Experimental Economics (2025), 1–25
doi:10.1017/eec.2025.10024

ORIGINAL PAPER

Beyond representation: the importance of the decision-making process in hiring decisions

Prachi Jain and Michelle M. Miller

Department of Economics, Loyola Marymount University, Los Angeles, CA, USA
Corresponding author: Prachi Jain; Email: prachi.jain@lmu.edu

This study was reviewed and determined to be exempt by Loyola Marymount University's IRB (2019 SP 72-R). It was pre-registered with the AEA RCT Registry, ID AEARCTR-0010775. A copy of our pre-registration can be found in Appendix F.

(Received 1 August 2024; revised 11 Februray 2025; accepted 2 July 2025)

Abstract

An abundance of statistics has shown gender disparity in hiring decisions. This paper shows that a previously unexplored factor, the decision-making process utilized by a hiring committee, plays a crucial role. Using a laboratory experiment, we find that gender disparity is eliminated when hiring decisions are made unanimously by a group. By comparison, we find that gender disparity is largest when decisions are made by a leader who volunteers. We do not find evidence of heterogeneity by gender as the results persist regardless of the number of women in the group or the leader's gender. The experimental design allows us to rule out several possible mechanisms including differences in leadership characteristics and communication styles.

Keywords: Decision-making; discrimination; gender; hiring processes; laboratory experiments; teams

JEL Codes: C92; D81; J23; J71; M51

1. Introduction

Each year, firms spend vast amounts of money and resources on their diversity programs in the hopes of improving representation. Indeed, in the United States alone, it is estimated that companies spend approximately $8 billion a year on diversity, equity, and inclusion related initiatives (Williams, 2021). Examples of such initiatives include diversity training of search committees, development of inclusive hiring practices, networking programs, and mentoring programs. However, many of these initiatives have been found to be ineffective. For example, Kalev et al. (2006) found that diversity training does not increase diversity. Indeed, Bohnet (2016) found that diversity training can backfire – training designed to raise awareness of gender inequality may end up making gender more salient and thereby actually highlight differences.

Without any clear tool to improve diversity, the way forward is uncertain. Women remain underrepresented in business, government, and academia. Indeed, despite comprising 57 percent of the labor force in the United States (Bureau of Labor Statistics, U.S. Department of Labor, 2024), women comprise only 10 percent of chief executive officers in Fortune 500 companies (Hinchliffe, 2023), 27 percent of the workforce in science, technology, engineering, and math (Martinez & Christnacht, 2021), 28 percent of Congress (Leppert & DeSilver, 2023), and 36 percent of full professors (Palmer, 2023). In order to improve gender disparity, we need to study how hiring decisions are made.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Studying how hiring decisions are made in real-world settings poses several challenges: the manner by which decisions are made, and the group interactions are difficult, if not impossible for the researcher to observe. To overcome these challenges, we employ a laboratory experiment. A controlled laboratory experiment with a chat interface allows us to randomize the decision-making processes of hiring committees while observing the free-form conversations of group members. Thus, with a laboratory experiment we are better able to examine how decision-making processes affect group dynamics than we would be with observational data.

In our experiment, participants were placed on a hiring committee and asked to 'hire' one candidate from a set of three. Participants were provided information about each candidate's gender, age, major, and past performance, thus mimicking real-world resumes in which gender, age, and past performance can be inferred from a candidate's name, professional experience, education, skills, certification, and so on. Participants were then able to chat with other members of their hiring committee using a chat interface before making their hiring decision.

Participants were assigned one of four decision-making treatments: majority vote, unanimous vote, a decision by a leader who is randomly assigned (i.e., exogenously appointed), and a decision by a leader who volunteers (i.e., a self-selected leader). We find this setup relevant for hiring decisions as they are are often made in groups, be it a hiring committee, search advisory committee, or interview by a panel. Sometimes these groups vote by majority or unanimous rule. Other times the hiring decisions are made by a leader, such as a hiring manager, academic dean, or business owner.

We find that the decision-making process crucially affects gender representation in hiring decisions. On average, a female candidate is 9.5 percentage points less likely to be hired than a comparable male candidate. Gender differences disappear when hiring decisions are made unanimously by a group. In contrast, gender differences are largest when decisions are made by a leader who volunteers. Indeed, when a decision is made by a leader who volunteers, a female candidate is 15.8 percentage points less likely to be hired than a comparable male candidate.

Since we observe the free-form conversations within groups, we are able to examine the mechanism by which the decision-making process affects hiring decisions. Because a unanimous vote requires agreement from all team members, this decision-making process elicits additional discussions that are more substantive in nature compared to a majority vote. Additionally, under the unanimous voting rule, group members interact using more positive sentiments. In contrast, when a decision is made by a leader, the leader's preference is often viewed as a default decision that can only be overturned by other clear improvements. As a result, group members may only voice their opinions if they are convinced that their preferred candidate is better, and they are willing to challenge the leader. Consistent with the literature, we find that groups have more substantive discussions under the leader treatments compared to decisions made using a majority vote. Exploring the mechanism by which leaders affect group decisions, we find selection into leadership; women and less risk-averse individuals are more likely to volunteer to be a leader. But we do not find differences in communication or leadership styles between randomly selected and volunteer leaders.

We also consider heterogeneity in our results. First, we examine whether the decision-making process has a varied impact depending on the number of women in the group. This analysis is inspired by the conflicting results regarding the impact of a group's gender composition on hiring decisions (Bagues et al., 2017; Dominguez, 2023; Mengel, 2021; De Paola & Scoppa, 2015). The literature suggests several reasons for this heterogeneity. For example, we will observe heterogeneous effects if women show preference for or are less discriminatory towards women (Carlsson & Eriksson, 2019; Chen & Li, 2009; Daskalova, 2020; Hewstone et al., 2002). Additionally, we will observe heterogeneity if women have different beliefs about women's performance. This could occur because they share similar backgrounds or skills (Cornell & Welch, 1996). Furthermore, heterogeneity could be driven by gender differences in group dynamics; prior work has shown that women are less likely to speak up (Karpowitz et al, 2012) or contribute their ideas (Coffman, 2014). Finally, we will observe heterogenous effects if the presence of women in the group affects the behavior of male group members

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

(Bagues & Esteve-Volart, 2010). We randomize the gender composition of the groups and find some evidence of such heterogeneity. When decisions are made by majority vote or a leader, gender disparity in hiring is similar regardless of the group's gender composition. However, when groups decide using unanimous voting, increasing the share of women in the group increases the likelihood that a female candidate is hired.

Second, we examine whether the decision-making process has a varied impact depending on the leader's gender. This type of heterogeneity is motivated by evidence that the leader's gender affects their decisions and that men are more likely to self-select into leadership (Ertac & Gurdal, 2012). Within a decision-making process, female and male leaders do not differ in their likelihood of hiring a female candidate. Meaning, discrimination is persistent and substantive in magnitude regardless of the leader's gender. Thus, gender differences in self-selection into leadership cannot fully explain our finding that gender differences are exacerbated when decisions are made by a leader who volunteers compared to decisions made by a randomly selected leader.

Our results suggest that firms can improve gender diversity by adjusting *how* hiring committees make decisions. Changing the decision-making process utilized by a group is relatively simple and straightforward. Our study shows it can yield immediate reductions in gender disparities in hiring. By comparison, policies that aim to 'break the glass ceiling' by improving gender diversity in leadership positions are larger institutional changes that may not have their intended effects and could take more time to yield downstream effects on the hiring or promotion of women.

2. Literature review

We consider how four decision-making processes impact hiring decisions: majority vote, unanimous vote, a decision by a leader who is randomly assigned (i.e., exogenously appointed), and a decision by a leader who volunteers (i.e., a self-selected leader). We base these treatments on the decision rules used in a range of real-world settings, including, but not limited to, hiring decisions. For example, majority voting rules are used to make decisions in the United States House of Representatives, Senate, and Supreme Court. On the other hand, in the United States, juries in criminal cases make decisions using a unanimous vote. Examples of random leaders include government agencies assigning lower-level officers to oversee local decision-making and a firm assigning an employee to organize a business event. Finally, in other settings, leaders are asked to volunteer – for example, academic department chairs are often selected among those who volunteer. While our random leader treatment is perhaps least representative of real-world decisions, we include this treatment to explore how self-selection into a leadership role affects hiring decisions (as in Ertac & Gurdal, 2012).

A large literature has examined the impact of the decision-making process in other contexts including risk-taking games (Ertac & Gurdal, 2012), public goods games (Arbak & Villeval, 2013), jury games (Goeree & Yariv, 2011), and sequential search (J. Chan et al., 2018; Mak et al., 2019). However, this paper is (to the best of our knowledge) one of the first to examine how the decision rule utilized by a hiring committee affects *who* is hired.[1]

The costs and benefits of these decision-making processes have been discussed in numerous theoretical and experimental works. For example, consider the most commonly employed voting rule: majority rule. Theory suggests that this rule has a conflicting impact on participation. On the one hand, there may be broad participation within the committee as everyone has an equal vote. Group members may feel more comfortable expressing their opinions (Austen-Smith & Feddersen, 2005), especially if they know that they are not the pivotal vote (Dessein, 2007). However, the members of the committee who hold that majority opinion may dominate the conversation, with minority opinion members contributing less and their contributions being undervalued (Lorenz et al., 2015; Stasser

[1] A related literature explores how institutional details or features of the decision-making process, such as the order of the candidates (Kessler et al., 2024), whether the committee deliberates by chatting with each other (Mengel, 2021), or the number of candidates shown (Batista Pereira, 2023; Bohnet et al., 2016), affects discrimination in hiring decisions.

& Abele, 2003). Because groups only need to reach a simple majority, conversations are predicted to be faster (Hastie & Kameda, 2005) and more polarized (Miller, 1985) than under a unanimous vote.

By comparison, because a unanimous vote requires agreement from all team members, many works theorize that participation will increase and be more equitable across group members (J. Chan et al., 2018). Thus, participants will be more likely to share their diverse perspectives. As a result, outcomes may be more efficient (Breitmoser & Valasek, 2017; Mak et al., 2019). However, some works suggest that a unanimous vote could have the opposite effect if participants feel pressure to conform (Guarnaschelli et al., 2000). For example, under a unanimous voting rule, group members may take their (perceived) group members' preferences into account (Daskalova, 2020), suppress their opinion if they do not think other group members share their views (Austen-Smith & Feddersen, 2005; Daskalova, 2020), and vote in line with the rest of the group. Despite these conflicting predictions, on average, prior works suggest that the unanimous voting rule will encourage inclusivity, produce higher quality discussions, elicit lengthier discussions, and increase group members' sentiment when compared to a majority voting rule.

How does this compare to decisions made by a leader? The leader has a preference about who to hire as well as the final authority to implement this decision. This preference can be viewed as a default decision that can only be overturned by other clear improvements (Dessein, 2007). As a result, the leader may dominate the discussion, with group members deferring to the leader's judgement and leaders skewing discussions in favor of their preference (Henningsen et al., 2004). Group members may only voice their opinions if they are convinced that their preferred candidate is better and if they are willing to challenge the leader (Dessein, 2007). Therefore, when decisions are made by a leader, there is typically less discussion, but the quality of discussion is better (Dessein, 2007). Additionally, the lack of discussion may mean that decisions will not integrate diverse perspectives (Schippers & Rus, 2021) and may be less efficient (Hastie & Kameda, 2005). (By comparison, under majority rule, there is no such default decision. As a result, under majority rule, all group members have an incentive to discuss their preferred candidate, regardless of their argument's merits. Thus, under the majority rule, there will be more discussion, but the quality of the discussion may be lower.)

We consider two leader processes in order to examine how self-selection into a leadership position impacts hiring decisions. In our setting, there is no financial incentive to become the leader. However, there are several reasons participants may want to become leaders, such as wanting to influence the group's hiring decision or enjoying the leadership role. Prior work has noted the numerous ways in which volunteer leaders differ from randomly appointed leaders. For example, prior work has shown that volunteer leaders have different personality traits. Volunteer leaders are more likely to exhibit Big-5 personality traits such as neuroticism and extraversion (Judge et al., 2002). Self-selected leaders are also more likely to be self-confident (Reuben et al., 2012), less likely to be risk averse (Ertac & Gurdal, 2012; K. Y. Chan et al., 2015),[2] and more likely to avoid public scrutiny (Alan et al., 2020). Additionally, individuals may be less likely to lead if they do not believe they will be influential (Born et al., 2022) or if they are sensitive to potential backlash (Chakraborty & Serra, 2024). We also consider two leader processes because prior work has shown that the source of leadership authority affects how leaders act. For example, random leaders encourage participation and try to be more democratic (Kocher et al., 2013). Finally, prior work has shown that the leader's source of authority impacts how the group members perceive the leader. In the random leader treatment, group members may be more likely to speak up and less likely to defer to the leader if they see the leader's role as arbitrary. Given these differences, we would expect that a group with a volunteer leader will have less discussion than a group with a randomly selected leader. Without a thorough discussion, decisions may not integrate diverse

[2] In our context, risk aversion may influence gender disparities in hiring decisions if female candidates are perceived to be riskier than similar male candidates. This may occur if there are (perceived) differences across groups' distributions (Stiglitz, 1973). Risk aversion may also affect the decision to volunteer to be a leader. For example, risk-averse individuals may want to make the decision to avoid the risk of someone else making the decision. Alternatively, risk-averse individuals may be less likely to take on the burden or risk of making decisions on behalf of others (Ertac & Gurdal, 2012).

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

perspectives and may be less efficient. However, ultimately, these outcomes depend on the leader's management style and communication (Ertac & Gurdal, 2019; Kocher et al., 2013). Additionally, the group's sentiment level may depend on how the leader and group members interact.

In this paper, we consider the impact of these decision-making processes in the context of hiring decisions – which of these decision-making processes will improve representation and which of these decision-making processes will exacerbate disparity? We also consider heterogeneity by gender – specifically, heterogeneity by the group's gender composition and heterogeneity by the leader's gender. We ask whether (1) increasing the number of women in a group has a different impact on discrimination that varies with the decision-making process and (2) whether having a female leader has a different impact on discrimination that varies with the leader's authority (randomly appointed versus volunteer). We consider such heterogeneity because prior works indicate an interaction between voting rules and gender in other settings. They showed that under the majority voting rule women participate more as the number of women in the group increases (Karpowitz et al., 2012). By comparison, because the unanimous voting rule is more inclusive of all minorities, women do not experience the same benefit from higher numbers under this rule. Instead, the unanimous rule benefits both genders when they are in the numerical minority (Karpowitz et al., 2012). We examine whether such heterogeneity exists in the context of hiring and gender disparity. Additionally, prior works indicated an interaction between the decision-making process and the leader's gender in other settings. Prior works have shown that men are more likely to volunteer (Arbak & Villeval, 2013) and that men have different leadership styles (Ertac & Gurdal, 2019). Thus, our second research question contributes to the literature by examining the interaction between the decision-making process (i.e., whether the leader is appointed or volunteers) and the leader's gender in the context of gender disparities in hiring.

3. Experimental design

In the main part of the experiment, participants played the role of evaluators. Similar to Dominguez (2023) and Bohnet et al. (2016), participants were given a pool of three resumes with information on a candidate's age, gender, college major, and signal of performance. Participants were then assigned to a three-person group (referred to as a hiring committee). The group was given the opportunity to communicate using a chat interface before hiring one candidate. Groups varied in terms of the decision-making process and their gender composition. In this section we provide detailed information on our experimental design.

3.1. Stage 1

To generate the resumes for the main part of the experiment we first implemented a study on Amazon's Mechanical Turk (MTurk).

In this first stage, 107 individuals participated as 'candidates.' All candidates were college graduates living in the United States. These candidates were asked to perform two rounds of a mathematical matrix search task (Cassar & Rigdon, 2021). For this task, candidates were presented with a series of three-by-three grids, with each cell containing a number with one decimal place. They were then asked to identify the two numbers in the grid that add up to 10.0 exactly. Candidates were asked to solve as many grids as possible in two minutes. (Candidates were paid a show-up fee of 50 cents plus 10 cents per correct answer.) Candidates also completed a short demographic questionnaire – information from this survey was used to create resumes for the second stage of the experiment. Instructions from the MTurk study are included in Appendix C. (MTurk participants were not informed about the second stage of the study.)

Similar mathematical tasks are commonly employed in the literature because they are male stereotyped but gender neutral in performance (Cassar & Rigdon, 2021; Niederle & Vesterlund, 2007). In

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Appendix B, we confirm that there is no significant difference in performance by gender and further validate the task's gender stereotype. Specifically, Table B1 shows that gender is insignificantly related to performance while Table B2 shows that the task is male stereotyped.

Of the 107 subjects that participated in the first stage, we generated 13 pools (i.e., sets) of resumes, with each pool containing three resumes. The resumes contained information on the candidate's age, gender, college major, and a signal of performance. The signal of performance consisted of the number of correct calculations done in the first round.[3] Each pool contained a female candidate and a male candidate with similar signals of performance, that is, with the male and female candidate scoring within one to two points of each other. Additionally, each pool contained a third 'dominated' candidate who had a signal of performance that was lower than the other two candidates.[4] In 8 of the 13 pools, the female candidate had the highest signal of performance. In these pools, the evaluator had to choose between a lower-performing candidate whose gender aligns with a favorable stereotype and a higher-performing candidate associated with an unfavorable stereotype – this is where discrimination is most likely to occur. Resume pools were constructed so that the candidate's relative age (whether the candidate was the youngest/oldest) was balanced across gender and the signal of performance. Finally, in each pool, the candidates' majors were in the same field.[5] Appendix Table C1 contains a list of the resume pools generated from this first stage.

3.2. *Stage 2*

The second stage was conducted with college students at the Experimental Economics Laboratory at Loyola Marymount University (LMU) and the Economics Laboratory at University of California, San Diego (UCSD). The experiment was programmed and conducted using z-Tree software (Fischbacher, 2007). Sample instructions for the second stage are included in Appendix D.

In the second stage, 463 individuals participated as evaluators. Evaluators were told about the candidate's task. Evaluators were then given a pool of three resumes and asked to indicate their individual hiring preference. Evaluators were incentivized to hire the candidate that they thought performed best on the second round of the task. Evaluators received $10 if they hired the candidate who performed best in the second round of the task, $5 if they hired the candidate who performed second best, and $1 if they hired the candidate that performed the worst.[6]

Participants were then randomly assigned to a three-person group, referred to as a hiring committee. As a group, evaluators revisited the same three candidates and were asked to jointly select one candidate to hire. Each session was randomly assigned to a decision-making process (between-subjects design). Groups were told they would be making their decision by majority vote, unanimous

[3] As discussed in Bohnet et al (2016), providing a signal of performance when candidates are evaluated jointly will reduce the chance of observing discrimination.

[4] We include a dominated candidate to mitigate experimenter demand effects by obscuring the gender focus of the experiment. The gender of the dominated candidate varied across pools; in 7 of the 13 pools the dominated candidate was female. In pilot studies, we explicitly asked participants to guess the purpose of the experiment. No participant guessed that the study was designed to explore gender discrimination in hiring decisions, with most guessing that we were generally interested in studying other factors that affect hiring decisions. While we cannot fully rule out experimenter demand effects related to gender, the presence of social desirability bias, in which participants try to appear less discriminatory, would reduce our likelihood of observing discrimination.

[5] Fields of study are Arts and Humanities, Engineering, Natural Science, and Social Science. Majors are classified based on the following list: https://undergrad.usc.edu/programs/major/list/; https://web.archive.org/web/20210222030045/ (accessed in 2021).

[6] To ensure that participants thoughtfully considered the resumes, they were first asked to indicate their individual hiring preferences. A large literature examines differences in individual and group decisions (Dominguez, 2023). Studying the difference between participants' individual decisions and group decisions is outside the scope of this paper. This paper instead focuses on group decisions.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

vote, by a member of their group who was randomly selected to be the leader, or by a member of their group who volunteered to be the leader.[7]

Each group had two minutes to chat using a chat box. This mimics the real world in which groups typically communicate before making a hiring decision. While some works have shown that allowing participants to communicate minimizes differences in decision-making rules and uniformly improves efficiency (Goeree & Yariv, 2011), others (Mengel, 2021) show that this increases gender discrimination. In the leader treatments, the leader of the group was announced before the group began chatting.

The groups varied in gender composition, with groups having zero to three women. Subjects were not informed of the gender composition of their group. However, the name and self-selected pronouns of each group member was shown when participants interacted using the chat box.[8] After chatting, the group was asked to select the candidate to hire, with the same incentives as the individual hiring decision.

This was repeated for another five or seven periods; in total, evaluators either participated in six rounds (spring of 2022) or eight rounds (fall of 2022 and spring of 2023). In each period, groups were randomized again (within-subject randomization). After completing all periods, evaluators answered a short questionnaire. The questionnaire collected information on the evaluator's characteristics, including gender, age, race, year of college, and GPA range. Additionally, as part of the questionnaire evaluators self-reported personality traits including risk preference (whether they are generally a person who is fully prepared to task risk or avoid risk) and public perception (how much it matters what other people think about them). Finally, the questionnaire elicited evaluators' feedback about their group's dynamics in the last round – it asked if in the last round evaluators felt that their input was heard, whether their group openly expressed ideas and opinions, and whether they believed their team made thoughtful decisions that all team members supported. Evaluators were compensated for both their individual and group decision in one randomly selected period. In total, evaluator earnings varied between $7 and $25, which included the show up fee.

The order in which the resume pools were evaluated was randomly implemented across sessions. Evaluators were informed about their group's decision at the end of each period. However, the information on earnings was not displayed until the end of the session.

We note that this experimental design is similar to Dominguez (2023). In Dominguez (2023) participants viewed resumes with information on the candidates age, gender, field of study, and signal of past performance. They were asked to select two candidates out of six to perform a mathematical task. Participants first made their decisions individually. Participants were then allocated into groups of three, with varied gender composition. Groups were able to chat for three minutes before making

[7] In the majority vote treatment, groups were told that if they did not reach a majority decision, a random candidate would be hired. In the majority vote, all groups reached a majority decision.

Similarly, in the unanimous vote treatment, groups were told that if they did not reach a unanimous decision, a random candidate would be hired. In the unanimous vote, seven groups out of 272 (2.6 percent) did not reach a unanimous decision. (When a group failed to reach a unanimous decision, the random candidate was chosen from the three candidates in the pool.) Not surprisingly, given the small percentage of groups that were affected, our results are similar if these seven groups are excluded from our sample.

In the volunteer leader treatment, participants were given the names of other group members before being asked if they wanted to volunteer to be a leader. Groups were told that if no one volunteered to be a leader, one would be randomly selected. This occurred in 31 out of 284 groups (10.9 percent). Results are similar if these 31 groups are excluded from our sample. Additionally, groups were told that if more than one person volunteered to be a leader, one of the volunteers would be randomly selected. This occurred in 151 out of 284 groups (53.2 percent).

[8] The use of personal pronouns was standard practice at both LMU and UCSD at the time of our study. Thus, it is unlikely that asking participants to select their pronouns revealed our secondary research question.

Eight participants identified as non-binary, selecting the pronouns they/them/theirs. Groups with these eight participants have been dropped from our sample. However, our results are similar if they are included in the analysis.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Table 1 Number of observations

	LMU		UCSD		
	Male	Female	Male	Female	Total
Majority vote	142 (20)	146 (21)	218 (31)	262 (38)	768 (110)
Unanimous vote	124 (17)	236 (31)	180 (26)	276 (42)	816 (116)
Random leader	275 (37)	211 (28)	119 (22)	133 (24)	738 (111)
Volunteer leader	206 (28)	202 (29)	138 (21)	306 (48)	852 (126)
Total # of observations	747 (102)	795 (109)	655 (100)	977 (152)	3,174 (463)

Note: Table 1 shows the number of observations.
Number of participants is shown in parentheses. Each participant was placed on a three-person hiring committee, viewed three candidates per period, and participated in either six or eight periods. Therefore, the number of observations = number of participants ÷ 3 participants per group * 3 candidates * 6 or 8 periods.

a decision using a majority voting rule. This was repeated for three periods. Our experimental design builds on Dominguez (2023) by varying the decision-making process across sessions.

Table 1 provides information on the sample size by gender, treatment, and school. Table 1 shows that our sample is well balanced across these three dimensions. Table 1 also shows the number of participants in parentheses.

4. Results

4.1. Hiring decisions

To examine whether the decision-making process affects gender disparities in hiring we estimate the following specification:

$$\begin{aligned} Pr\left(Hire_{gcp}=1\right) &= \beta_0+\beta_1 Female_c+\beta_2 Unanimous_g+\beta_3 RandomLeader_g \\ &+\beta_4 VolunteerLeader_g+\ \beta_5 Female_c * Unanimous_g+\beta_6 Female_c \\ &* RandomLeader_g+\beta_7 Female_c * VolunteerLeader_g \\ &+\delta X_c+\alpha X_g+\gamma Period_p+\varepsilon_{gcp} \end{aligned} \tag{1}$$

where $Hire_{gcp}$ is a dummy variable that equals one if group g hires candidate c in period p. $Female_c$ is a dummy variable that equals one if candidate c is female. $Unanimous_g$, $RandomLeader_g$, and $VolunteerLeader_g$ are dummy variables denoting each of the decision-making treatments, with the majority vote treatment as the omitted group. X_c is a vector of candidate c's other characteristics (a dummy variable that equals one if the candidate is the youngest candidate in the pool, a dummy variable that equals one if the candidate is the highest scoring in the pool, the candidate's major, and the candidate's position on the screen) and the share of the candidates in the pool that are female.[9] X_g is

[9] Our results are robust to alternative measures of candidate c's characteristics. For example, our results are robust to including two categories of the candidate's relative ranking in the pool, such as dummy variables to denote the highest and median scoring candidate (relative to the lowest scoring) and dummy variables to denote the youngest and median aged candidate (relative to the oldest candidate).

Our data includes three observations per group as each group evaluates three candidates. Since each pool of resumes includes a 'dominated' candidate, whose purpose was to obscure the gender focus of the experiment, this may inflate sample size. We find that results are similar if we exclude the 'dominated' candidates from our analysis.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

a vector of group characteristics (namely the number of women in group g), and $Period_p$ is a vector of period fixed effects.

Our key coefficients are β_1, β_5, β_6, and β_7. These coefficients capture whether gender differences in hiring decisions vary across decision-making processes. β_1 represents the average difference in the probability that a female candidate is hired as compared to a male candidate in the majority rule treatment (after controlling for candidate characteristics, group characteristics, and period effects). β_5, β_6, and β_7 represent the average difference in the probability that a female candidate is hired as compared to a male candidate under the unanimous, randomly appointed leader, and volunteer leader treatments respectively (as compared to the majority rule treatment).

We note that this specification is consistent with Dominguez (2023) and Bohnet et al. (2016); Dominguez (2023) estimates a probit specification in which the dependent variable is a dummy variable that equals one if group g hires candidate c in period p. His analysis is also conducted at the candidate level, enabling the author to control for candidate characteristics, namely age, major, signal of performance, and position on the screen. Other controls include the share of female candidates in the pool and period fixed effects. The primary difference between the specification used in Dominguez (2023) and the specification used here are the interaction terms. Unlike Dominguez (2023), in this experiment groups were assigned one of four decision-making processes; thus, we interact the gender of the candidate with dummies for the decision-making processes. We also differ from Dominguez (2023) by using a linear probability model rather than a probit model; probit or logit specifications are not appropriate in this setting due to the interaction terms (Ai & Norton, 2003).

We begin by estimating equation (1) without treatment fixed effects or interaction terms. When treatment fixed effects and interaction terms are excluded, β_1 represents the average difference in the probability that a female candidate is hired as compared to a male candidate (after controlling for candidate characteristics, group characteristics, and period effects). Because our analysis controls for candidate characteristics, any remaining difference in the likelihood that a female candidate is hired as compared to a male candidate captures gender discrimination. A significant β_1 indicates that groups are using uninformative information (gender) to make hiring decisions, thus making their decisions inefficient. As seen in the first column of Table A1, the probability a female candidate is hired is 0.095 lower than men. This indicates that on average a female candidate is 9.5 percentage points less likely to be hired than a comparable male candidate. This difference is statistically significant at the 1 percent level. This coefficient is similar in size and significance to Dominguez (2023), whose experimental design is most comparable to ours.

Next, we estimate the linear probability model specified in equation (1). The first column of Appendix Table A1 reports select coefficients from regression with robust standard errors in parentheses.[10](For reference, these results are also in Table 2, discussed below.) Mainly, the table reports coefficients β_1, β_5, β_6, and β_7, as well as the implied level of gender disparities within each treatment, that is β_1, $\beta_1 + \beta_5, \beta_1 + \beta_6$, and $\beta_1 + \beta_7$. Fig. 1 utilizes these regression coefficients to display the predicted probability that a male is hired (in blue) and the predicted probability that a female is hired (in red). Fig. 1 also displays 95 percent confidence intervals. Since participants are asked to hire one candidate out of a pool of three, the average probability that a candidate is hired is one third in all

[10] In our preferred specification we include treatment fixed effects and use robust standard errors. As discussed in Kim (2020), clustering at the session level is not a remedy for session fixed effects. As the treatment effect is assigned at the session level an alternative could be clustering at the session level, as is done in Dominguez (2023). We obtain similar p-values if the standard errors are clustered at the session level, with clustered standard errors derived using wild bootstrap (Kline & Santos, 2012) due to the low number of clusters.

The experimental design could also justify clustering at the group level. As observations are more positively related at the session level as opposed to the group level, we believe this is less appropriate. However, we note that results are quantitatively similar if the errors are instead clustered at the group level.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Table 2 Heterogeneity over time

	(1)	(2)	(3)	(4)	(5)
Dummy = 1 if candidate is female	−0.090***	−0.084*	−0.076	−0.105***	−0.116***
	(0.031)	(0.047)	(0.080)	(0.034)	(0.034)
Dummy = 1 if candidate female * unanimous voting treatment	0.048*	0.048*	0.027	0.048*	0.070***
	(0.028)	(0.028)	(0.112)	(0.027)	(0.027)
Dummy = 1 if candidate female * random leader treatment	0.002	0.002	0.102	−0.007	0.009
	(0.030)	(0.030)	(0.113)	(0.030)	(0.028)
Dummy = 1 if candidate female * volunteer leader treatment	−0.068**	−0.068**	−0.118	−0.060**	−0.025
	(0.029)	(0.029)	(0.104)	(0.029)	(0.027)
Implied level of discrimination:					
ß_1: Majority voting treatment	−0.090***	−0.084*	−0.076	−0.105***	−0.116***
	(0.031)	(0.047)	(0.080)	(0.034)	(0.034)
$\text{ß}_1 + \text{ß}_5$: Unanimous voting reatment	−0.0418	−0.036	−0.049	−0.057*	−0.046
	(0.032)	(0.049)	(0.086)	(0.034)	(0.035)
$\text{ß}_1 + \text{ß}_6$: Random leader treatment	– 0.0878**	−0.082*	0.026	−0.111***	−0.107***
	(0.034)	(0.048)	(0.086)	(0.035)	(0.034)
$\text{ß}_1 + \text{ß}_7$: Volunteer leader treatment	– 0.158***	−0.152***	−0.195***	−0.165***	−0.141***
	(0.034)	(0.048)	(0.073)	(0.037)	(0.036)
Observations	3,174	3,174	3,174	2,727	2,727
Adjusted R-squared	0.611	0.610	0.609	0.661	0.664
Additional interaction terms	None	Female * Period	Female * Period * Treatment	Highest Scoring * Last Period Female Highest Scoring & Youngest * Last Period Female Youngest	Highest Scoring * Last Period Female Highest Scoring * Treatment & Youngest * Last Period Female Youngest * Treatment

Note: Table reports select regression coefficients from a linear probability regression. Controls include: candidate's characteristics (dummy variable = 1 if youngest, dummy variable = 1 if highest scoring, major, position on screen), share of female candidates in the pool, number of females in the group, and period fixed effects.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

treatments. Discrimination is measured as the difference between the predicted probability that a male candidate is hired and the predicted probability that a female candidate is hired.

The first panel in Fig. 1 depicts the level of discrimination when a majority voting rule is used. Under a majority voting rule, a female candidate is 9.0 percentage points less likely to be hired than a similar male candidate. This difference is statistically significant at the 1 percent level.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

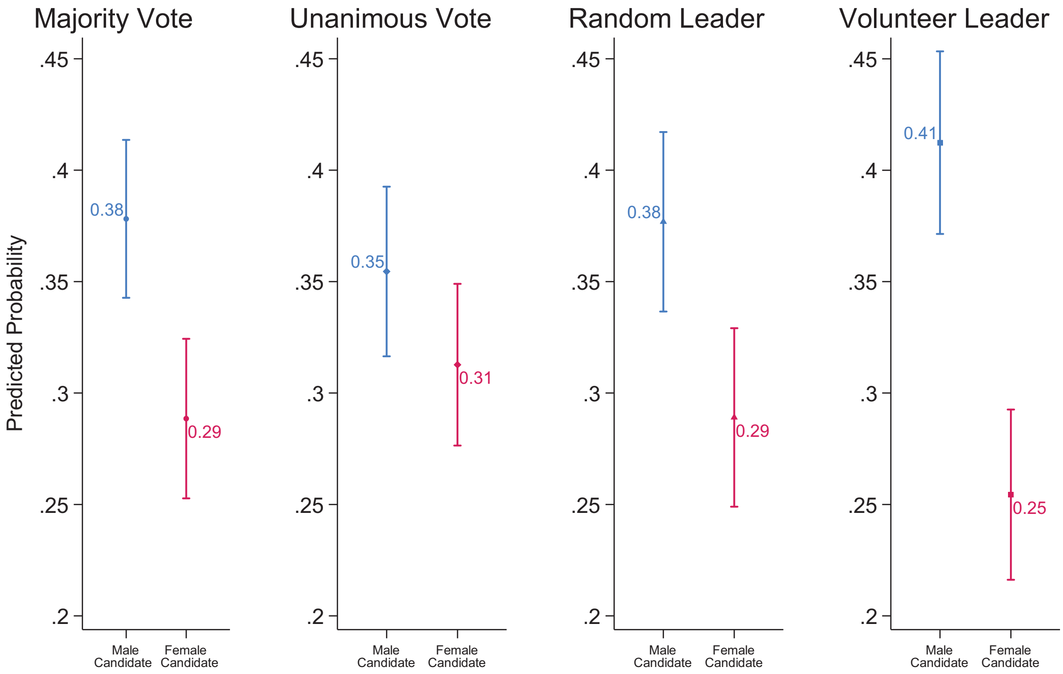

Fig. 1 Predicted probability a candidate is hired by gender and treatment
Note: Predicted probability (and 95 percent confidence interval) a candidate is hired by the candidate's gender and treatment. Predicted probabilities are based on our primary specification seen in column 3 of Table A1. All other covariates are held at their mean value.

The second panel in Fig. 1 shows that the unanimous voting rule closes the gender discrimination gap. The interaction term β_5 is positive and significant at the 10 percent level. As a result, when a unanimous voting rule is utilized, male and female candidates do not have a statistically significant difference in the predicted probability of being hired. This is consistent with the findings from Mak et al. (2019) and J. Chan et al. (2018), which found that a unanimous decision-making rule is more efficient than a majority decision-making rule.

The third panel of Fig. 1 examines the level of discrimination when a decision is made by a randomly appointed leader. In this decision-making process, discrimination is similar to that seen under the majority voting rule – males are 8.8 percentage points more likely to be hired than similar female candidates. This is consistent with Hastie and Kameda (2005) that discussed how, when a leader decides for a group, the implicit decision rule is usually a majority rule.

The final panel of Fig. 1 depicts the level of discrimination when the hiring decision is made by a leader who volunteers. The interaction term β_7 is negative and statistically significant at the 5 percent level. Thus, this treatment exacerbates discrimination; a male candidate is 15.8 percentage points more likely to be hired than a comparable female candidate. This is nearly a two-fold increase in the level of discrimination. This is consistent with the literature referenced above that details how leadership willingness may be trait-dependent, and the literature that details how these traits can correlate with the leader's style and decisions (Ertac & Gurdal, 2019; Kocher et al., 2013). For example, the type of person who volunteers to be a leader may be less responsive to others' preferences, or less democratic. In this setting, this leads to additional discrimination.

As detailed in Table A1, our results are robust to the inclusion of session-fixed effects, field of study fixed effects, and interactions between candidate characteristics and treatment effects.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

4.2. Heterogeneity by gender

In this section we first consider heterogeneity by the group's gender composition to explore whether the decision-making process has a differential impact on hiring decisions that depends on the number of women in the group. Then, for our treatments in which a decision is made by a leader, we consider heterogeneity in our results by the leader's gender.

4.2.1. Heterogeneity by group composition

To examine heterogeneity by the group's gender composition, we interact our dummy variable that equals one if the candidate is female with a continuous variable that denotes the number of females in the group. As before, other controls include the candidate's characteristics (dummy variable that equals one if the candidate is female, dummy variable that equals one if the candidate is the youngest, dummy variable that equals one if the candidate is the highest scoring, major, position on the screen), the share of women in the pool of candidates, group characteristics (the number of females in the group), and period fixed effects. Select coefficients are reported in Table A2 with robust standard errors in parentheses.

When using the full sample, the coefficient on the interaction term is 0.026; this coefficient is significant at the 5 percent level. This suggests that on average, the gender composition of the group impacts the level of discrimination. Specifically, each additional woman in the group is associated with a 2.6 percentage point increase in the likelihood that the female candidate is hired. This is consistent with some of the literature discussed in Section 2, including De Paola and Scoppa (2015), which found that hiring committees with women are more likely to hire female candidates than all-male committees.

Next, we perform the same analysis separately for each of our decision-making treatments. While this analysis allows for comparisons across treatments, we interpret the coefficients with caution given the limited sample size. Given our sample size, the baseline likelihood that a female candidate is hired across all treatments, and power of 0.9, we can only detect effect sizes bigger than 0.8. Coefficients from these regressions are used to calculate the predicted probability that a candidate is hired by the group's gender composition and decision-making process – these calculated probabilities are displayed in Fig. 2. Fig. 2 displays the predicted probability that a male is hired (in blue) and the predicted probability that a female is hired (in red) by the number of women in the group for each treatment. The figure also displays 95 percent confidence intervals.

Fig. 2 shows that under three of the four decision-making processes, increasing the number of women in a group does not have a statistically significant impact on the predicted difference in the probability that a male versus female candidate is hired.[11] Meaning, in all decision-making processes except the unanimous vote, gender discrimination is relatively consistent regardless of the group's gender composition. However, when the unanimous voting rule is applied, increasing the share of women in a group significantly increases the likelihood that the hired candidate is female. Thus, Fig. 2 does not find compelling evidence of heterogeneity by the group's gender composition for decision rules other than the unanimous vote.

4.2.2. Heterogeneity by leader's gender

For our two treatments in which a decision is made by a leader, we consider heterogeneity in our results by the leader's gender. We consider this heterogeneity because the leader's gender may influence their leadership style and/or their decisions about which candidate to hire. For example,

[11] Our insignificant results could be caused by non-linear effects (Dominguez, 2023). In results not shown, we consider such non-linearity by interacting the dummy variable that equals one if the candidate is female with three discrete measures: a dummy variable if there are one, two, or three females in the group. This is done separately for each treatment. Results are quantitatively similar to those seen in Table A2 and Fig. 2. Thus, it is unlikely that our results are driven by non-linear effects.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

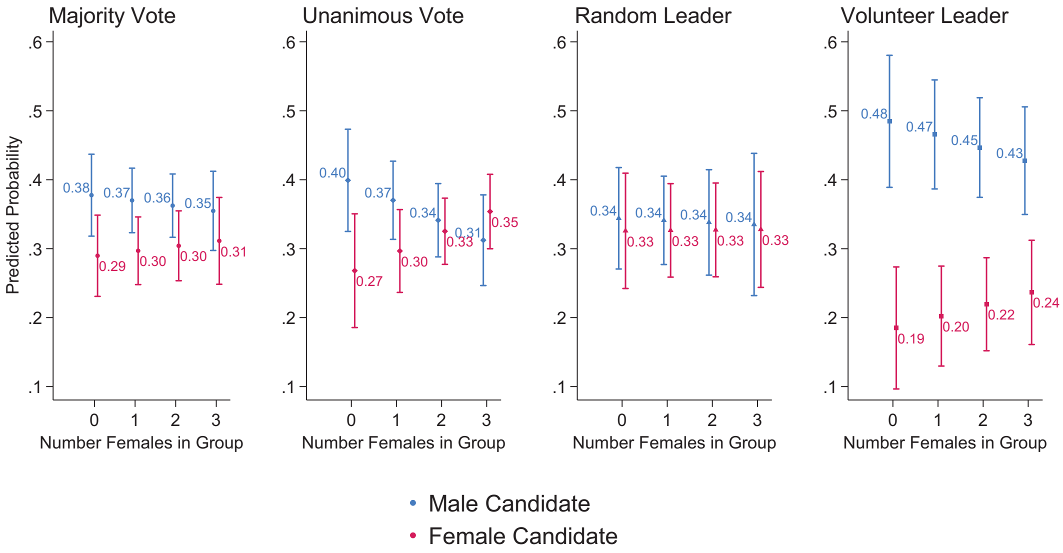

Fig. 2 Predicted probability a candidate is hired by group composition and treatment
Note: Predicted probability (and 95 percent confidence interval) a candidate is hired by the group's gender composition and the treatment. Predicted probabilities are based on coefficients from Table A2, columns 2–5. All other covariates are held at their mean value.

female leaders may be more responsive to others' preferences or more democratic than male leaders (Ertac & Gurdal, 2019); this may influence the level of discrimination in hiring decisions.

To examine heterogeneity by the leader's gender we interact our dummy variable that equals one if the candidate is female with a dummy variable that equals one if the leader is female. Other controls include a dummy variable that equals one if the leader is a female, the candidate's characteristics, the share of women in the pool of candidates, group characteristics (the number of females in the group), and period fixed effects. Select coefficients are reported in Table A3 with robust standard errors in parentheses. Again, we may be underpowered to detect statistically significant effects with this analysis. With our sample size in the leader treatments, the baseline likelihood that a female candidate is hired in the leader treatments, and power of 0.9, we are able to detect effect sizes bigger than 0.09.

When using data from both leader treatments, the coefficient on the interaction term is not significant, indicating that on average the leader's gender does not impact the level of discrimination. This is in line with previous works such as Williams and Ceci (2015), which did not find evidence that male and female decision-makers differ in hiring preferences.

Next, we perform the same analysis separately for each of our leader treatments. These coefficients are used to calculate the predicted probability that a candidate is hired by the leader's gender and decision-making process – these calculated probabilities are displayed in Fig. 3. Fig. 3 also displays 95 percent confidence intervals.

Consistent with Fig. 1, Fig. 3 illustrates that discrimination is intensified when a hiring decision is made by a leader who volunteers as compared to a leader who is randomly selected. When a decision is made by a randomly selected leader, the difference in the predicted probability that a male and female candidate is hired remains statistically similar regardless of the leader's gender. Specifically, the figure shows that regardless of the leader's gender, a female candidate is two percentage points less likely to be hired compared to a similar male candidate.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

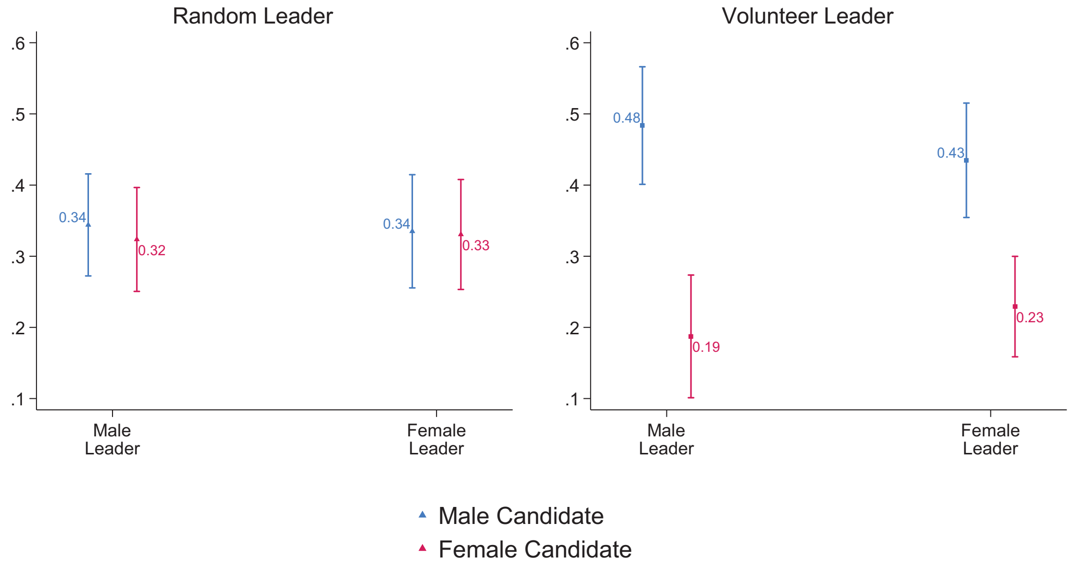

Fig. 3 Predicted probability a candidate is hired by leader's gender and treatment
Note: Predicted probability (and 95 percent confidence interval) a candidate is hired by the leader's gender and treatment. Predicted probabilities are based on coefficients from Table A3, columns 2–3. All other covariates are held at their mean value.

When a decision is made by a volunteer leader, the figure shows that the hiring decisions of male and female leaders differ; female leaders are more likely to hire female candidates. Specifically, we find that male leaders are 29 percentage points less likely to hire a female candidate and female leaders are 20 percentage points less likely to hire a female candidate. This 9 percentage point difference is significant at the 5 percent level. Thus, gender differences in self-selection into leadership cannot fully explain our finding that gender differences are exacerbated when decisions are made by a leader who volunteers as compared to decisions made by a randomly selected leader.

4.3. Heterogeneity over time

As participants evaluated pools of candidates over multiple rounds, in this section we consider heterogeneity in the level of discrimination over time. To explore heterogeneity over time, we include an additional interaction term – we interact our dummy variable that equals one if the candidate is female with a vector of period fixed effects. Regression results are shown in Table 2 with robust standard errors in parentheses. (To assist readers, results from our primary specification are shown in column 1.) As seen in the second column of Table 2, when the additional interaction terms are included, the coefficients of interest and the implied levels of discrimination remain unchanged. Furthermore, the additional interaction terms are all statistically insignificant and the adjusted R^2 falls. Together, this suggests that, across treatments, the level of discrimination remains consistent over time.

However, there could be heterogeneity in discrimination over time within treatments. To consider this possibility, in the third column we include a triple interaction term, interacting the dummy variable that equals one if the candidate is female with a vector of treatment fixed effects and a vector of period fixed effects. With these additional interaction terms, the implied level of discrimination remains similar in size. Furthermore, none of the interaction terms are statistically significant and the adjusted R^2 falls again. The results from the second and third column indicate that our results are not driven by dynamics over time and within treatment.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Finally, we consider whether there are sequential spillovers in discrimination (Kessler et al., 2024). For example, we examine whether groups discriminate against the highest scoring candidate if the female candidate was the highest scoring candidate in the previous period. Similarly, we consider whether groups discriminate against the youngest performing candidate if the female was the youngest candidate last period. To explore this mechanism, we include two additional interaction terms: (1) we interact our dummy variable that equals one if the candidate is the highest scoring with a dummy variable that equals one if the female candidate was the highest scoring in the prior period and (2) we interact our dummy variable that equals one if the candidate is the youngest candidate with a dummy variable that equals one if the female candidate was the youngest in the prior period. For this analysis, our sample size is reduced as we must exclude group decisions made in the first period. As seen in the fourth column of Table 2, when we include these additional terms, the primary coefficients and the implied level of discrimination remain similar. In the final column, we examine whether the sequential spillovers differ by treatment – we interact the terms described above with treatment fixed effects. The coefficients of interest and implied level of discrimination remain unchanged, suggesting that our findings are not caused by sequential spillovers.

4.4. Mechanisms

To explore the mechanisms by which decision-making processes affect group decisions, we analyze the group chats as well as the data collected in the post-experimental survey.[12]

The group chats differed in quantity, content, and tone. For example, consider the initial message exchanged by a group: approximately one half of the initial messages were proposals (for example, 'I think we should hire person B'), while the other half of the initial messages were trivial (not related to the task, for example: 'Hi,' 'How's it going,' or 'Who do you think we should hire?'). The average group exchanged 13 messages with 3 messages being trivial. Of the non-trivial messages, on average six messages agreed with the original proposal (for example, 'ok that works for me' or 'I also think person A is good because of the education') and four messages countered the original proposal (for example, 'I say A then B' after someone else said 'I think B is better'). We note that all participants sent at least one message during the chat period. In total, the average group spent 80 seconds chatting. In this section we analyze the chat data to examine whether the various decision-making processes affect the chat dynamics. To explore this effect, the unit of analysis changes; our analysis is performed at the group-period level. Since our dataset includes one observation per group per period, we have less power to detect statistically significant effects in this analysis.

We also analyze data from the post-experimental survey, which solicited participants feedback on how their group worked in the final round. Because the post-experimental survey solicited participants feedback about the last round completed, the sample is restricted to the final period.

Select regression coefficients from the chat data analysis can be found in Table 3. In all specifications, our independent variables include our treatment dummy variables (with the majority vote

[12] To minimize measurement error, three research assistants who were unaware of the purpose of the study, independently coded the chat data. These research assistants had several tasks. First, they classified all messages into four categories: proposal, agreement, counter, and trivial. A message is considered a proposal if the group member is putting forward a candidate's name for the first time or if a group member provides justification for a candidate they already proposed. A message is considered an agreement if a different group member is explicitly agreeing with and/or providing a justification for the candidate(s) first proposed in the conversation. A message is considered a counter if a group member is suggesting or proposing a different candidate than the candidate first proposed in the conversation. A message is considered trivial if the message is not relevant to the study. For our analysis, a message is classified in one of these categories if two of the three research assistants coded it as such. However, our results are robust to alternative definitions such as the average of the three counts. Second, the research assistants counted the number of candidates discussed by a group. For our analysis, we created a dummy variable that equals one if two of the three research assistants indicated that more than one candidate was discussed. Again, our results are robust to alternative definitions. Finally, the research assistants counted the number of characteristics the group discussed. Our dependent variable is the median of the counts from the three research assistants. As before, our results are robust to alternative definitions such as the average of the three counts.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

decision-making process serving as the omitted group), resume pool characteristics (a dummy variable that equals one if the female candidate is the youngest, a dummy variable that equals one if the female candidate is the highest scoring, a vector of dummy variables denoting the candidates' field of study, and the share of female candidates in the pool), the number of females in the group, and period fixed effects.

In the first three columns, we examine how the quantity of discussion varies by decision-making process. To measure the quantity of discussions, we examine the number of messages and the number of characters.

Columns 4–10 examine how the decision-making process impacts the content of the conversations. To capture the content of the chat messages, we consider six variables. First, we measure the substantive quantity of the discussions, calculated as the number of non-trivial messages. Second, we create a dummy variable that equals one if the group discussed more than one candidate. Additionally, we count the number of the candidates' characteristics that the group discussed. (This variable ranges from zero to four as resumes included four characteristics: the candidate's age, gender, college major, and a signal of performance.) Next, we examine how inclusive the discussions are. This is measured as whether one group member dominated the discussion (dummy variable that equals one if one group member entered more than 50 percent of the characters into the chat box) and whether one group member is relatively quiet (dummy variable that equals one if one group member enters less than 10 percent of the characters into the chat box). Finally we measure the sentiment of the discussions. This measure is based on the AFINN model (Nielsen, 2011). (This variable is the average sentiment score of all the messages exchanged by a group. Score values range between − 5 and 5 with higher values corresponding to words that generate more positive sentiments. Positive scores occur if evaluators use words with positive tone, agree with one another, or discuss positive aspects of the candidates.)

In the last four columns we analyze the post-experimental survey data to examine whether perceptions of a group's dynamics differ across treatments. First, we measure the percentage of the group that felt their input was heard, based on the percent of participants that responded yes to the question, '*Do you feel that your input was heard in the group decision-making?*.' Second, we take the average score of responses to the question, '*Do you believe that your team members openly expressed their ideas and opinions?*.' Scores range 1–4 with higher scores indicating that the respondent more strongly agreed with the statement. Third, we similarly take the average score of individual responses to the question, '*Do you believe that your team was able to make thoughtful decisions that all team members supported?*.' Again, scores range 1–4 with higher scores indicating that the respondent more strongly agreed with the statement.

With so many outcomes being tested, readers may be concerned about multiple hypothesis testing. To ensure that our results are not due to chance, we implement two standard approaches. First, we create three summary index measures to reduce the number of tests; one index measures the quantity of discussions (Column 3), another the content of discussions (Column 10), and the third measures the perceptions of discussions (Column 14). Each index pools several outcomes into a single measure thereby providing a statistical test for whether a treatment has a general effect on a set of outcomes. Following Anderson (2008), when creating the index, we define outcomes so that a higher number corresponds to a 'better' outcome. The index is then calculated as the mean of standardized outcomes weighted by the inverse of their correlation matrix. Additionally, our analysis adjusts the p-values to correct for multiple comparisons. We use False Discovery Rate (FDR) adjusted q-values that adjust p-values by dividing the significance level by the number of hypotheses tested, taking into account the rank of the variable according to its p-value within the index (Anderson, 2008). Table 3 reports the unadjusted p-values in brackets and the adjusted p-values in braces. Stars denoting significance levels are based on the adjusted p-values.

As detailed in Section 2, because a unanimous vote requires agreement from all team members, prior works have maintained that this decision-making process will elicit additional discussions compared to the majority voting rule. Our results support this premise – the first column of Table 3 shows

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Table 3 Mechanisms

	Quantity of discussions			Content of discussions							Perceptions of discussions			
	Number of messages	Number of characters	Quantity index	Number of non-trivial messages	Dummy = 1 if more than one candidate discussed	Number of CV characteristics discussed	Dummy = 1 if one participant dominates discussion	Dummy = 1 if one participant is quiet	Sentiment score	Content index	Percent of group that felt their input was heard	Encouraged to express different points of view	Thoughtful decision that all members supported	Perceptions index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment fixed effects														
Unanimous vote	1.563**	17.427	0.189	1.785***	0.125**	0.214	0.091	0.020	0.068	0.094	0.025	0.053	0.005	0.127
	[0.005]	[0.151]	[0.025]	[0.000]	[0.002]	[0.027]	[0.037]	[0.599]	[0.043]	[0.036]	[0.530]	[0.717]	[0.962]	[0.563]
	{0.045}	{0.279}	{0.114}	{0.002}	{0.024}	{0.114}	{0.119}	{0.814}	{0.129}	{0.118}	{0.796}	{0.838}	{0.962}	{0.814}
Random leader	−1.645**	−5.332	−0.152	0.189	0.089	0.18	0.107	0.068	0.037	0.015	0.004	0.068	−0.008	0.070
	[0.001]	[0.661]	[0.057]	[0.645]	[0.036]	[0.064]	[0.016]	[0.088]	[0.309]	[0.734]	[0.922]	[0.601]	[0.950]	[0.738]
	{0.019}	{0.817}	{0.151}	{0.817}	{0.118}	{0.154}	{0.112}	{0.177}	{0.522}	{0.838}	{0.962}	{0.814}	{0.962}	{0.838}
Volunteer leader	0.907	20.443	0.147	1.374**	0.09	0.222	0.080	0.018	0.004	0.044	−0.014	0.196	0.136	0.201
	[0.071]	[0.073]	[0.056]	[0.000]	[0.026]	[0.018]	[0.066]	[0.626]	[0.905]	[0.332]	[0.762]	[0.153]	[0.311]	[0.420]
	{0.154}	{0.154}	{0.151}	{0.015}	{0.114}	{0.111}	{0.154}	{0.817}	{0.962}	{0.536}	{0.842}	{0.279}	{0.522}	{0.654}
Observations	1058	1058	1058	1058	1058	1058	1058	1058	1058	1058	149	149	149	149
Mean in majority vote	12.42	205.12	−0.05	8.96	0.34	1.03	0.46	0.25	0.53	−0.04	0.94	3.41	3.69	−0.10
Adjusted R2	0.03	0.03	0.03	0.05	0.06	0.03	0.01	0.01	0.01	0.04	−0.05	0.02	−0.03	−0.03

Note: Table reports select regression coefficients from a linear regressin model. Additional controls include CV characteristics (a dummy variable that equals one if the female candidate is the highest scoring, a dummy variable that equals one if the female candidate is the youngest, the share of female candidates in the pool, and a vector of dummy variables denoting the candidates' field of study), the number of females in the group, and period fixed effects.

In column 3, the dependent variable is a score which standardizes the two measures of quantity in columns 1 and 2. Similarly the content index in column 10 is a score which standardizes the six measures in columns 4 to 9 while the perceptions index in column 14 is a score that standardizes the three measures in columns 11 to 13. All indexes are created such that a positive coefficient reflects 'better' outcomes, with each measure being weighted by the inverse of the correlation matrix.

Unadjusted p-values are in brackets and FDR adjusted p-values are in braces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ based on FDR adjusted p-values.

that groups exchange 1.6 more messages under a unanimous voting rule than a majority voting rule; this difference is statistically significant at the 5 percent level. Given that groups exchanged 12.4 messages on average when making decisions by majority vote, this coefficient corresponds to a 13 percent increase. In addition to exchanging more messages, participants exchanged 17 more characters (an eight percent increase) in the unanimous voting treatment than the majority voting treatment, though the difference is not statistically significant. Our index measure suggests that the unanimous voting role leads to additional discussions (compared to the majority voting rule), although again the difference is not statistically significant.

Moreover, prior works have maintained that compared to the majority voting rule, the unanimous voting rule will increase the quality/content of discussions. The fourth column of Table 3 shows that when decisions are made by a unanimous voting rule, groups exchange 1.8 additional messages with non-trivial content, which corresponds to a 20 percent increase compared to a majority vote. This difference is significant at the 1 percent level. When decisions are made by a unanimous voting rule, groups are also more likely to discuss more than one candidate. The probability that they discuss more than one candidate increases by 0.13, which represents a 37 percent increase. This difference is significant at the 5 percent level. Additionally, these groups discuss 0.21 more characteristics from the resumes, a 21 percent increase. However, this difference is not statistically significant. Qualitatively these discussions are less inclusive. One participant is 9.1 percentage points more likely to dominate the discussion, an increase of 20 percent. Additionally, one participant is 2 percentage points more likely to be relatively quiet (an 8 percent increase). Under the unanimous voting rule, groups interact more positively, with an average sentiment score that is 0.07 points higher, which represents a 13 percent increase. This is not surprising as sentiment scores may capture how inclusive the conversation is, whether evaluators agree with each other, or whether evaluators focus on positive aspects of the candidates being evaluated. Our index measure also suggests that the content of the discussion is more robust compared to the majority voting rule.

Previous works suggest that on average, perceived satisfaction will be higher under a unanimous voting rule than under a majority voting rule because group members are more willing to share their perspectives and constructively work through differences. Qualitatively, our results in the final four columns support this assertion. Compared to a majority voting rule, under a unanimous voting rule, group members are more likely to feel that their input was heard, feel encouraged to express different points of view, and believe their group made a thoughtful decision that all group members supported. The index measure also shows that groups have a higher perceived satisfaction under the unanimous voting rule compared to the majority voting rule. However, none of these differences are statistically significant. Taken together, our results from Fig. 1 and Table 3 align with prior research that found the unanimous voting rule leads to additional and more thorough discussions as well as more efficient (i.e., less discriminatory) decisions than the majority voting rule.

By comparison, as discussed above, when a decision is made by a leader, the leader's preference is often viewed as a default decision that can only be overturned by other clear improvements. As a result, group members will only voice their opinions if they are convinced that their preferred candidate is better, and they are willing to challenge the leader. Thus, we would expect to find fewer but higher quality discussions (as compared to a majority vote). Our chat data largely supports these predictions, particularly when decisions are made by a randomly appointed leader. As seen in the first column of Table 3, when a decision is made by a randomly appointed leader, 1.6 fewer messages are exchanged, representing a 13 percent decrease. This difference is significant at the 5 percent level. Surprisingly, groups in which the decision is made by a volunteer leader exchanges 0.9 more messages (a 7 percent increase) compared to groups in which decisions are made by a majority vote. However, this difference is not statistically significant. The effects on the number of characters and the quantity index follow a similar pattern, albeit insignificantly.

Additionally, the second section of Table 3 indicates that the content of the discussion improves in both leader treatments. Qualitatively, groups with a leader making the decision exchanged more

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

non-trivial messages; groups with a random leader exchange 0.19 (2 percent) more non-trivial messages and groups with a volunteer leader exchange 1.4 (15 percent) more non-trivial messages than in the majority vote treatment. This difference is insignificant for groups with random leaders and significant for groups with volunteer leaders. Groups with a leader have an increased likelihood of discussing more than one candidate; the probability of discussing more than one candidate increases by 0.09 (26 percent) although the difference is not statistically significant. Additionally, groups with a leader discuss 0.2 (19 percent) more characteristics from the resumes although the difference is not statistically significant. However, conversations are less inclusive. We find an increase in the probability that one participant dominates the conversation in both leader treatments, as does the probability that one participant is relatively quiet. This is consistent with prior research that suggests that there is less diversity in ideas when a leader makes the decision. Prior works have also suggested that a leader can have a divergent outcome on a group's sentiment (Ertac & Gurdal, 2019; Kocher et al., 2013); the direction of the effect depends on how the leader interacts with their group. In total, the index measure indicates that groups have higher quality discussions when decisions are made by a leader as compared to the majority vote (although the difference is not statistically significant).

In the last four columns we find no statistically significant effects of having a leader make the decision on the satisfaction reported in the post-experimental surveys. Qualitatively, groups with a random leader express higher satisfaction levels based on whether they felt their input was heard and whether they were encouraged to express different points of view. Similarly, groups that made decisions using a volunteer leader express higher satisfaction based on whether they were encouraged to express different points of view and whether their group made thoughtful decisions. The index measure indicates that under both leader treatments, group members' perceptions are higher than when decisions are made using a majority voting rule. But again these differences are not statistically significant.

To summarize, the results from the random leader treatment align with prior research. Results from Table 3 show that there are fewer but better discussions when decisions are made by a randomly appointed leader. As seen in Fig. 1, this leads to a decision that is equally efficient (i.e., equally discriminatory) as in the majority vote. When decisions are made by a volunteer leader, there are more discussions, and the discussions are of a higher quality. But as seen in Fig. 1, this comes with decisions that are *less* efficient (i.e., more discriminatory) than in the majority vote. The fact that conversations look somewhat comparable across leader treatments yet result in dramatically different hiring outcomes suggests that there may be further differences across these leader treatments that are not captured by the results presented. These findings align with prior research, which suggests that whether a leader improves decision quality depends on the leadership style exhibited by the leader as well as self-selection into leadership, that is, who is the leader. Thus, in the next section we further examine the leader treatments.

4.5. Why does the leader matter?

In this section we explore the mechanisms by which leaders affect group decisions. First, we examine selection into leadership, that is, who volunteers to be a leader. Then, we examine whether leaders differ in their leadership style by considering how they influence the group's conversation and how they influence the group's decision.

To examine selection into leadership, Table 4 presents select coefficients from a linear probability model in which the dependent variable is a dummy variable that equals one if a participant volunteers to be a leader in a given period. This analysis is performed at the participant-period level with one observation per participant per period. As the analysis is restricted to the volunteer leader treatment, data is only available from 126 participants. Robust standard errors are reported in parentheses.

In column 1, we include participant characteristics obtained from the post-experiment survey as well as period fixed effects. These characteristics include the participant's gender, age, and

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Table 4 Who volunteers to be a leader?

	(1)	(2)	(3)
Dummy = 1 if participant is female	0.074**	0.110**	0.111**
	(0.037)	(0.047)	(0.047)
Risk preference	0.033***	0.031***	0.031***
	(0.011)	(0.011)	(0.011)
What other people think matters	−0.026	−0.026	−0.025
	(0.021)	(0.021)	(0.021)
Number of females in the group		−0.031	−0.030
		(0.024)	(0.024)
Observations	852	852	852
Adjusted R2	0.0609	0.0618	0.0616
Additional participant characteristics	Yes	Yes	Yes
Resume pool characteristics	No	No	Yes

Note: Table reports select coefficients from a linear probability regression with 126 participants. All regressions include controls for additional participant characteristics (age, dummy variables for: race, year of college, and GPA range) and period fixed effects. In addition some regressions include resume pool characteristics (share of female candidates in the pool, dummy variable if the female candidate is the highest scoring, dummy variable if the female candidate is the youngest, and dummy variables for field of study).
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

dummy variables for: race, year of college, and GPA range, as well as self-reported personality traits. Participants self-report their risk preference, with a score of zero indicating that they are unwilling to take risks and a score of 10 indicating that they are fully prepared to take risks. They also report the importance of public perception, with higher scores indicating that they care more about what other people think (scores range between one and five).

In column 2 we also control for the number of females in the group since participants know their group members before they volunteer to be a leader. Consistent with our finding that the gender composition of the group does not affect discrimination, we find that the number of females in the group does not significantly affect whether the participant volunteers to be a leader.[13]

In column 3 we control for resume pool characteristics: the share of female candidates in the pool, a dummy variable if the female candidate is the highest scoring, a dummy variable if the female candidate is the youngest, and dummy variables for the field of study. Our results are robust to the inclusion of these variables. This suggests that participants do not differentially volunteer to be leaders based on the resume pool characteristics. For example, participants do not differentially volunteer when the female candidate is the highest performing or when the female candidate is the youngest.[14]

Across all three specifications, we find that controlling for demographic characteristics, women are *more* likely to volunteer to be a leader. Specifically, in our preferred specification (column 3), women are 11.1 percentage points more likely to volunteer to be a leader; this difference is statistically

[13] Motivated by Born et al.'s (2022) finding that the gender composition of the group affects women's willingness to lead, we explore the interaction between participants' gender and the number of females in the group. In results not shown, we find no significant interactions between the participants' gender and the number of females in the group.

We also interact risk aversion with the number of females in the group and with a dummy variable that equals one if the participant is female to examine whether women are more likely to volunteer if they are risk averse or there are more females in their group. These interactions terms are not statistically significant.

[14] In results not shown, we interact these resume pool characteristics with a dummy variable that equals one if the participant is female. All interaction terms are insignificant. Additionally, the inclusion of these interaction terms decreases the adjusted R^2. These findings indicate that female participants do not differentially volunteer to be the leader, depending on the resume pool characteristics. For example, female participants do not differentially volunteer when the female candidate is the highest performing or the youngest.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

significant at the 5 percent level. Indeed, in our sample, men volunteered to be the leader 50 percent of the time and women volunteered to be the leader 54 percent of the time. This finding differs from Ertac and Gurdal (2012), which found that women are less willing to make a risky decision for a group than men.

Additionally, in all three specifications, we find that participants who are more willing to take risks are more willing to volunteer to be a leader; this difference is significant at the 1 percent level.[15] This is consistent with Brenner (2015), who found that senior managers in the United States are less risk averse than non-senior managers. It is also consistent with K. Y. Chan et al. (2015) and Ertac and Gurdal (2012), which showed that individuals who want to be leaders are less likely to be risk averse. Our result is consistent with the hypothesis that risk-averse individuals are less likely to want to take on the burden or risk of making decisions on behalf of others (Ertac & Gurdal, 2012). This coefficient may also be positive because risk tolerance is correlated with Big-5 personality traits (Judge et al., 2002); Judge et al. (2002) showed that Big-5 personality traits are correlated with wanting to be a leader.

In the third row, we examine whether participants' willingness to be leader is related to the importance of public perception. Across all three specifications, we find that whether participants' volunteer to lead is insignificantly related to how much they care about what other people think about them.

To summarize, we find selection into leadership, showing that less risk-averse and female participants are more likely to volunteer to be a leader in our setting. Because participants with these characteristics may interact with their groups differently, and because volunteer leaders may generally interact differently with their groups as compared to randomly appointed leaders, we next examine the impact that leaders have on the group's discussion and the group's decision.

In Table 5, we examine a leader's impact on both the group's discussion (columns 1–4) and the group's hiring decisions (columns 5–6). The analysis is performed at the group-period level. As the analysis is restricted to the leader treatments, data is only available for 530 groups. Specifically, for this analysis, we consider whether a volunteer leader has a differential impact than a randomly assigned leader. In all specifications, our independent variables are our treatment dummy variable (a dummy variable that equals one in the volunteer leader treatment and zero in the randomly assigned leader treatment), resume pool characteristics (the share of female candidates in the pool, a dummy variable that equals one if the female candidate is the youngest, a dummy variable that equals one if the female candidate is the highest scoring, and a vector of dummy variables denoting the candidates' field of study), and period fixed effects. Additionally, some specifications include the leader's characteristics (a dummy variable that equals one if the leader is female, the leader's risk preference, how much the leader cares about what other people think, age, and dummies for race, year of college, and GPA category). Table 5 shows select coefficients from our regression. As before, robust standard errors are shown in parentheses.

The first two columns examine whether the leader is the first to speak in the group conversation. The leader speaks first in 42 percent of the conversations. There are no statistically significant differences across the leader treatments, with a randomly selected and volunteer leader equally likely to speak first in the group conversation. However, we note that qualitatively, the volunteer leader is more likely to speak first. Additionally, qualitatively, we find that after we control for the leader's characteristics, this likelihood decreases.

Next, we examine whether the leader dominates the conversation. Prior research suggests that there is less diversity in ideas when a leader makes the decision – indeed, Table 3 shows that in both leader treatments, there was a higher probability that one participant dominates the conversation. In

[15] Female participants have higher risk scores than male participants. The average risk score for female participants is 6.68 while the average risk score for male participants is 5.81, with the difference being significant at the 1 percent level. However, it is unlikely that our results are biased by multicollinearity as the variance inflation factor is only 3.3.

Additionally, female and male participants have divergent college majors. However, controlling for college major does not alter our findings.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Table 5 Leader's impact

	Dummy = 1 if leader talked first		Dummy = 1 if leader dominates conversation		Dummy = 1 if leader does not change their vote	
	(1)	(2)	(3)	(4)	(5)	(6)
Volunteer leader	0.058 (0.043)	0.024 (0.048)	0.070* (0.039)	0.049 (0.044)	0.022 (0.036)	−0.008 (0.041)
Select leader characteristics						
Dummy = 1 if female = 1		−0.016 (0.048)		0.024 (0.042)		−0.082** (0.039)
Risk preference		−0.006 (0.016)		0.001 (0.014)		0.005 (0.012)
What other people think matters		−0.012 (0.028)		0.010 (0.025)		−0.027 (0.022)
Observations	530	530	530	530	530	530
Mean	0.417	0.417	0.277	0.277	0.781	0.781
Adjusted R2	−0.00112	−0.0159	0.0138	0.0232	0.0194	0.0448
CV characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Additional leader characteristics	No	Yes	No	Yes	No	Yes

Note: Table reports select coefficients from a linear probability regression.
All regressions include controls for resume pool characteristics (share of female candidates in the pool, dummy variable if the female candidate is the highest scoring, dummy variable if the female candidate is the youngest, dummy variables for field of study), and period fixed effects. Additionally, some regressions include controls for leader characteristics (age, dummy variables for: race, year of college, GPA range).
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

columns 3 and 4 we specifically examine whether it is the leader that is dominating the conversation. In these columns, the dependent variable is a dummy variable that equals one if the leader entered more than 50 percent of the characters into the chat box. We find that the leader only dominates the conversation in 28 percent of the groups. Column 3 shows that the leader is 7 percentage points more likely to dominate if they volunteered; this difference is marginally significant at the 10 percent level. This difference is smaller and not significant once we control for the leader's characteristics.

Finally, we examine whether the leader typically enacts their own preferences or changes their decision after the group's conversation. In columns 5 and 6 the dependent variable is a dummy variable that equals one if the leader's individual vote (elicited before the group conversation and selection of a leader) is equal to the leader's vote submitted after the group discussion, on behalf of the group. Seventy-eight percent of leaders' decisions correspond to their own initial preferences. To provide context, we note that in the majority voting rule, 61 percent of participants submit the same vote before and after the group discussion. Similarly, in the unanimous voting rule, 54 percent of participants submit the same vote before and after the group discussion. (Appendix Table E1 provides detailed statistics on the number of participants who change their vote.) Thus, leaders are more likely to follow their own preferences than individuals are in a majority or unanimous vote. Columns 5 and 6 show that randomly selected and volunteer leaders are similarly likely to enact their own preferences.

To summarize, we find that leaders who are randomly appointed are similar to volunteer leaders in whether they speak first, dominate the conversation, and whether they ultimately hire the candidate that aligns with their own preferences. This suggests that the results are unlikely to be driven by the source of the leader's authority. Additionally, it suggests that it is unlikely that the leader's source of authority impacts how the group members perceive the leader. Instead, differences across leadership treatments are likely driven by self-selection – indeed, we note that across all regressions, the coefficient of interest declines when leader characteristics are included in our specification.

6. Conclusion

Women remain underrepresented in many fields, especially those that are traditionally male dominated. And while the statistics are well known, the solution remains uncertain. Firms partake in education and training programs, governments have instituted quotas, and yet gender discrimination persists. Using a laboratory experiment, we show the important role the decision-making process plays in discrimination in hiring. In a context in which groups communicate via chat, we find that gender discrimination is eliminated when hiring decisions are made unanimously by a group and that gender discrimination is largest when decisions are made by a leader who volunteers. We do not find heterogeneity by the group's gender composition or by the leader's gender. Thus, we provide further evidence that quotas that increase the representation of women in decision-making bodies are not sufficient to eliminate discrimination.

Instead, our results suggest that firms should focus on how hiring committees make decisions, indicating that the structure of the decision-making process itself may meaningfully influence disparities in hiring outcomes. Because such procedural changes are relatively simple to implement and may yield more immediate effects, they represent a promising avenue for institutional reform. While our experimental setting cannot fully capture potential drawbacks – such as logistical challenges in applying decision rules or differences between online and face-to-face deliberations – the results underscore the need for greater scholarly and policy attention to the role of group decision-making processes in shaping hiring decisions.

How hiring committees make decisions is a black box, seldom studied or discussed in the literature or in practice. While our paper shows that *how* hiring committees make decisions has a significant impact on gender diversity in hiring, the mechanisms for *why* remain unclear. Our experimental design allows us to rule out several explanations, including differences in leader characteristics and communication styles. Our findings suggest that self-selection into leadership roles may exacerbate gender disparities. Future research should investigate the underlying channels through which decision-making processes shape hiring outcomes with particular attention to self-selection into leadership roles.

Our findings have extensions beyond hiring. Groups utilize each of these four decision-making processes in a range of different real-world settings. Our findings suggest that committees need to carefully consider the decision-making process they employ.

Replication Packages. To obtain replication material for this article, https://doi.org/10.17605/OSF.IO/BHMWC.

Funding statement. The authors received funding from Loyola Marymount University for this research.

Competing interests. The authors declare that there are no conflicts of interest related to the work submitted for publication.

Ethical Approval. This study was reviewed and determined to be exempt by Loyola Marymount University's Institutional Review Board (2019 SP 72-R).

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economic Letters*, *80*(1), 123–129.

Alan, S., Ertac, S., Kubilay, E. & Loranth, G. (2020). Understanding gender differences in leadership. *The Economic Journal*, *130*(626), 263–289.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and early training projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495.

Arbak, E., & Villeval, M.-C. (2013). Voluntary leadership: Motivation and influence. *Social Choice & Welfare*, *40*(3), 635–662.

Austen-Smith, D., & Feddersen, T. (2005). Deliberation and voting rules. In D. Austen-Smith & J. Duggan (Eds.), *Social choice and strategic decisions: Essays in honor of Jeffrey S. Banks* (pp. 269–316). Springer.

Bagues, M., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, *77*(4), 1301–1328.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). Does the gender composition of scientific committees matter?. *American Economic Review*, *107*(4), 1207–1238.

Batista Pereira, F. (2023). Electoral rules and voter bias against female candidates in Brazilian congressional elections. *Journal of Elections, Public Opinion and Parties*, *33*(1), 22–40.

Bohnet, I. (2016). *What works: Gender equality by design*. Harvard University Press.

Bohnet, I., van Green, A., & Bazerman, M. (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science*, *62*(5), 1225–1234.

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

Breitmoser, Y., & Valasek, J. (2017). A rationale for unanimity in committees. *WZB Discussion Paper Number SP11 2017-308*.

Brenner, S. (2015). The risk preferences of US executives. *Management Science*, *61*(6), 1344–1361.

Bureau of Labor Statistics, U.S. Department of Labor (2024). *The Economics Daily*, Labor force participation rate for women highest in the District of Columbia in 2022 at https://www.bls.gov/opub/ted/2023/labor-force-participation-rate-for-women-highest-in-the-district-of-columbia-in-2022.htm (visited May 09, 2024).

Carlsson, M., & Eriksson, S. (2019). In-group gender bias in hiring: Real-world evidence. *Economic Letters*, *185*, 108686.

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

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

Chan, J., Lizzeri, A., Suen, W., & Yariv, L. (2018). Deliberating collective decisions. *Review of Economic Studies*, *85*(2), 929–963.

Chan, K. Y., Uy, M. A., Chernyshenko, O. S., Ho, M. H. R., & Sam, Y. L. (2015). Personality and entrepreneurial, professional and leadership motivations. *Personality and Individual Differences*, *77*, 161–166.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, *99*(1), 431–457.

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

Cornell, B., & Welch, I. (1996). Culture, information, and screening discrimination. *Journal of Political Economy*, *104*(3), 542–571.

Daskalova, V. (2020). In-Group Favoritism in Collective Decisions. *Working Paper*.

De Paola, M., & Scoppa, V. (2015). Gender discrimination and evaluators' gender: Evidence from Italian academia. *Economica*, *82*(325), 162–188.

Dessein, W. (2007). Why a group needs a leader: Decision making and debate in committees. *CEPR Discussion Paper No. DP6168*.

Dominguez, J. J. (2023). Diversified committees in hiring processes: Lab evidence on group dynamics. *Journal of Economics Psychology*, *97*, 1–15.

Ertac, S., & Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, *83*(1), 24–30.

Ertac, S., & Gurdal, M. Y. (2019). Preference Communication and Leadership in Group Decision Making. *Journal of Behavioral and Experimental Economics*, *80*, 130–140.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Goeree, J. K., & Yariv, L. (2011). An experimental study of collective deliberation. *Econometrica*, *79*(3), 893–921.

Guarnaschelli, S., McKelvey, R. D., & Palfrey, T. R. (2000). An experimental study of jury decision rules. *The American Political Science Review*, *94*(2), 407–423.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*(2), 494–508.

Henningsen, D. D., Henningsen, M. L. M., Jakobsen, L., & Borton, I. (2004). It's good to be leader: The influence of randomly selected and systematically selected leaders on decision-making groups. *Group Dynamics: Theory, Research, and Practice*, *8*(1), 62–76.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604.

Hinchliffe, E. (2023). Women CEOs run 10.4% of Fortune 500 companies. A quarter of the 52 leaders became CEO in the last year. *Fortune 500*. June 5, 2023.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*(4), 589–617.

Karpowitz, C. F., Mendelberg, T., & Shaker, L. (2012). Gender inequality in deliberation participation. *American Political Science Review*, *106*(3), 533–547.

Kessler, J. B., Low, C., & Shan, X. (2024). Lowering the playing field: Discrimination through sequential spillover effects. *The Review of Economics and Statistics*, 1–28.

Kim, D. G. (2020). Clustering standard errors at the "session" level. *CESifo Working Paper No 8386*.

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press

Kline, P., & Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, *1*(1), 23–41.

Kocher, M. G., Pogrebna, G., & Sutter, M. (2013). Other-regarding preferences and management styles. *Journal of Economic Behavior & Organization*, *88*, 109–132.

Leppert, R., & DeSilver, D. (2023). 118th Congress has a record number of women. *PEW Research Center*. January 3, 2023.

Lorenz, J., Rauhut, H., & Kittel, B. (2015). Majoritarian democracy undermines truth-finding in deliberative committees. *Research & Politics*, *2*(2), 1–10.

Mak, V., Seale, D. A., Rapoport, A., & Gisches, E. J. (2019). Voting rules in sequential search by committee: Theory and experiments. *Management Science*, *65*(9), 4349–4364.

Martinez, A., & Christnacht, C. (2021). Women are nearly half of U.S. workforce but only 27% of STEM workers. *U.S. Census Bureau: Stats for Stories*. January 26, 2021.

Mengel, F. (2021). Gender bias in opinion aggregation. *International Economic Review*, *62*(3), 1055–1080.

Miller, C. E. (1985). Group decision making under majority and unanimous decision rules. *Social Psychology Quarterly*, *48*(1), 51–61.

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

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*

Palmer, K. (2023). Workplace climate pushing female professors out. *Inside Higher Ed*. October 24, 2023.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, *83*(1), 111–117.

Schippers, M. C., & Rus, D. C. (2021). Majority decision-making works best under conditions of leadership ambiguity and shared task representations. *Frontiers in Psychology*, *12*, 1–16.

Stasser, G., & Abele, S. (2003). Group creativity and collective choice. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation Through Collaboration* (pp. 85–109). Oxford University Press.

Stiglitz, J. E. (1973). Approaches to the economics of discrimination. *The American Economic Review*, *63*(2), 287–295.

Williams, J. (2021). We need real metrics not heartfelt conversations, to tackle workplace diversity. *Fortune 500*, December 7, 2021.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences*, *112*(17), 5360–5365.

Cite this article: Jain, P., & Miller, M. M. (2025). Beyond representation: the importance of the decision-making process in hiring decisions. *Experimental Economics*, 1–25. https://doi.org/10.1017/eec.2025.10024

https://doi.org/10.1017/eec.2025.10024 Published online by Cambridge University Press