their spokesperson. The spokesperson will give a 5-minute presentation explaining why their group chose the ranking order that they did. Whoever makes the most convincing case will earn another $50 for their group. These "presentations" will occur at a later time after all groups have had an opportunity to participate in the lab activity. The spokesperson will be chosen by majority vote. In the case where no member receives majority vote, a spokesperson will be randomly chosen from the selected group.

Please mark who you would like to select as your group's spokesperson. (Indicate using the letter on the nameplate in front of the group members or refer to the seating chart sheet attached to this packet. Even if you think multiple people could serve as an effective spokesperson, please mark only *one* option. Choose the person you think would be the best spokesperson for your group. If you are in a group of four members, do not mark option E.)

☐ A ☐ D
☐ B ☐ E
☐ C

End of Stage 3.
You may exit the room. Take your belongings. Leave the clipboard and pen on your chair, and bring this Stage 3 Packet to RA.

Note: Packet also contained extra pages and space for notes.

3.3 Ethics Statement: IRB Approval and Informed Consent

Study 1 was reviewed and approved by the [REDACTED] Institutional Review Board as Project X15228 in 2015. It was later renewed as Project X17352. Study 2 was reviewed and approved by the [REDACTED] Institutional Review Board as Project X18455. Participants in both studies were fully informed that they were participating in a research study, and no deception was used in the experiment. All participants could freely grant or deny consent for their work to be included in the research study.