



ROCKWOOL Foundation Berlin

Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES

282/26

Committee Deliberation and Gender Differences in Influence

Jonas Radbruch, Amelie Schiprowski

Committee Deliberation and Gender Differences in Influence

Authors

Jonas Radbruch, Amelie Schiprowski

Reference

JEL Codes: D71, J16, M51

Keywords: Committee Decision-Making, Gender Quotas, Hiring

Recommended Citation: Jonas Radbruch, Amelie Schiprowski (2026): Committee Deliberation and Gender Differences in Influence. RFBerlin Discussion Paper No. 282/26

Access

Papers can be downloaded free of charge from the RFBerlin website: <https://www.rfberlin.com/discussion-papers>

Discussion Papers of RFBerlin are indexed on RePEc: <https://ideas.repec.org/s/crm/wpaper.html>

Disclaimer

Opinions and views expressed in this paper are those of the author(s) and not those of RFBerlin. Research disseminated in this discussion paper series may include views on policy, but RFBerlin takes no institutional policy positions. RFBerlin is an independent research institute.

RFBerlin Discussion Papers often represent preliminary or incomplete work and have not been peer-reviewed. Citation and use of research disseminated in this series should take into account the provisional nature of the work. Discussion papers are shared to encourage feedback and foster academic discussion.

All materials were provided by the authors, who are responsible for proper attribution and rights clearance. While every effort has been made to ensure proper attribution and accuracy, should any issues arise regarding authorship, citation, or rights, please contact RFBerlin to request a correction.

These materials may not be used for the development or training of artificial intelligence systems.

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com



Committee Deliberation and Gender Differences in Influence

Jonas Radbruch[†] Amelie Schiprowski[§]

October 16, 2026

Abstract

This paper provides empirical evidence on the aggregation of information in hiring committees. We use data from the hiring process of a large company in which committee members first conduct independent one-on-one interviews and then deliberate to reach a collective hiring decision. Results show that the mapping of committee members' initial assessments to final decisions is not deterministic, leaving scope for deliberation and individual influence. Committee decisions are systematically less aligned with recommendations made by women than with those made by men, despite similar qualifications, experience, and evaluation behavior. This gender difference in influence is strongest when disagreement is high and when women are in the minority. Further results suggest that it has implications for the gender composition of new hires.

Keywords: Committee Decision-Making, Gender Quotas, Hiring

JEL Codes: D71, J16, M51

[†] Humboldt University Berlin, Spandauer Straße 1, D-10178 Berlin. Email: jonas.radbruch@hu-berlin.de

[§] University of Bonn, Adenauerallee 24-42, D-53113 Bonn. Email: amelie.schiprowski@uni-bonn.de

We thank several seminar and workshop audiences, as well as Stefano DellaVigna, Pierre Deschamps, Simon Jäger, Paul Heidhues, Stephan Lauermann, Friederike Reichel and Nina Roussille, for helpful comments and suggestions. Marcel Wachter and Zsigmond Szajbely provided excellent research assistance. Support by the German Research Foundation (DFG) through EXC 2126/2–390838866, CRC TR 224 (Project A05, Schiprowski) and CRC TR 190 (Project A06, Radbruch) is gratefully acknowledged. No third party had the right to pre-review this paper, except for the purpose of verifying the absence of confidential information on the hiring process of the anonymous company.

1 Introduction

Many important decisions are taken by committees or boards, which need to aggregate the individual information and preferences of their members. A growing number of policies and initiatives target the representation of specific groups in this process. One prominent example is the gender quota, which has been explicitly or implicitly introduced on various boards and committees (see, e.g., Azmat & Boring, 2020). Gender quotas often aim at incorporating the perspectives of women into important organizational decisions. In hiring committees, for example, a common conjecture is that women may evaluate job candidates differently than men and thereby influence the composition of new hires.

While quotas mechanically increase numerical representation, their effects on outcomes depend on women’s influence in the decision-making process. An important bottleneck is the deliberation stage, during which committees aggregate potentially divergent information and preferences. A large experimental literature documents systematic behavioral differences between women and men that may be relevant in this context. For example, it has been shown that women are less inclined to promote their ideas or assume leadership in group discussions (e.g., Carli, 1989; Coffman, 2014; Born et al., 2022; Hardt et al., 2025). In committee deliberations, such behavioral differences may translate into gender differences in influence over decision outcomes.

In this paper, we use data on private-sector hiring committees to examine how committees aggregate individual candidate assessments into collective hiring decisions. The data stem from the German-speaking branch of a large consulting company that selects employees for entry-level permanent positions and internships. The company’s hiring process is organized through interview days held at various locations throughout the year, with varying numbers of candidates and evaluators. Every candidate is first independently interviewed and assessed by two or three distinct evaluators from the committee. After all one-on-one interviews have been completed, evaluators reveal and discuss their individual hiring recommendations in a committee meeting. They then reach a joint decision through deliberation, without a fixed cutoff or voting rule. This structure allows us to measure committee members’ influence by exam-

ining how individual pre-deliberation assessments align with collective post-deliberation decisions. The data include about 8,300 recommendations, resulting in about 3,000 committee decisions. We exploit the institutional setting to address three main research questions. First, how do hiring committees aggregate individual recommendations? Second, are recommendations made by women less likely to be reflected in committee decisions than those made by men? Third, do differences in influence affect the gender composition of new hires?

We first describe how men and women on the committee compare in terms of their observable characteristics, candidate assignments, and evaluation behavior. Men and women have similar observable characteristics, including experience and seniority. Women are more likely to evaluate female candidates because the HR department aims to assign each female candidate to at least one female evaluator. Because this guideline is non-binding and the availability of female evaluators varies across interview days, candidates of both genders are evaluated by committees ranging from all-male to female-majority. Conditional on a candidate's gender and interview day, observable candidate characteristics are balanced across evaluator gender. While recommendations made by women are slightly more favorable on average, they are as informative about predetermined candidate quality as those made by men. Moreover, male and female evaluators exhibit similar levels of disagreement with their co-evaluators.

We then address our first research question concerning the aggregation of individual recommendations into hiring decisions. As expected, committees are more likely to make job offers when candidates receive more positive recommendations. However, these decisions cannot be summarized by a deterministic cutoff or majority-voting rule, implying that initial recommendations are strong but imperfect predictors of hiring decisions. This is particularly true in cases involving disagreement among evaluators, which we refer to as 'deliberation cases'.

Motivated by the scope for deliberation, we next examine whether recommendations made by women and men are equally likely to be reflected in committee decisions. A first descriptive analysis suggests systematic differences: positive (negative) recommendations made by women are less (more) likely to result in a job offer than comparable recommendations made by men. To analyze these differences more formally, we construct a binary indicator measuring whether an initial recommendation aligns with the committee's final decision. We then esti-

mate how the probability of alignment varies between men and women. When considering all hiring decisions, the raw gender gap in alignment amounts to about 8%. Roughly 55% of this gap persists after controlling for deliberation conditions, including the level of disagreement among committee members. This residual gap corresponds to a two standard deviation (≈ 3 years) difference in hiring experience. It is entirely driven by deliberation cases, where the initial recommendations leave scope for heterogeneity in influence. Within this subsample, the residual alignment gap increases to about 13%. More generally, it widens with the degree of initial disagreement, reaching up to 60% in cases with the highest possible level of disagreement. Finally, we find suggestive evidence that the gap is largest when candidates are discussed by two men and one woman. By contrast, no gap emerges when women are in the majority.

Several placebo and robustness checks support the interpretation that the observed gender differences arise during deliberation rather than from differences in evaluation behavior. In particular, the gender gap in influence disappears under hypothetical cutoff rules that mechanically aggregate committee members' initial assessments into hiring decisions.

We discuss potential mechanisms underlying the results. We find no evidence that women are less influential because they produce less informative evaluations. Moreover, the gender gap is not a simple artifact of women's minority status, as it does not flip when men are in the minority. Instead, the findings point to behavioral differences during deliberation as a plausible mechanism. According to previous experimental findings, women and men differ systematically in how they communicate their information and in how their communication is perceived by others. While the exact dimensions of behavior leading to differences in influence are difficult to distinguish in observational data, they have in common that they are unrelated to the informational content of assessments. The results therefore suggest that behavioral factors beyond information quality shape the aggregation of information into collective decisions.

In a final step, we examine potential implications for the effect of gender quotas on hiring decisions. A common rationale for increasing female representation on hiring committees is to improve the hiring outcomes of female candidates. This also applies to our setting, where female candidates at the interview stage are on average 17% less likely to receive an offer than male candidates, despite similar pre-determined characteristics and ability measures.

Whether the presence of women on committees can alter the gender balance of new hires depends both on committee members' individual evaluation behavior and on the aggregation of these evaluations at the committee level. Our data allow for a separate analysis of these two levels. At the individual level, women evaluate female candidates significantly more favorably than their male colleagues. The recommendation gap that favors male candidates is about 50% smaller among female than among male committee members. At the same time, positive recommendations by women are significantly less likely to translate into an offer decision by the committee. At the committee level, the presence of a 'lone' woman does not affect the gender composition of new hires. This changes markedly when women become the majority, in which case the gender gap in offer rates falls to zero. Overall, the analysis suggests that deliberative dynamics play a critical role in determining how committee diversification affects gender equality in hiring outcomes.

The results in this paper document a systematic gender gap in the influence of individuals' recommendations on committee decisions with high stakes. They contribute to a growing literature on gender differences arising in the context of group decision-making. The vast majority of the existing evidence comes from laboratory experiments, showing that women are more likely to be overruled by others (Guo & Recalde, 2023), and less likely to become group leaders (e.g., Alan et al., 2019; Coffman et al., 2021; Born et al., 2022) or contribute ideas (e.g., Coffman, 2014; Hardt et al., 2025).¹ These differences have been found to be especially pronounced in male-dominated decision environments. Based on a field experiment with accounting students, Karpowitz et al. (2024) provide detailed evidence that lone women have a lower influence on their team compared with women in majority-women teams.² Overall, the existing literature has primarily focused on identifying behavioral differences between men

¹More generally, there exists broad evidence on gender differences in other dimensions of behavior, which are indirectly related to committee deliberation. For example, it has been shown that women exhibit a lower degree of competitiveness (e.g., Gneezy et al., 2003), take and receive less credit for group work (e.g., Sarsons et al., 2021; Koffi, 2025), are more likely to accept unpopular tasks (e.g., Babcock et al., 2017), ask for lower wages (e.g., Roussille, 2024) and engage less in negotiation (Bowles et al., 2007) and self-promotion (e.g., Exley & Kessler, 2022) than men.

²Additional descriptive evidence shows that women are less likely than men to interrupt others and more likely to be interrupted in supreme court examinations (Jacobi & Schweers, 2017), ask fewer questions in academic seminars (Carter et al., 2018), and speak up less frequently in a classroom setting (Burzryn et al., 2017).

and women in experimental group settings. Our key contribution is to document the relevance of these differences for high-stakes decision processes in the labor market.

Second, we provide new evidence on barriers to the effectiveness of gender quotas. Prior results have shown that individual female managers improve the employment and wage outcomes of women (e.g., Kunze & Miller, 2017; Adams-Prassl et al., 2024), whereas the presence of female board or committee members does not generate comparable effects (e.g., Bagues & Esteve-Volart, 2010; Bertrand et al., 2018; Williams & Ceci, 2015; Bagues et al., 2017; Maida & Weber, 2022; Deschamps, 2024). Our findings advance this literature by opening the ‘black box’ of how committees aggregate pre-deliberation views into post-deliberation decisions.³ We show that women are, on average, more favorable toward female candidates at the individual level, but that this advantage is often offset at the committee level due to gender differences in influence during deliberation. Moreover, we find that a shift from minority to majority status is associated with a substantial increase in the gender balance of new hires.⁴ These results raise new questions for the design of committee decision rules. Since it is often not feasible to increase the share of women to a critical mass, organizations that want to ensure an equal voice of all committee members may need to create procedural solutions or target behavioral changes in the deliberation process. This relates to experimental evidence by Karpowitz and Mendelberg (2014), showing that the relative speaking time of women versus men depends on the subsequent voting procedure.

More broadly, this paper adds to the literature on information aggregation in committees. A mostly theoretical strand of this literature analyzes the strategic behavior of committee members and the optimal design of voting rules (see Hao & Suen, 2009, for an overview). Particularly related to our setting are models of committee decision-making with pre-voting communication (e.g., Gerardi & Yariv, 2007; Coughlan, 2000; Austen-Smith & Feddersen., 2006), and informal election models without ex-ante commitment to a voting rule (e.g., Morgan &

³While Bagues et al. (2017) have shown that women on scientific promotion committees do not vote differently than men *after* deliberation has taken place, pre-deliberation views of committee members have not been observable in previous studies.

⁴Consistent with this result, Brito et al. (2024) show that women participate more actively in male-dominated political decision-making bodies as the share of women increases, and Ronchi and Salvestrini (2026) find that female-only judicial courts reach systematically different decisions than gender-mixed courts.

Stocken, 2008). Empirical evidence on committee decision-making is scarce. Goeree and Yariv (2011) and Burdea and Woon (2025) examine how voting with and without communication performs in a laboratory setting, Chan (2021) studies the role of seniority for information aggregation in physician teams, and Iaryczower et al. (2018) structurally estimate the effectiveness of deliberation in appellate courts.⁵ We contribute new empirical insights into how committees aggregate information in the absence of formal voting procedures. Our results show that the aggregation outcomes of committees are shaped not only by their institutional rules but also by the individual influence of their members.

The remainder of this paper is structured as follows. Section 2 provides information on the institutional setting. Section 3 describes the data source. Section 4 documents the aggregation of information in hiring committees. In Section 5, we estimate the gender gap in influence on committee decisions, investigate its heterogeneity, and discuss potential mechanisms. Section 6 analyzes implications for the gender balance of new hires. Section 7 concludes.

2 Institutional Setting

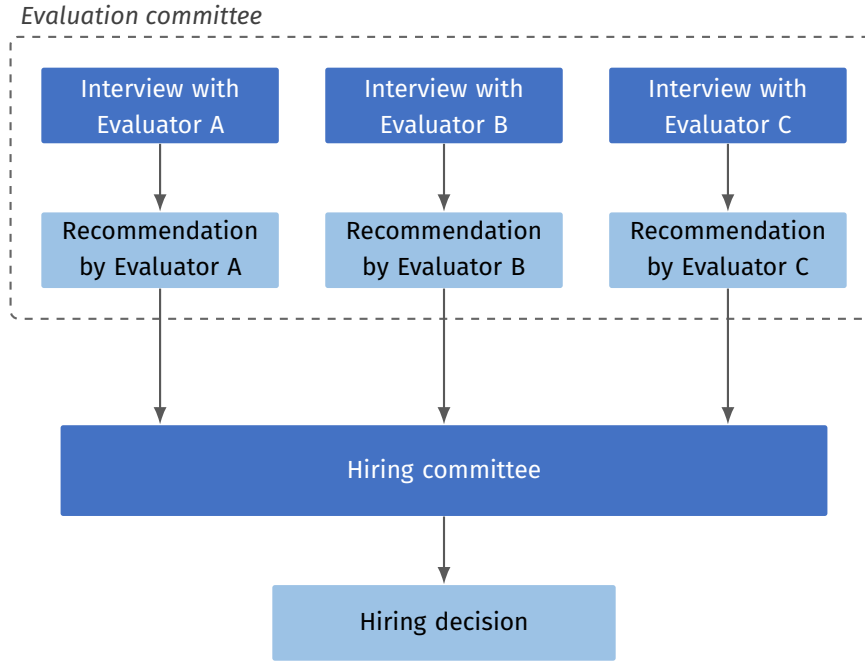
Background Our dataset covers the interview process of a large consulting company that hires candidates for paid internships (33%) and entry-level permanent positions (67%) at its German-speaking branch. Starting salaries for permanent positions fall within the top 10% of the German wage distribution. The process is organized through interview days held at different locations, with varying numbers of candidates and evaluators.⁶ Evaluators do not face a hiring quota at the interview-day level, as the company hires on a rolling basis.

Prior to the interview stage, candidates have been pre-screened and selected by the HR department based on their written applications. Candidates for internships usually hold an undergraduate degree, while candidates for permanent positions usually hold a Master’s degree. Evaluators on the hiring committee are consultants at the company and have undergone pro-

⁵More generally, a few empirical studies analyze how individual recommendations affect the decisions of third parties, such as journal editors (Card et al., 2020; Card & DellaVigna, 2020) or admission committees (Bai et al., 2022).

⁶The number of candidates (evaluators) per interview day ranges from 2 to 12 (3 to 13).

Figure 1: Illustration of Hiring Decision Process



Note: This figure illustrates the process through which hiring committees decide on the outcome of a given candidate. Each candidate is independently interviewed by three evaluators; together they constitute the candidate's *evaluation committee* (dashed frame). The *hiring committee* comprises all evaluators who interviewed at least one candidate on the respective interview day, as well as an HR representative and the local head of the interview team. It deliberates and reaches a hiring decision on all evaluated candidates. Evaluation committees are subsets of the hiring committee.

fessional interviewer training. The committee composition varies across interview days. Although evaluators are assigned to a primary interview location, their participation on a given day depends on their availability and the required size of the hiring committee. Moreover, evaluators commonly serve on hiring committees at other locations. While interview days are usually dedicated to either internships or permanent positions, the interview structure and assessment criteria are the same for both types of positions.

Figure 1 illustrates the structure of the committee's decision process. In the first stage, three evaluators conduct independent one-on-one interviews with each candidate, each resulting in an individual hiring recommendation formed prior to any discussion with the other evaluators. We refer to these three evaluators as the candidate's *evaluation committee*. In the second stage, all evaluators present on the interview day convene to discuss all candidates and reach

a joint hiring decision. We refer to this larger group as the *hiring committee*. It comprises all evaluators who conducted interviews that day, as well as an HR representative and the chair of the committee. Evaluation committees are therefore subsets of the hiring committee. In the following, we provide a more detailed description of the interview and deliberation stages of the hiring process.

Interview Stage During the interview stage, each candidate undergoes two or three one-on-one interviews with different evaluators.⁷ Evaluators typically interview and assess three candidates per day.

The HR department centrally assigns candidates to the evaluators participating on a given interview day. Because the interview process is highly standardized and the company recruits for generalist consulting positions, candidates and evaluators are not matched based on their fields of expertise. The only systematic assignment rule is that, whenever feasible given the pool of evaluators available on a particular interview day, each female candidate is assigned to at least one female evaluator.⁸ Conditional on this requirement, the remaining interviews are assigned at random, without regard to candidate or evaluator characteristics. This procedure is intended to ensure ex-ante fairness in candidate assignment.

The main outcome of each interview is a recommendation on whether to hire the candidate, expressed on a three-point scale (1 = rejection, 2 = potential offer, 3 = offer). This recommendation is derived from sub-ratings of the candidate's cognitive and non-cognitive skills, assessed through structured interview procedures developed by the company. Each interview evaluates a number of cognitive and a number of non-cognitive skills. The cognitive skill ratings are based on the candidate's performance in solving a business case and focus on mathematical and analytical skills, whereas the non-cognitive ratings capture dimensions related to achievements, motivation, leadership, and teamwork. Each interview assesses the same cognitive skills but a different subset of non-cognitive skills. The sub-ratings mechanically

⁷The regular case in our data is three interviews per candidate. Toward the end of our sample period, the number of interviews per candidate was reduced to two, affecting 12% of the dataset.

⁸This practice does not involve encouraging greater female participation in the interview process. Consistent with this, women do not conduct more interviews than men, as shown in Appendix Figure A.1.

translate into a recommendation according to a fixed rule that assigns equal weight to all sub-ratings and leaves no discretion to evaluators. Evaluators are instructed to form their assessments independently, without discussion or consultation with other committee members.

Deliberation Stage After all evaluators have entered their recommendations and sub-ratings into the computer system, the hiring committee convenes to deliberate and decide on the outcomes of all candidates based on the individual recommendations. The HR representative determines the order in which evaluators reveal and explain their recommendations. As a general rule, positive recommendations are presented first. There is no deterministic mapping from recommendations to committee decisions. Because recommendations are not unanimous in the vast majority (80%) of cases, there is substantial scope for deliberation to affect the committee's decision. The discussion of a given candidate primarily involves the evaluators who conducted the corresponding interviews (i.e., the evaluation committee), although other committee members may also participate. The deliberation process results in a decision on the candidate's hiring outcome without a formal voting procedure. The committee's decision is usually final. Candidates for permanent positions may be referred to a final interview with a senior manager, who is asked to validate the decision in cases of doubt.⁹

The offer probability amounts to about 30% among candidates for permanent positions and about 40% among candidates for internships. The hiring committee is not involved in determining the terms of the offer. More than 95% of offers made by the company are accepted.

3 Data

This section first describes the data source and the main variables used in the analysis (Section 3.1). It then provides evidence on the assignment of candidates to male and female committee members (Section 3.2), and on committee members' evaluation behavior (Section 3.3).

⁹About 15% of candidates in the data are referred to a final interview. We observe and use the final hiring decision. Our results are robust to excluding candidates with final interviews from the analysis (see Table D.6).

Table 1: Summary Statistics: Outcomes of the Interview Process

	All CM		Male CM			Female CM			Difference	
	Mean	SD	Mean	SD	N	Mean	SD	N	Diff.	p-value
Recommendation: Reject	0.42	0.49	0.44	0.50	5722	0.39	0.49	2600	-0.046	0.02
Recommendation: Potential Offer	0.27	0.44	0.27	0.44	5722	0.28	0.45	2600	0.014	0.36
Recommendation: Offer	0.31	0.46	0.30	0.46	5722	0.33	0.47	2600	0.033	0.07
Avg. Rating of Cognitive Skills	1.88	0.47	1.87	0.48	5722	1.90	0.45	2600	0.027	0.17
Avg. Rating of Non-Cognitive Skills	2.01	0.41	2.00	0.41	5722	2.04	0.40	2600	0.048	0.01
Offer	0.33	0.47	0.34	0.47	5722	0.32	0.47	2600	-0.016	0.28

Note: CM = committee member. The unit of observation is the interview (i.e., the committee member–candidate pair). p-values are based on standard errors clustered at the committee member level. The data include 359 distinct committee members (127 women and 232 men).

3.1 Data Source and Description of Main Variables

The data cover all interviews conducted between 2017 and 2021 at the German-speaking branch of the company.¹⁰ Our final dataset comprises 8,322 interviews across 444 interview days, involving 2,985 candidates and 359 committee members.¹¹ The data contain interview-level assessment outcomes (recommendations and skill ratings), the committee’s final hiring decision, as well as different candidate and committee member characteristics.

In the following, we describe the main variables used in our analysis. We distinguish between interview and hiring outcomes (3.1.1), committee member characteristics (3.1.2), and candidate characteristics (3.1.3).

3.1.1 Interview Assessments and Hiring Outcomes

Table 1 presents summary statistics for interview and hiring outcomes at the level of the interview (i.e., the committee member–candidate pair). Overall, 42% of interviews result in a reject recommendation, 27% in a potential-offer recommendation, and 31% in an offer recommendation. Female committee members are 4.6 percentage points less likely to recommend

¹⁰Interviews were conducted in person before the outbreak of the Covid-19 pandemic and moved online at the beginning of 2020, remaining in that format until the end of the observation period.

¹¹We excluded 726 interviews with missing information on recommendations, committee member gender, or candidate gender. Moreover, we dropped 574 interviews in which candidates did not complete the scheduled number of interviews, indicating attrition during the interview process. Finally, we excluded interview days that were specifically designed to evaluate only female candidates (270 interviews).

rejecting a candidate and correspondingly more likely to issue a positive recommendation. Appendix Figure B.1 shows that the recommendation behavior of men and women does not vary across evaluation committees with different gender compositions.

As described in Section 2, recommendations summarize cognitive and non-cognitive skill ratings measured on a three-point scale. The average cognitive and non-cognitive ratings are 1.9 and 2.0, respectively, with female committee members assigning slightly higher ratings on average. Appendix Figure B.2 shows that the mapping from average skill ratings to recommendations does not vary by committee member gender, consistent with the company's standardized aggregation rule.

Following the interviews, committees aggregate individual recommendations into a final hiring decision via deliberation. Overall, about 34% of interviews conducted by men and 32% of interviews conducted by women result in an offer. As reported in Appendix Figure B.3, offer rates vary widely across interview days, ranging from 0 to 100%.

3.1.2 Committee Member Characteristics

For each committee member, we observe gender, hiring experience measured as the number of years since the individual's first interview day,¹² an indicator for holding a management position, an indicator for being on temporary educational leave from the consulting role (e.g., to pursue an MBA), and an indicator for whether the interview takes place at the committee member's main location. Table 2 reports summary statistics for these characteristics, overall and by gender. The average interview is conducted by a committee member with approximately two years of hiring experience; about one-third are conducted by committee members in management positions.¹³ About 70% of interviews take place at the committee member's primary location, and 42% are conducted by committee members on temporary educational leave. None of these characteristics exhibits significant gender differences.

¹²We use time since the first interview day because the date of first participation is observed precisely for all committee members, whereas measures based on the number of prior interviews may be affected by incomplete data coverage before the start of our observation period.

¹³Beyond the management indicator, the data include anonymized job-title identifiers. Appendix Table B.2 reports their distribution by committee member gender; neither the individual title shares nor the overall distribution differ significantly between women and men.

Table 2: Summary Statistics: Committee Member Characteristics

	All CM		Male CM			Female CM			Difference	
	Mean	SD	Mean	SD	N	Mean	SD	N	Diff.	p-value
Hiring Experience (in years)	2.02	1.68	2.03	1.67	5722	2.00	1.71	2600	-0.026	0.89
Manager	0.34	0.47	0.33	0.47	5722	0.36	0.48	2600	0.023	0.67
Int Day at Main Location	0.70	0.46	0.70	0.46	5722	0.70	0.46	2600	-0.005	0.87
On Leave	0.42	0.49	0.43	0.49	5722	0.40	0.49	2600	-0.031	0.60

Note: CM = committee member. The unit of observation is the interview (i.e., the committee member–candidate pair). p-values are based on standard errors clustered at the committee member level. All characteristics are measured at the time of the interview. “Int Day at Main Location” takes the value one if the interview day takes place at the committee member’s assigned main location. The data include 359 distinct committee members (127 women and 232 men).

3.1.3 Candidate Characteristics

The data report several candidate characteristics that are predetermined with respect to the interview day. These include gender, field of study, whether the candidate applied for an internship or a permanent position, high-school GPA and final mathematics grade, as well as a CV score assigned by the HR department to screen candidates for interview invitations.

Summary statistics on candidate characteristics, both overall and by candidate gender, are reported in Appendix Table B.1. A third of both male and female candidates apply for internships. The majority of candidates (59%) studied Economics or Business Administration. The remaining candidates are distributed across Engineering and IT (11%), Mathematics, Physical Sciences, and Chemistry (7%), Life Sciences (6%), Law (3%), Social Sciences, and Humanities (5%), and an unknown/other category (9%). The field-of-study composition is broadly similar across genders, although a higher share of men has a background in STEM fields, whereas women are, for example, more prevalent in the Life Sciences as well as in the Social Sciences and Humanities. Male and female candidates exhibit broadly similar predetermined ability measures, with economically small differences: female candidates have a slightly better overall GPA performance, whereas male candidates hold slightly better mathematics grades and CV scores. In both groups, 14% of candidates achieved the best possible high-school GPA.

As a measure of candidates’ ex-ante likelihood of success in the hiring process, we estimate the probability of receiving an offer based on high-school GPA, mathematics grade, and

the CV score.¹⁴ Appendix Figure B.4 plots the distribution of predicted offer probabilities by candidate gender. The predicted probabilities exhibit meaningful variation, with a standard deviation of about 6 percentage points relative to an average probability of 34%. At the same time, the distributions are highly similar for male and female candidates, indicating comparable predicted probabilities of success based on pre-interview ability measures.

3.2 Assignment of Committee Members to Candidates

As described in Section 2, the HR department assigns each candidate to two or three committee members conducting one-on-one interviews. The only systematic assignment criterion is candidate gender: whenever feasible, each female candidate is assigned to at least one female committee member. Conditional on this requirement and the pool of evaluators at a given interview day, the remaining interviews are assigned as-good-as randomly, without taking candidate or committee member characteristics into account.

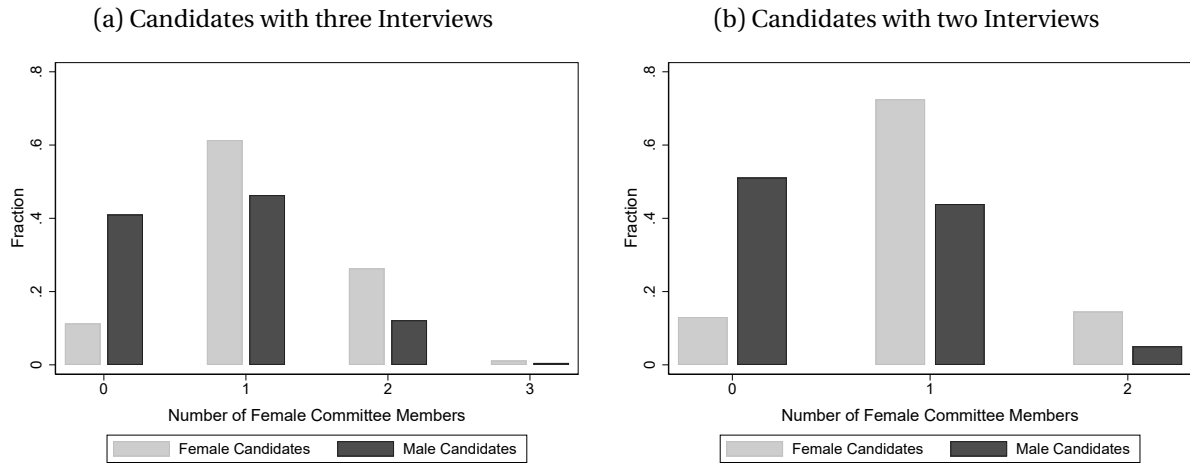
In the following, we first describe the resulting distribution of female committee members per candidate (3.2.1) and then assess whether the assignment procedure leads female and male committee members to evaluate comparable candidates conditional on candidate gender (3.2.2).

3.2.1 Number of Female Committee Members per Candidate

Figure 2 shows the distribution of female committee members across male and female candidates. Among candidates with three interviews (panel a), the most common committee composition consists of one woman and two men, applying to about 45% of male candidates and 60% of female candidates. Approximately 40% of male candidates and 10% of female candidates are interviewed exclusively by men, whereas all-female committees are rare. A similar

¹⁴Specifically, we estimate a linear probability model linking the final offer decision to high-school GPA, final mathematics grade, and the CV score, while excluding a candidate's own interview day from the estimation sample. We then use the estimated coefficients to predict the candidate's offer probability.

Figure 2: Number of Female Committee Members per Candidate



Note: The unit of observation is the candidate. N=2,500 in panel (a), N=485 in panel (b).

pattern holds for candidates with two interviews (panel b).¹⁵

3.2.2 Balance of Candidate Characteristics by Committee Member Gender

A core implication of the assignment procedure is that—within hiring committees (i.e., interview days) and conditional on candidate gender—female and male committee members evaluate comparable candidates. To assess whether this holds in the data, Table 3 examines the balance of candidate characteristics across male and female committee members. Specifically, it relates the probability that an interview is conducted by a female committee member (columns 1 and 2), as well as the candidate-level number of female committee members (columns 3 and 4), to a rich set of observable candidate characteristics. As expected, female candidates are significantly more likely than male candidates to be interviewed by a woman. Beyond candidate gender, however, candidate characteristics do not predict the assignment of a female committee member. In particular, a joint F-test fails to reject the null that candidate characteristics are jointly unrelated to the gender of the committee member. This holds in the

¹⁵Appendix Figure B.5 further reports the share of female committee members by interview day, which averages approximately 30%, corresponding to the share of female consultants at the relevant career level in the company.

Table 3: Balance of Candidate Characteristics Across Male and Female Committee Members

	P(Interview w/ Female CM)		Number of Female CM per Cand.	
	(1)	(2)	(3)	(4)
Female	0.162*** (0.010)	0.181*** (0.010)	0.442*** (0.029)	0.501*** (0.027)
Internship Application	-0.002 (0.013)	0.012 (0.014)	-0.007 (0.038)	0.026 (0.039)
CV Score (std.)	0.003 (0.004)	-0.002 (0.004)	-0.002 (0.012)	-0.006 (0.011)
High School GPA (Std.)	-0.006 (0.005)	-0.001 (0.005)	-0.022 (0.015)	-0.003 (0.015)
High School GPA: Top Grade	0.002 (0.015)	0.007 (0.013)	-0.013 (0.040)	0.025 (0.036)
High School Math Grade (Std.)	0.006 (0.006)	0.007 (0.005)	0.018 (0.016)	0.019 (0.014)
Study Field: Engineering/IT	-0.005 (0.015)	-0.016 (0.013)	-0.021 (0.042)	-0.045 (0.038)
Study Field: Maths/Physical Sciences/Chemistry	0.002 (0.017)	0.018 (0.014)	-0.026 (0.044)	0.050 (0.039)
Study Field: Life Sciences	0.013 (0.018)	-0.004 (0.017)	0.074 (0.054)	-0.006 (0.048)
Study Field: Law	0.022 (0.026)	0.015 (0.024)	0.052 (0.074)	0.050 (0.065)
Study Field: Social Sciences/Humanities	-0.006 (0.020)	-0.002 (0.019)	0.005 (0.054)	0.010 (0.052)
Study Field: Unknown/Others	-0.035** (0.017)	-0.008 (0.025)	-0.082 (0.050)	-0.022 (0.073)
Hiring Committee FE	No	Yes	No	Yes
p-value joint significance (excl. gender)	0.68	0.80	0.63	0.81
Outcome Mean	0.31	0.31	0.87	0.87
N	8322	8322	2985	2985

Note: In columns (1) and (2), the dependent variable is an indicator for whether the interview is conducted by a female committee member (CM), defined at the interview level. In columns (3) and (4), the dependent variable is the number of female committee members assigned to the candidate, defined at the candidate level. The omitted study field is economics/business. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

unconditional specifications (columns 1 and 3) and remains true in our main specification with hiring committee fixed effects (columns 2 and 4).

Appendix B provides further evidence consistent with the assignment procedure being conditionally independent of candidate or evaluator characteristics. Appendix Table B.3 shows that the balance of candidate characteristics also holds within the subsamples of male and female candidates. Appendix Table B.4 reports that female and male committee members are

not systematically assigned to evaluate different non-cognitive skill dimensions. Finally, Appendix Table B.5 shows that committee member characteristics are unrelated to candidate gender once hiring committee fixed effects are included.

3.3 Evaluation Behavior of Male and Female Committee Members

The previous subsection showed that female and male committee members are assigned comparable candidates. We now examine whether they evaluate candidates in comparable ways. Specifically, we ask whether recommendations by men and women are equally informative about candidate quality and whether they exhibit similar levels of disagreement with their co-evaluators.

3.3.1 Correlation of Recommendations with Predetermined Offer Probability

In Table 4, we examine whether recommendations by female and male committee members are equally informative about candidates' ex-ante quality. Specifically, we relate individual recommendations to our prediction of a candidate's probability to receive an offer based on predetermined ability measures (see Section 3.1.3). The results show that female committee members give more favorable recommendations on average, consistent with the leniency pattern documented in Table 1. This is particularly pronounced for female candidates (column 4), a pattern we return to in Section 6. Conditional on this level difference, recommendations by female and male committee members are equally correlated with the independent measure of candidate quality. Thus, we find no evidence that recommendations by female committee members are more or less informative than those made by men.

3.3.2 Disagreement

We next examine whether men and women differ in the extent to which they disagree with the other evaluators assessing the same candidate. We use two interview-level measures of disagreement: the share of co-evaluators giving a different recommendation and the average

Table 4: Correlation between Predicted Offer Probability and Recommendations

	Offer Recommendation (3-Point Scale)			
	(1) All	(2) All	(3) Male Cand.	(4) Female Cand.
Pred. Offer Probability (Std.)	0.089*** (0.012)	0.093*** (0.013)	0.079*** (0.017)	0.127*** (0.025)
Female CM	0.077*** (0.020)	0.061*** (0.020)	0.071** (0.030)	0.128*** (0.028)
Pred. Offer Probability (Std.) × Female CM	0.003 (0.018)	0.008 (0.018)	0.001 (0.024)	0.002 (0.028)
Hiring Committee FE	No	Yes	Yes	Yes
Outcome Mean	1.89	1.89	1.94	1.80
N	8322	8322	5030	3292

Note: The table reports OLS estimates from regressions of the committee member's (CM) recommendation on the candidate's predicted offer probability, interacted with committee member gender. The recommendation is specified on a 1–3 scale, with 1=reject, 2=potential offer, 3=offer. The offer probability is predicted based on predetermined ability measures (high-school GPA, mathematics grade, CV scores), see Section 3.1.3 for details. Columns (1) and (2) include all candidates, column (3) includes male candidates and column (4) includes female candidates. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

absolute difference, measured in recommendation points, between an individual's own recommendation and those of the co-evaluators.¹⁶

Table 5 shows that disagreement is common. On average, 57% of co-evaluators have made a different recommendation, with an average recommendation difference of 0.75 points. Women exhibit a slightly higher share of disagreement than men (58% versus 56%), although the difference is neither statistically significant nor reflected in the average recommendation difference. These patterns are similar across candidate gender. Appendix Figure B.6 further documents the absence of major differences in disagreement across evaluation committees with different gender compositions. Overall, we find no evidence of large differences in the disagreement encountered by female and male committee members.

¹⁶For example, if a committee member recommends rejection (1 point) while the two other evaluators recommend an offer (3 points), the share of disagreement equals one and the average recommendation difference equals two points.

Table 5: Disagreement Across Committee Members

	All CM		Male CM			Female CM			Difference	
	Mean	SD	Mean	SD	N	Mean	SD	N	Diff.	p-value
Share of Disagreement	0.57	0.39	0.56	0.39	5722	0.58	0.38	2600	0.013	0.15
Share of Disagreements, Male Cand.	0.58	0.38	0.57	0.38	3785	0.59	0.37	1245	0.015	0.16
Share of Disagreements, Female Cand.	0.55	0.40	0.55	0.40	1937	0.57	0.40	1355	0.021	0.17
Avg. Disagreement in Points	0.75	0.58	0.74	0.59	5722	0.76	0.57	2600	0.012	0.34
Avg. Disagreement in Points, Male Cand.	0.76	0.58	0.76	0.58	3785	0.77	0.56	1245	0.017	0.31
Avg. Disagreement in Points, Female Cand.	0.73	0.59	0.72	0.59	1937	0.74	0.59	1355	0.022	0.31

Note: CM = committee member. The unit of observation is the interview (i.e., the committee member–candidate pair). “Share of disagreement” denotes the share of co-evaluators who gave a different recommendation on the same candidate. “Avg. Disagreement in Points” denotes the average absolute difference in recommendation points (1–3 scale) between an individual’s own recommendation and those of the co-evaluators. p-values are based on standard errors clustered at the committee member level. The data include 359 distinct committee members (127 women and 232 men).

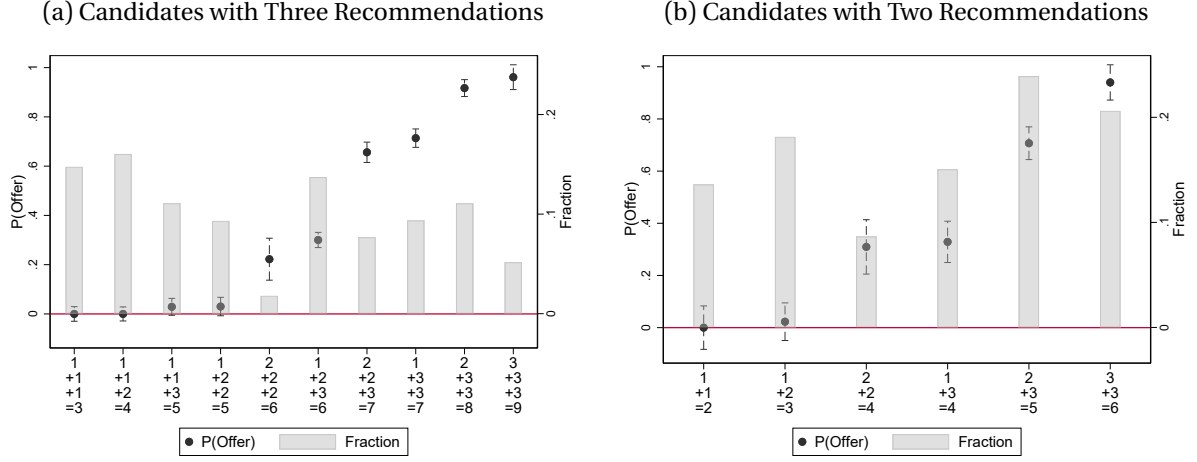
4 Aggregation of Information

In this section, we examine how individual recommendations are aggregated into committee decisions. Specifically, we analyze the extent to which initial recommendations predict hiring outcomes and assess the scope for deliberation in the decision process.

Committee members share the common objective of hiring candidates who meet the company’s predefined selection criteria. Because they independently observe the same candidate in separate one-on-one interviews, there remains substantial scope for disagreement, for example due to the subjective nature of interviews or circumstantial factors.¹⁷ Moreover, each interview assesses a different, quasi-randomly assigned subset of non-cognitive skills (see Section 2 and Appendix Table B.4). While the company does not prescribe a deterministic cutoff or majority voting rule for aggregating divergent recommendations, committees may nevertheless behave as if such rules were in place. We therefore begin by examining the empirical mapping from recommendation profiles to hiring decisions.

¹⁷An example of circumstantial factors influencing interview evaluations is provided in Radbruch and Schiprowski (2024), showing that interviewers tend to evaluate candidates relative to the previous candidate in the interview sequence.

Figure 3: Aggregation of Information in the Committee



Note: The unit of observation is the candidate. $N=2,500$ in panel (a), $N=485$ in panel (b). The x-axis shows the combination of initial recommendations per candidate, where 1='reject'; 2='potential offer'; 3='offer'. The left y-axis shows the probability that a given combination results in a positive offer decision by the hiring committee. The right y-axis shows the fraction of a given combination in the data. Dashed lines show 95% confidence intervals.

Figure 3 illustrates the probability of receiving an offer conditional on the combination of individual recommendations. The figure also reports the empirical frequency of each combination in the data. As expected, offer probabilities increase with the total number of recommendation points. However, the relationship is not deterministic: the probability of receiving an offer does not jump from zero to one upon reaching a particular threshold or a majority of positive recommendations. Instead, the pattern suggests a distinction between three types of cases. About 50% of candidates receive very low initial support (≤ 5 points in panel a; ≤ 3 points in panel b), resulting in an offer probability close to zero. About 15% receive strong support (≥ 8 points in panel a; 6 points in panel b) and obtain an offer with probability close to one. The remaining 35% of candidates receive an intermediate level of support (6 or 7 points in panel a; 4 or 5 points in panel b), for which offer probabilities lie well inside the unit interval. In these cases, the final decision is difficult to predict based on the initial recommendations alone, suggesting greater scope for deliberation to affect the outcome. We refer to these cases as 'deliberation cases'.

As additional evidence of non-deterministic aggregation, Appendix Figures C.1 and C.2 show that similar patterns emerge when using candidates' average sub-ratings instead of recommendation profiles or when ranking candidates within interview days based on their ratings. Appendix Table C.1 further quantifies the predictive power of initial recommendations and ratings. The recommendation profile alone explains about 60% of the variation in offer decisions in the full sample and 16% among deliberation cases. After controlling for hiring committee fixed effects, candidate characteristics (gender, field of study, high-school grades, CV score, and indicator for internship application), as well as sub-ratings of cognitive and non-cognitive skills, these shares increase to 63% and 28%, respectively, leaving a substantial share of the variation unexplained.

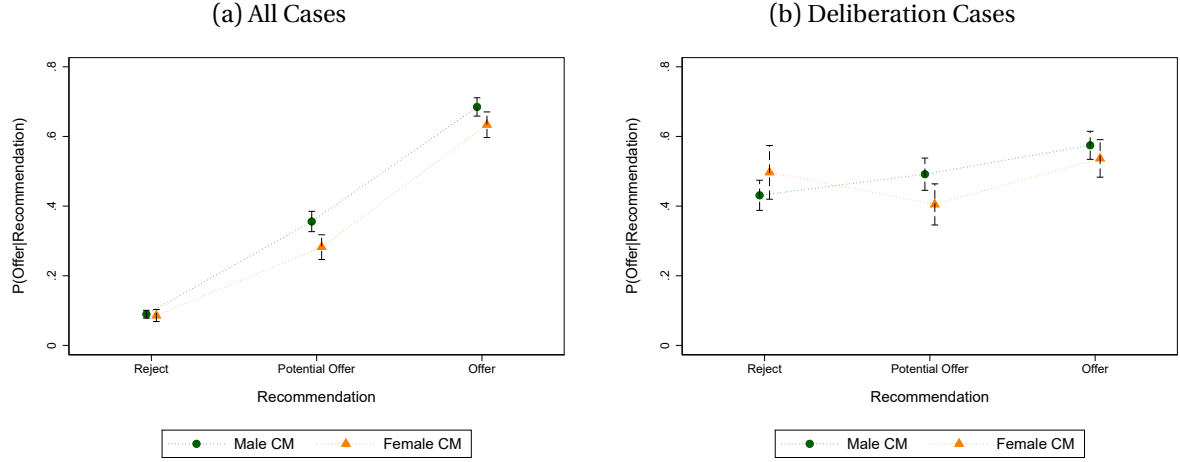
Taken together, individual recommendations are strong but imperfect predictors of final hiring outcomes.¹⁸ As a result, committee decisions cannot be summarized by voting or cut-off rules that assign equal weight to each recommendation, leaving room for deliberation to influence how individual recommendations are aggregated into final hiring decisions.

5 Gender Differences in Influence

The previous section has shown that the aggregation of individual recommendations into committee decisions is not deterministic, leaving scope for deliberation. We now examine how recommendations made by female and male committee members influence decisions. We first provide descriptive evidence on how recommendations translate into final decisions by committee member gender. Next, we formally estimate the average gender gap in the probability that a recommendation aligns with the final committee decision, holding deliberation conditions constant. We then examine the determinants of this alignment gap. Finally, we discuss how prior experimental evidence on behavioral differences between men and women may help explain the results.

¹⁸The observed patterns could in principle be explained by a cutoff rule using precision-weighted recommendations. However, this would require that committee members can quantify and communicate the true precision of their interview signals, which is unlikely to hold in practice.

Figure 4: Translation of Recommendations into Offers (Raw Data)



Note: CM = committee member. The unit of observation is the interview recommendation (i.e., candidate–committee member pair). $N=8,322$ in panel (a); $N=2,851$ in panel (b). Deliberation cases refer to candidates for whom the final hiring decision is ex-ante ambiguous given the initial recommendations (see Section 4). 95% confidence intervals, with standard errors clustered at the level of the hiring committee ($N=444$).

5.1 Descriptive Evidence

Figure 4 illustrates the probability that a given recommendation translates into an offer, by committee member gender. Panel (a), based on the full sample of committee decisions, shows that the relationship between recommendations and offer probabilities is flatter for female committee members. This difference is driven by positive recommendations, which are 5 to 7 percentage points less likely to result in an offer when made by a woman. Panel (b) restricts the analysis to deliberation cases, in which the committee decision is ex-ante ambiguous given the initial recommendations (see Section 4). In these cases, the gender difference in alignment also extends to negative assessments: a reject recommendation made by a woman is about 7 percentage points more likely to be overturned and result in an offer than a reject recommendation made by a man. Moreover, the relationship between recommendations and offer probabilities remains positive for men but not for women.

In Appendix Table D.1, we complement the descriptive evidence by estimating gender differences in recommendation–decision slopes conditional on committee member fixed effects and detailed controls for both own and co-evaluators’ cognitive and non-cognitive skill rat-

ings. The recommendation–decision slope is consistently 8–13% flatter for women than for men, even when controlling for the underlying support as measured by the skill ratings. Moreover, the table shows that this difference is driven by non-cognitive skill assessments, whose correlation with offers is about 25% flatter among women. A possible explanation is that non-cognitive evaluations leave greater scope for disagreement because committee members assess different non-cognitive skills, but the same cognitive skills (see Section 2).

5.2 Empirical Strategy

The presented evidence suggests that recommendations made by women are less likely to align with final hiring decisions than those made by men. One possible interpretation is that women’s assessments have less influence during deliberation. However, differences in alignment may also reflect differences in deliberation conditions, such as the recommendation constellations under which men and women enter deliberation.

In the following, we seek to isolate the component of the gender gap that remains after holding deliberation conditions constant. We first introduce our econometric specification and successive sets of controls. We then discuss the assumptions under which the resulting residual gap can be related to gender differences in influence during deliberation.

Econometric Specification As a summary measure of recommendation–decision alignment, we define an indicator that equals one when a reject recommendation translates into a rejection and when a (potential) offer recommendation translates into a job offer.¹⁹ We then estimate how the alignment outcome varies with committee member gender using the following linear specification:

$$(1) \quad A_{i,j} = \beta F_j + \omega_{c(i,j)} + X'_{i,j} \delta + \epsilon_{i,j}$$

The outcome $A_{i,j}$ indicates whether the committee’s decision on candidate i aligns with

¹⁹Results are robust to excluding ‘potential offer’ recommendations, which arguably express increased uncertainty in the assessment (see Panel B of Appendix Table D.6).

the initial recommendation made by committee member j . F_j equals one if committee member j is female. The vector $\omega_{c(i,j)}$ includes hiring committee fixed effects, corresponding to the level at which candidates are assigned to evaluators and at which deliberation takes place. Since hiring committees coincide with interview days, $\omega_{c(i,j)}$ also absorbs all other time- and location-specific factors.

The coefficient of interest, β , measures the gender gap in the probability of alignment. Its interpretation depends on the set of control variables included in $X_{i,j}$.

As a first set of controls, we include committee member and candidate characteristics (see Table 2 and Appendix Table B.1), together with indicators for the type of non-cognitive skills assessed in the interview.²⁰ Importantly, this set includes candidate gender, which influences the assignment of committee members to candidates (see Sections 2 and 3.2). Under this specification, the estimate of β captures the overall gender difference in recommendation–decision alignment before accounting for potential differences in recommendation behavior or case-specific disagreement.

Second, we control for leniency using the committee member’s average recommendation on the 1–3 scale, leaving out the recommendation on candidate i . This control accounts for the fact that women give more favorable recommendations on average (see Section 3.1), which may be less likely to be adopted by the committee.

Third, we condition on the exact recommendation constellation that committee members face when entering deliberation on a given candidate. Specifically, we include case-type indicators for every possible combination of initial recommendations (corresponding to the bins in Figure 3). We also account for the fact that committee members within the same case may face different opportunities for alignment depending on their disagreement with their co-evaluators. For example, in a case with one negative and two positive recommendations, the dissenting committee member faces a lower ex-ante probability of alignment than the other two evaluators. To capture this variation, our most demanding specification interacts the case-type indicators with an individual-specific measure of disagreement, defined as the average

²⁰Because the non-cognitive skill dimensions are assessed in a fixed order, the skill-type indicators also control for interview order.

pairwise absolute difference between the individual's own recommendation and those of the other committee members who interviewed the same candidate.²¹

Under the full set of controls, we interpret β as the average gender difference in recommendation–decision alignment that arises during deliberation.

Main Assumptions Our successive interpretation of β relies on two key assumptions. First, differences in alignment between female and male committee members should not be driven by systematic differences in the candidates they evaluate. This assumption would be violated, for example, if women were systematically assigned to candidates with lower hiring prospects or noisier skill signals. As described in Section 2, the assignment of candidates to evaluators within hiring committees is determined by the HR department and depends only on candidate gender and the pool of available evaluators. Consistent with this assignment rule, Section 3.2 has shown that observable candidate characteristics are unrelated to committee member gender conditional on candidate gender and hiring committee fixed effects (Tables 3 and B.3), and that female and male committee members are not assigned to assess different non-cognitive skills (Appendix Table B.4). The robustness analyses presented later in this section further assess the sensitivity of the results to relaxing assumptions about candidate assignment.

Second, after accounting for case type and disagreement, we assume that remaining gender differences in alignment arise during the deliberation process. We will assess the plausibility of this assumption by examining whether the residual alignment gap is concentrated in settings where deliberation is most relevant, remains robust to alternative controls for deliberation conditions, and disappears under hypothetical decision processes that aggregate the same interview evaluations using a fixed cutoff rule instead of deliberation.

Table 6: Gender Difference in the Alignment of Recommendations with Committee Decisions

	Alignment Probability					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.051*** (0.011)	-0.053*** (0.012)	-0.047*** (0.012)	-0.035*** (0.011)	-0.028*** (0.008)	-0.005 (0.006)
Female x Delib. Case						-0.066*** (0.020)
Experience (Std.)		0.030*** (0.006)	0.030*** (0.006)	0.029*** (0.005)	0.012*** (0.004)	0.002 (0.004)
Experience (Std.) x Delib. Case						0.026*** (0.010)
Hiring Committee FE	No	Yes	Yes	Yes	Yes	Yes
Candidate & CM Characteristics	No	Yes	Yes	Yes	Yes	Yes
Leniency	No	No	Yes	Yes	Yes	Yes
Case Type	No	No	No	Yes	Yes	Yes
Case Type x Disagreement	No	No	No	No	Yes	Yes
R-Squared	0.00	0.01	0.01	0.22	0.57	0.57
Outcome Mean	0.68	0.68	0.68	0.68	0.68	0.68
N	8322	8322	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated by OLS based on equation (1), with varying sets of control variables as indicated at the bottom of the table. The controls for candidate and committee member (CM) characteristics include all variables reported in Table 2 and Appendix Table B.1, as well as indicators for the set of non-cognitive skills assessed in the interview. The leniency control is the committee member's leave-one-out mean recommendation on the 1–3 scale. Case type controls include indicators for each combination of initial recommendations (the bins shown in Figure 3). Disagreement is measured as the average pairwise absolute difference between a committee member's own recommendation and the recommendations of the other two members for a given candidate. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Main Results

Table 6 reports estimates of the gender gap in recommendation–decision alignment based on equation (1). Column (1) presents the raw alignment gap, which amounts to about 5 percentage points (8% relative to the male average). Column (2) adds hiring committee fixed effects, candidate and committee member characteristics, and indicators for the (non-cognitive) skill

²¹For example, if one committee member recommended rejection while the other two recommended an offer (1 and 3, respectively), the measure takes the value $\frac{|3-1|+|3-1|}{2} = 2$. This measure only varies within cases with three non-unanimous recommendations.

set assessed in the interview. Consistent with the quasi-random assignment of candidates and the similar observable characteristics of female and male committee members, these controls have little effect on the estimate.

At this stage, the estimated alignment gap captures overall gender differences in the extent to which individual recommendations are reflected in the company's hiring decisions. As discussed above, part of this gap may arise from differences in recommendation behavior or in the recommendation constellations that committee members face when entering deliberation. We now successively control for these factors to isolate the residual alignment gap. Column (3) introduces the committee member's leave-one-out mean recommendation as a measure of leniency, leading to a small but statistically insignificant reduction in the estimated gap. Columns (4) and (5) introduce increasingly flexible controls for individual deliberation conditions. Column (4) adds indicators for the full combination of initial recommendations (corresponding to the bins in Figure 3), while Column (5) additionally interacts these indicators with individual-specific disagreement. After accounting for these factors, the estimated gap falls to about 2.8 percentage points, implying that about 55% of the raw gender gap remains unexplained.

As documented in Section 4, the scope for deliberation is greatest when committee members disagree in their initial recommendations. We therefore expect the residual alignment gap to be concentrated in these deliberation cases. The results in Column (6) are in line with this conjecture. Within deliberation cases, the residual alignment gap increases to about 7 percentage points, corresponding to a 13% difference relative to the male average within this subsample.

Table 6 also reports on the role of experience, measured as the time since the committee member's first interview day. In the full specification, a one standard deviation increase in experience is associated with a 1.2 percentage point higher probability of alignment (column 5). As with the gender gap, this association is concentrated in deliberation cases, where the corresponding estimate increases to about 2.6 percentage points. Relative to experience, the gender gap is substantial: it corresponds roughly to a two standard deviation decrease in experience (≈ 3 years), both in the full sample and among deliberation cases. Appendix Table D.2 shows

Table 7: Alignment under Deliberation vs. Hypothetical Cutoff Rule

	Alignment Probability under Different Decision Rules					
	Deliberation		Cutoff: Top 35% per Year		Cutoff: Top 35% per Quarter	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.028*** (0.008)	-0.005 (0.006)	-0.004 (0.008)	0.001 (0.007)	-0.004 (0.009)	0.000 (0.007)
Female x Delib. Case		-0.066*** (0.020)		-0.015 (0.020)		-0.013 (0.020)
R-Squared	0.57	0.57	0.53	0.53	0.52	0.52
Outcome Mean	0.68	0.68	0.71	0.71	0.71	0.71
N	8322	8322	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision (actual or hypothetical) aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1), including the full set of controls. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). Columns (1) and (2) replicate columns (5) and (6) of Table 6. In columns (3) to (6), we assign hypothetical offers to candidates whose average ratings are in the top 35% of a given year (columns 3 to 4) or quarter (columns 5 to 6). We then compute the alignment of recommendations with the hypothetical offer decision. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that the remaining committee member characteristics are not significantly associated with alignment in the full specification.

Placebo Check: Alignment Outcomes Under a Hypothetical Cutoff Rule The previous analysis estimated the gender gap in recommendation–decision alignment that remains after accounting for observable deliberation conditions. We now provide an additional assessment of whether this residual gap arises during deliberation. To this end, we compare the observed committee decisions with hypothetical hiring outcomes implied by a cutoff rule that only takes pre-deliberation interview evaluations into account. The idea is that a gender gap driven by deliberation should disappear under a mechanical aggregation rule that shuts off the deliberation stage.

Specifically, we assign 'placebo offers' to candidates whose average interview rating falls within the top 35% of candidates interviewed in the same year (columns 3 and 4) or the same quarter (columns 5 and 6). Importantly, both the hypothetical cutoff rule and the observed

committee decisions are based on the same underlying interview evaluations, thereby holding fixed any gender differences in assessments. The only difference is that the hypothetical rule replaces committee deliberation with a mechanical aggregation procedure.

Table 7 shows that the hypothetical cutoff rules do not generate any significant gender gap in alignment. This suggests that the gap observed in the data is not a mechanical consequence of gender differences in evaluation behavior, but arises during the deliberation process. Appendix Table D.3 shows that this result is robust to using alternative cutoff rules.

Additional Heterogeneity & Robustness Appendix D provides further heterogeneity analyses and robustness checks. Appendix Table D.4 examines how the alignment gap varies with committee member and candidate characteristics. Candidate gender is the only dimension with a statistically significant heterogeneity, with the alignment gap being larger among female candidates. We return to this finding and its implications in Section 6.

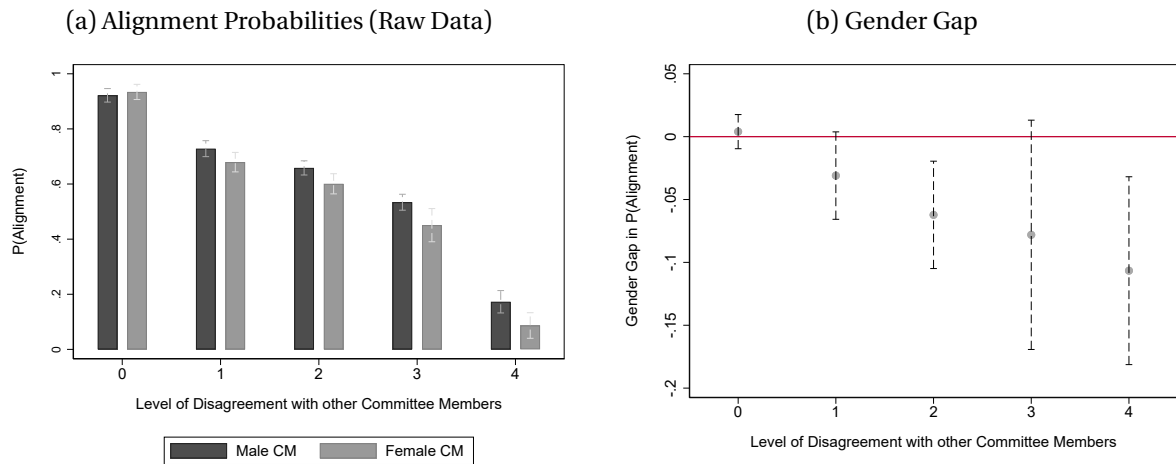
Appendix Table D.5 demonstrates that the estimated gender gap is unaffected by alternative sets of control variables. In particular, it is robust to flexible controls for the ratings underlying recommendations and to the inclusion of candidate (i.e., evaluation committee) fixed effects, under which the gender gap is identified solely from within-candidate comparisons. Finally, Appendix Table D.6 documents robustness to excluding “potential offer” recommendations, candidates who undergo an additional final interview,²² candidates with two interviews, and interview days with a single female interviewer.

5.4 Role of Disagreement

The previous results have shown that the residual alignment gap is concentrated in cases with scope for deliberation. We further investigate the role of disagreement in Figure 5. If the gap indeed arises during deliberation, it should be largest when committee members disagree most strongly about a candidate.

²²The smaller coefficient in the full sample (column 1) reflects the fact that candidates forwarded to a final interview are disproportionately drawn from deliberation cases, where the gender gap in alignment is concentrated. Importantly, excluding candidates with final interviews does not affect the estimated gap within deliberation cases (column 2).

Figure 5: Gender Gap in Recommendation–Decision Alignment, by Level of Disagreement



Note: CM = committee member. The analysis only includes candidates with three recommendations ($N = 7,354$). The alignment outcome equals one if the committee's final hiring decision aligns with the committee member's initial recommendation. The level of disagreement is measured as the sum of the pairwise absolute differences between a member's own recommendation and the other two members' recommendations on the same candidate. Results in panel (b) are based on OLS estimates of equation (1) including hiring committee fixed effects, committee member and candidate controls, leniency, and case-type indicators, where the female indicator is interacted with indicators for the sum of pairwise disagreement between a committee member's own recommendation and those of the other evaluators. 95% confidence intervals, with standard errors clustered at the level of the hiring committee ($N = 375$).

Figure 5 relates the gender gap in alignment to disagreement, defined as the sum of the pairwise absolute differences between a committee member's own recommendation and those of the co-evaluators assessing the same candidate.²³ Panel (a) reports raw alignment probabilities for male and female committee members. Alignment generally declines with disagreement, reflecting that individual recommendations become less predictive of final hiring decisions the more evaluators disagree. Panel (b) reports the estimated gender gap based on equation (1). As expected, there is no gender gap when all committee members agree in their recommendations. The gap then increases monotonically with disagreement, reaching approximately 10 percentage points at the highest level of disagreement (60% relative to men).

²³To ensure that the measure is comparable across decisions, we restrict this analysis to cases with three recommendations per candidate (88% of the data).

Table 8: Role of the Gender Composition

	All Cases		Deliberation Cases	
	(1)	(2)	(3)	(4)
Male Majority $\times$ Female	-0.055*** (0.015)	-0.029*** (0.008)	-0.073*** (0.027)	-0.092*** (0.025)
Parity $\times$ Female	-0.048 (0.032)	-0.042 (0.032)	-0.046 (0.042)	-0.033 (0.043)
Female Majority $\times$ Female	0.014 (0.029)	0.006 (0.022)	0.042 (0.060)	0.015 (0.059)
Hiring Committee FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Outcome Mean	0.68	0.68	0.52	0.52
p-value: Male Maj.= Female Maj.	0.03	0.12	0.08	0.09
N	8322	8322	2851	2851

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1), with the female committee member indicator interacted with the gender composition of the evaluation committee. Male majority = 2 men, 1 woman (N=3,846). Parity = 1 man, 1 woman (N=544). Female majority = 2 women, 1 man (N=1,317). The remaining observations come from all-male and all-female evaluation committees. They are included in the estimation sample but do not contribute to the identification of the gender-gap interactions. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5 Role of the Gender Composition

As in many high-level decision-making bodies, men are in the majority in most of the evaluation committees. Specifically, about 50% of candidates are assessed by two men and one woman (see Figure 2). If influence during deliberation depends on the gender composition of the committee, we would expect the alignment gap to vary across these settings. We therefore compare decisions reached under a 2:1 male majority, a 2:1 female majority, and parity (1:1).²⁴

Table 8 reports the results. Columns (1) (committee fixed effects only) and (2) (full set of controls) include all cases. The results suggest that the gender difference in alignment is driven

²⁴As described in Section 2, the full hiring committee includes more evaluators than the three members of the evaluation committee who interview a given candidate. While the gender composition of the broader hiring committee can also influence the deliberation dynamics, we focus on variation in the composition of the evaluation committee for two reasons. First, its composition results from a quasi-random assignment process. Second, anecdotal evidence suggests that the evaluators who interviewed the candidate are most important in the deliberation.

by committees with two men and one woman. By contrast, no gap in either direction emerges when women are in the majority. Due to the limited number of parity cases, the estimates are imprecise for this group, but provide suggestive evidence of a negative gap. Columns (3) and (4) show that the same pattern holds when restricting the sample to deliberation cases. In these cases, the gender gap increases to approximately 9 percentage points under a male majority.²⁵ We interpret these findings as suggestive evidence that differences in influence arise primarily in deliberation environments where women constitute a minority.

5.6 Discussion of Potential Mechanisms

The results in this section have shown that committee decisions are systematically less aligned with the initial recommendations of women than with those of men. Since women in our setting possess similar training and experience in conducting interviews, it is unlikely that they produce less informative assessments. Consistent with this interpretation, recommendations made by male and female committee members are similarly correlated with predicted offer probabilities based on predetermined ability measures (Table 4). Moreover, the alignment gap does not appear to be a simple consequence of minority status. While women are less influential when being in the minority, this does not hold for men (Table 8). Finally, survey responses from a subset of committee members indicate that men and women hold similar levels of confidence in the quality of their recommendations (Appendix Figure D.1).

Altogether, the findings suggest that the gender gap in alignment is unlikely to be driven by differences in information quality and instead reflects differences in influence during deliberation. A large literature in psychology and economics has documented gender differences in behaviors that are relevant for deliberations. Laboratory studies in psychology find that women exert less influence on group decisions than men and identify resistance to competent women as an important mechanism (e.g., Butler & Geis, 1990; Carli, 1989, 2001). Consistent with this interpretation, Guo and Recalde (2023) show that experimental participants are

²⁵Given the institutional design of the interview process described in Section 2, it is unlikely that initial recommendations respond strategically to the gender composition of the evaluation committee. Consistent with this, Appendix Figure B.1 shows that the recommendations of male and female evaluators do not vary systematically with the committee's gender composition.

more likely to override a woman's opinion than a man's in situations of disagreement, while Bowles et al. (2007) document that women face penalties when engaging in negotiation. Experimental evidence from economics further suggests that women are less likely to voice and defend their opinions (e.g., Coffman, 2014; Hardt et al., 2025) or assume leadership roles in group discussions (e.g., Born et al., 2022; Alan et al., 2019). These behavioral differences may be amplified when women are in the minority (e.g., Brito et al., 2024; Karpowitz et al., 2024) or face pressure to conform to gender norms (e.g., Burztnyn et al., 2017), consistent with our finding that the alignment gap is strongest in male-majority committees. Finally, advocating for or against a candidate can be viewed as a form of competition, particularly when disagreement about candidate quality is substantial. Gender differences in alignment would therefore also be consistent with the well-documented literature on gender differences in competitiveness (e.g., Gneezy et al., 2003; Niederle & Vesterlund, 2007).

Taken together, the experimental literature documents a range of behavioral differences between men and women that could plausibly translate into differences in influence within committees. While the underlying mechanisms are closely related and difficult to distinguish in observational data, they have in common that they are unrelated to the informational content of individual assessments. The findings therefore suggest that the aggregation of individual information is influenced by behavioral factors that are irrelevant to the actual decision problem.

6 Implications for the Effects of Gender Quotas

Many organizations target the representation of women on hiring committees with the goal of improving the gender balance of their hires. This also applies to the setting we study, where female candidates at the interview stage are about 17% less likely to receive an offer than male candidates. However, previous evidence points to zero or even negative effects of committee gender quotas on the outcomes of female candidates (e.g., Bagues et al., 2017; Deschamps, 2024). This could reflect that women do not assess female candidates more favorably. Alternatively, women's influence on committees might not be sufficient to alter the final hiring de-

cision.²⁶ These channels are difficult to distinguish in standard data sources, which typically do not report the pre-deliberation evaluations of individual committee members.

In the following, we use the structure of our data to independently analyze the effects of female committee members on individual evaluations from their effects on committee-level decisions. We first estimate to what extent female committee members exhibit a smaller gender gap in their recommendations. We then examine how the gender gap in final offer decisions varies with the representation of women at the committee level. The analysis relies on the quasi-random assignment of committee members to candidates within interview days, conditional on candidate gender (see Section 3.2). This assignment generates exogenous variation in committee gender composition among both male and female candidates. Consistent with the assignment rule, Tables 3 and B.3 show that candidate characteristics are balanced across male and female committee members, both overall and within candidate gender.

Individual Evaluation Behavior We start by estimating whether the probability that an interview results in a positive recommendation (≥ 2 points) differs between male and female candidates, and how this difference varies with the gender of the committee member. Column (1) of Table 9 reports that female candidates are on average about 8 percentage points ($\approx 13\%$) less likely to receive a positive recommendation than male candidates. Column (2) shows that this average gap masks substantial heterogeneity with respect to committee member gender. Specifically, female committee members exhibit a 50% smaller gender gap than their male colleagues. This result persists after introducing hiring committee fixed effects (column 3) and control variables (column 4). Moreover, it is robust to a specification with candidate fixed effects (column 5), which compares the recommendations of male and female committee members evaluating the same candidate.

As a placebo check, Appendix Table E.1 reports specifications in which the outcome is replaced with the candidate's ex-ante offer probability as predicted from predetermined ability measures (see Section 3.1.3 for details). As expected under the quasi-random assignment

²⁶A further possible reason is that men might become less supportive of female candidates when women enter the committee (see suggestive evidence in Bagues et al., 2017; Deschamps, 2024).

Table 9: Gender Evaluation Gap Among Male and Female Committee Members

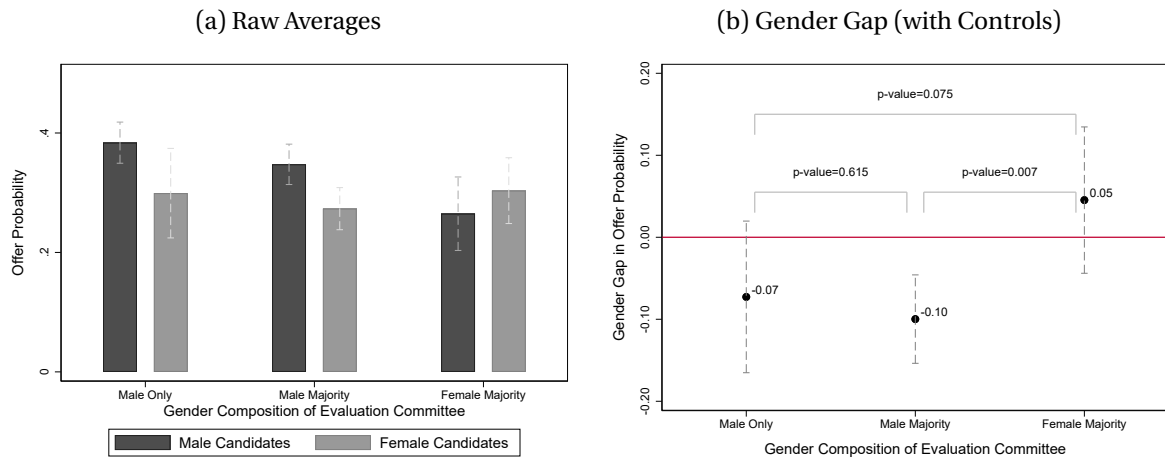
	P(Positive Offer Recommendation)				
	(1)	(2)	(3)	(4)	(5)
Female Candidate	-0.084*** (0.012)	-0.112*** (0.013)	-0.114*** (0.014)	-0.116*** (0.014)	
Female CM		0.039** (0.016)	0.034** (0.017)	0.029* (0.017)	0.028 (0.018)
Female Candidate x Female CM		0.054** (0.023)	0.048** (0.023)	0.055** (0.023)	0.053** (0.024)
Hiring Committee FE	No	No	Yes	Yes	No
Controls	No	No	No	Yes	Yes
Candidate FE	No	No	No	No	Yes
Outcome Mean	0.58	0.58	0.58	0.58	0.58
N	8322	8322	8322	8322	8322

Note: CM = committee member. The dependent variable is an indicator equal to one if the committee member gave a positive offer recommendation (≥ 2 points on the 1–3 recommendation scale). Control variables include committee member and candidate characteristics as reported in Table 2 and Appendix Table B.1, as well as indicators for the set of non-cognitive skills assessed in the interview. In column (5), controls for candidate characteristics are excluded due to their collinearity with candidate fixed effects. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of committee members, we find no analogous differences. This supports the interpretation that the gender differences in recommendations arise from differences in evaluation behavior rather than from the differential assignment of candidate types to male and female committee members.

Effect of Committee Composition on Offer Decisions Given the evidence presented so far, it remains unclear how the presence of female committee members affects the gender composition of new hires. While recommendations by female committee members exhibit a lower gender gap, Section 5 showed that they have a lower influence on committee decisions. Appendix Table E.2 additionally reports that this holds particularly true for female candidates. Thus, while adding a female committee member increases the likelihood that a female candidate receives a positive recommendation, such a recommendation is less likely to translate

Figure 6: Gender Composition of the Committee and Gender Balance in Hiring Decisions



Note: Panel (a) reports raw offer probabilities by candidate gender and committee composition. The estimates of the gender gap in Panel (b) are based on OLS regression with hiring committee fixed effects and controls for candidate characteristics as reported in Table B.1. The unit of observation is the candidate ($N=2,985$). Parity and all-female committees are included in the estimation but omitted from the figure because of the small number of candidates. Dashed lines indicate 95% confidence intervals, with standard errors clustered at the level of the hiring committee ($N=444$).

into a job offer.

To estimate the net effect of these two opposing forces, we analyze how the gender gap in offer probabilities varies with the gender composition of the evaluation committee. Panel (a) of Figure 6 reports raw offer probabilities for male and female candidates by evaluation committee composition. Panel (b) plots the corresponding gender gaps after controlling for hiring committee fixed effects and candidate characteristics.

Among candidates evaluated by all-male committees, the gender gap in offer rates amounts to about 7 percentage points, relative to a male offer rate of 38%. Replacing one male committee member with a woman does not significantly change this gap. This suggests that the presence of a lone woman does not measurably affect the gender composition of hired candidates, consistent with previous findings in the literature (e.g., Bagues et al., 2017). The pattern changes markedly once women constitute the majority of the evaluation committee. In these cases, the gender gap in offer rates disappears, reflecting both a lower offer rate among male

candidates and a higher offer rate among female candidates.²⁷ This pattern parallels the results in Table 8, which show that the gender gap in influence disappears when women constitute the majority. Overall, the pronounced non-linearity suggests that deliberative dynamics play an important role for how committee diversification affects hiring outcomes.

7 Conclusion

Committee decision-making is ubiquitous in firms and organizations. This paper documents systematic gender differences in individuals' influence on the outcomes of committee deliberation. After accounting for pre-deliberation conditions, committee decisions remain less aligned with the initial recommendations of women than with those of men. The residual gender gap in alignment is comparable in magnitude to the effect of a two standard deviation decrease in committee members' experience. It is most pronounced when disagreement is high and when deliberation takes place in male-majority committees. These findings indicate that gender differences in group behavior documented in prior experimental research are relevant in the context of high-stakes decision processes.

The results inform policies aimed at increasing diversity in committees and on corporate boards. Systematic gender differences in influence may reduce the effectiveness of such policies, particularly when quotas introduce "lone women." More generally, differences in influence that are unrelated to the quality of information may prevent committees from efficiently aggregating all available information.

Given the prevalence of group decision-making in the labor market, awareness about potential gender differences in influence is important. However, designing suitable policy responses is not straightforward. Avoiding women's minority status appears to be a natural approach, but is often not feasible, particularly in environments where women represent a small share of the workforce. Moreover, such policies may increase inequalities in the allocation of committee work and thereby exacerbate existing gender disparities in non-promotable tasks

²⁷Appendix Figure E.1 reports the corresponding placebo analysis, replacing the actual offer decision with the candidate's predicted offer probability based on predetermined ability measures. The placebo outcome exhibits no significant gender gap across committee compositions.

(Babcock et al., 2017). A possible alternative is to design decision procedures that allow for a balanced participation among committee members. Understanding how organizational procedures shape the distribution of influence within committees, and how such changes ultimately affect the quality of downstream decision outcomes, remains an important question for future research.

References

- Adams-Prassl, A., Huttunen, K., Nix, E., & Zhang, N. (2024). Violence against Women at Work. *The Quarterly Journal of Economics*, 139(2), 937–991.
- Alan, S., Ertac, S., Kubilay, E., & Loranth, G. (2019). Understanding Gender Differences in Leadership. *The Economic Journal*, 130(626), 263–289.
- Austen-Smith, D., & Feddersen, T. J. (2006). Deliberation, preference uncertainty, and voting rules. *American Political Science Review*, 100(2), 209–217.
- Azmat, G., & Boring, A. (2020). Gender diversity in firms. *Oxford Review of Economic Policy*, 36(4), 760–782.
- Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3), 714–747.
- Bagues, M., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? evidence from a repeated randomized experiment. *Review of Economic Studies*, 77(4), 1301–1328.
- Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, 107(4), 1207–1238.
- Bai, J., Esche, M., MacLeod, B., & Shi, Y. (2022). Subjective evaluations and stratification in graduate education. *NBER Discussion Paper No. 30677*.
- Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2018). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies*, 86(1), 191–239.
- 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.
- Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1), 84–103.
- Brito, E., Bruhn, J., Choon, T. H., & Weber, E. A. (2024). *Gender composition and group behavior: Evidence from us city councils* (tech. rep.).
- Burdea, V., & Woon, J. (2025). Getting it right: Communication, voting, and collective truth finding. *European Journal of Political Economy*, 90, 102768.
- Burztnyn, L., Fujiwara, T., & Pallais, A. (2017). Acting wife: Marriage market incentives and labor market investments. *American Economic Review*, 107(11), 3288–3319.

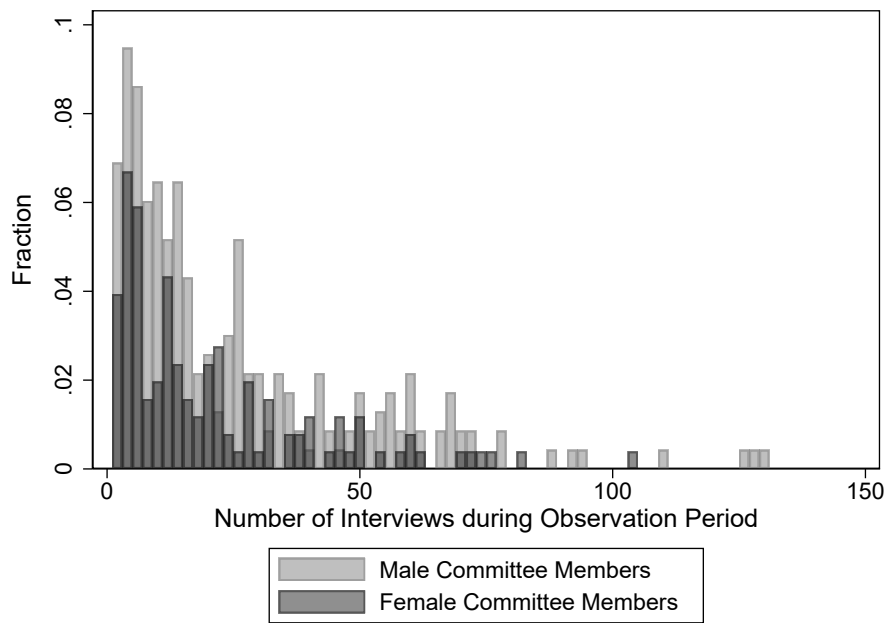
- Butler, D. L., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: Implications for leadership evaluations. *Journal of Personality and Social Psychology*, 58(1), 48–59.
- Card, D., & DellaVigna, S. (2020). What do editors maximize? Evidence from four economics journals. *Review of Economics and Statistics*, 102(1), 195–217.
- Card, D., DellaVigna, S., Funk, P., & Iriberri, N. (2020). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics*, 135(1), 269–327.
- Carli, L. L. (1989). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology*, 56(4), 565.
- Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, 57(4), 725–741.
- Carter, A. J., Croft, A., Lukas, D., & Sandstrom, G. M. (2018). Women’s visibility in academic seminars: Women ask fewer questions than men. *PloS one*, 13(9), e0202743.
- Chan, D. C. (2021). Influence and information in team decisions: Evidence from medical residency. *American Economic Journal: Economic Policy*, 13(1), 106–137.
- Coffman, K. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129 (4), 1625–1660.
- Coffman, K., Flikkema, C. B., & Shurchkov, O. (2021). Gender stereotypes in deliberation and team decisions. *Games and Economic Behavior*, 129, 329–349.
- Coughlan, P. J. (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *American Political Science Review*, 94(2), 375–393.
- Deschamps, P. (2024). Gender quotas in hiring committees: A boon or a bane for women? *Management Science*, 70(11), 7486–7505.
- Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3), 1345–1381.
- Gerardi, D., & Yariv, L. (2007). Deliberative voting. *Journal of Economic theory*, 134(1), 317–338.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118 (3), 1049–1074.
- Goeree, J. K., & Yariv, L. (2011). An experimental study of collective deliberation. *Econometrica*, 79(3), 893–921.
- Guo, J., & Recalde, M. P. (2023). Overriding in teams: The role of beliefs, social image, and gender. *Management Science*, 69(4), 2239–2262.
- Hao, L., & Suen, W. (2009). Viewpoint: Decision-making in committees. *Canadian Journal of Economics/Revue canadienne d’économique*, 42(2), 359–392.

- Hardt, D., Mayer, L., & Rincke, J. (2025). Who does the talking here? The impact of gender composition on team interactions. *Management Science*, 71(5), 4131–4152.
- Iaryczower, M., Shi, X., & Shum, M. (2018). Can words get in the way? The effect of deliberation in collective decision making. *Journal of Political Economy*, 126(2), 688–734.
- Jacobi, T., & Schweers, D. (2017). Justice, interrupted: The effect of gender, ideology, and seniority at supreme court oral arguments. *Va. L. Rev.*, 103, 1379.
- Karpowitz, C., & Mendelberg, T. (2014). The silent sex: Gender, deliberation, and institutions. In *The silent sex: Gender, deliberation, and institutions*. Princeton University Press.
- Karpowitz, C., O’Connell, S. D., Preece, J., & Stoddard, O. (2024). Strength in numbers? Gender composition, leadership, and women’s influence in teams. *Journal of Political Economy*, 132(9), 3077–3114.
- Koffi, M. (2025). Innovative ideas and gender (in)equality. *American Economic Review*, 115(7), 2207–2236.
- Kunze, A., & Miller, A. R. (2017). Women helping women? evidence from private sector data on workplace hierarchies. *The Review of Economics and Statistics*, 99(5), pp. 769–775.
- Maida, A., & Weber, A. (2022). Female leadership and gender gap within firms: Evidence from an Italian board reform. *ILR Review*, 75(2), 488–515.
- Morgan, J., & Stocken, P. C. (2008). Information aggregation in polls. *American Economic Review*, 98(3), 264–96.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *Quarterly Journal of Economics*, 122(3), 1067–1101.
- Radbruch, J., & Schiprowski, A. (2024). Interview sequences and the formation of subjective assessments. *Review of Economic Studies*, 92(2), 1226–1256.
- Ronchi, M., & Salvestrini, V. (2026). Gender diversity and decision-making in teams. *mimeo*.
- Roussille, N. (2024). The role of the ask gap in gender pay inequality. *Quarterly Journal of Economics*, 139(3), 1557–1610.
- Sarsons, H., Gërkhani, K., Reuben, E., & Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, 129(1), 101–147.
- 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.

Appendix (For Online Publication)

A Additional Material: Institutional Setting

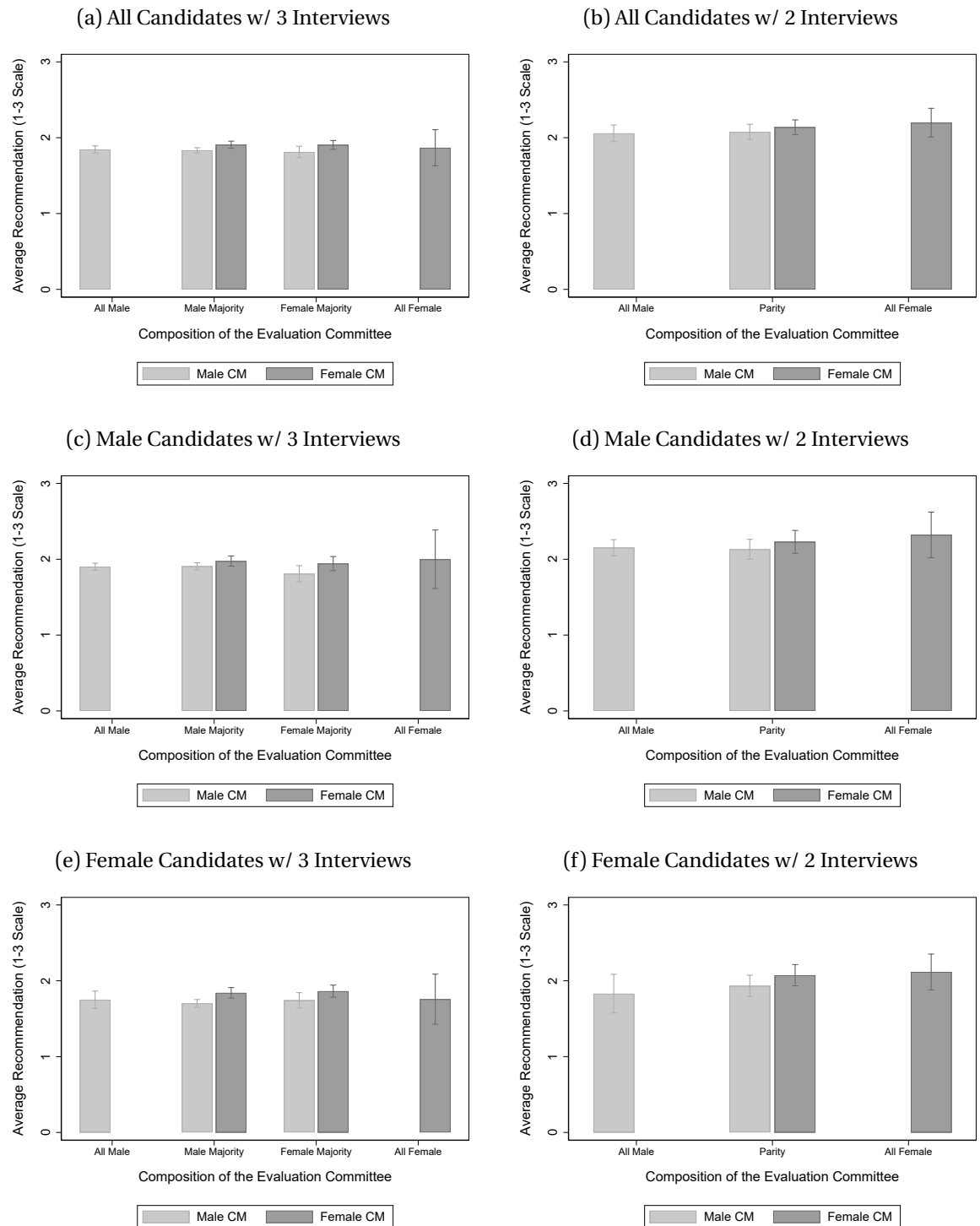
Figure A.1: Number of Interviews per Committee Member



Note: The unit of observation is the committee member (N=359). The average number of interviews is 25 (20) for male (female) committee members.

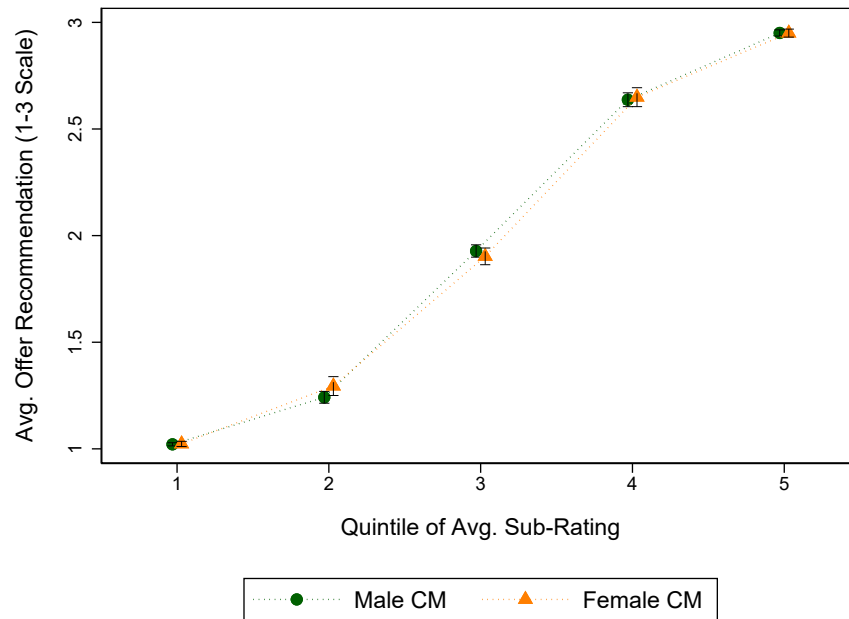
B Additional Material: Data

Figure B.1: Average Recommendations and Composition of the Evaluation Committee



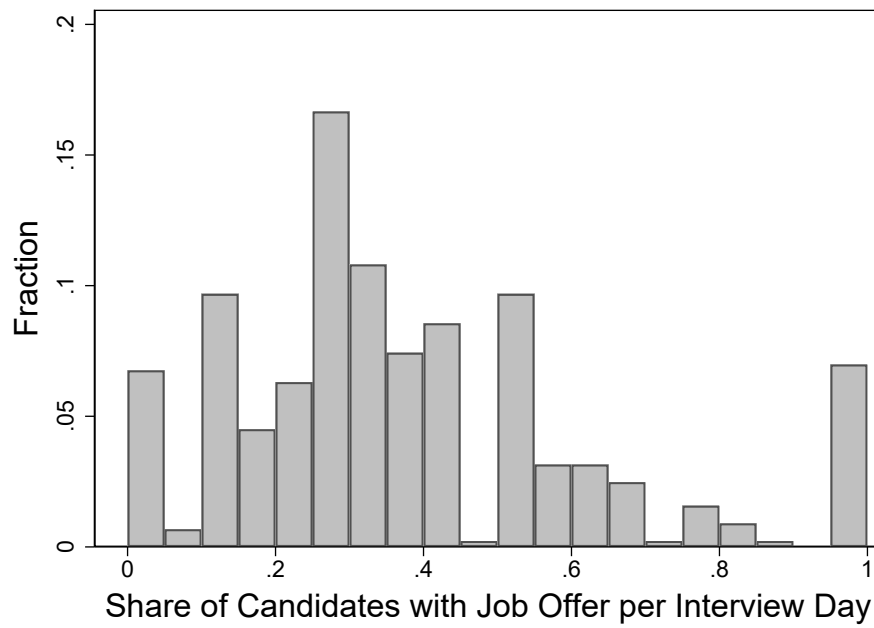
Note: The unit of observation is the interview. Male Majority= candidate is evaluated by 1 female & 2 male committee members. Female Majority= candidate is evaluated by 1 male & 2 female committee members. 95% confidence intervals, with standard errors clustered at the level of the hiring committee (N=444).

Figure B.2: Translation of Interview Sub-Ratings into Recommendations



Note: CM = committee member. The x-axis denotes quintiles of the average sub-rating. The y-axis denotes the average offer recommendation on a 1–3 scale. N=8,322. 95% confidence intervals, with standard errors clustered at the level of the hiring committee (N=444).

Figure B.3: Share of Candidates with a Job Offer per Interview Day



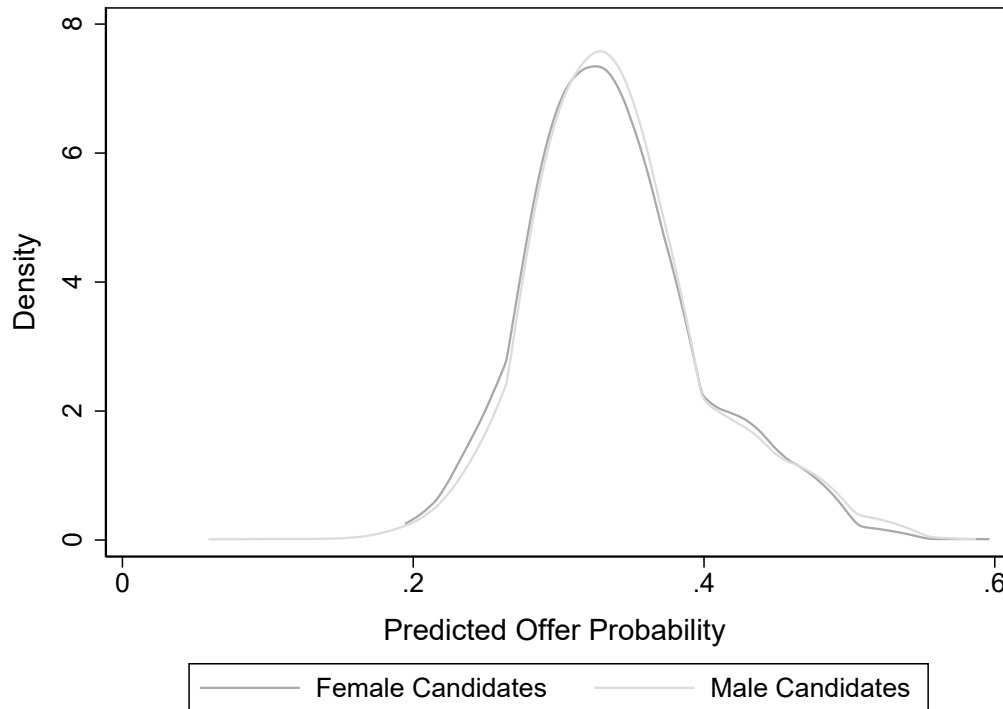
Note: The unit of observation is the interview day/hiring committee (N=444). The average share is 0.37, with a standard deviation of 0.25.

Table B.1: Summary Statistics on Candidate Characteristics

	All Candidates		Male Candidates			Female Candidates			Difference	
	Mean	SD	Mean	SD	N	Mean	SD	N	Diff.	p-value
Internship Application	0.33	0.47	0.33	0.47	1800	0.33	0.47	1185	-0.009	0.60
Study Field: Economics/Business	0.59	0.49	0.57	0.49	1800	0.61	0.49	1185	-0.034	0.06
Study Field: Engineering/IT	0.11	0.32	0.14	0.34	1800	0.08	0.28	1185	0.052	0.00
Study Field: Maths/Physical Sciences/Chemistry	0.07	0.26	0.08	0.27	1800	0.06	0.24	1185	0.019	0.04
Study Field: Life Sciences	0.06	0.24	0.05	0.21	1800	0.08	0.27	1185	-0.034	0.00
Study Field: Law	0.03	0.16	0.03	0.17	1800	0.02	0.15	1185	0.005	0.37
Study Field: Social Sciences/Humanities	0.05	0.22	0.04	0.19	1800	0.07	0.26	1185	-0.035	0.00
Study Field: Unknown/Others	0.09	0.28	0.10	0.30	1800	0.07	0.26	1185	0.026	0.01
High School GPA	1.38	0.32	1.39	0.33	1800	1.36	0.30	1185	0.031	0.01
High School GPA: Top Grade	0.14	0.35	0.14	0.34	1800	0.14	0.35	1185	-0.009	0.50
High School Math Grade	1.46	0.28	1.45	0.28	1800	1.48	0.28	1185	-0.025	0.02
CV Score	2.00	0.21	2.02	0.21	1800	1.98	0.21	1185	0.042	0.00

Note: The unit of observation is the candidate. High-school GPA and mathematics grades are reported on the German grade scale from 1.0 (excellent) to 4.0 (satisfactory). The indicator “High School GPA: Top Grade” equals one if the candidate achieved the best possible GPA score (1.0 on the German scale). The CV score is an average of three CV assessments made by the HR department prior to the interview day on a 1 (low) to 3 (high) scale.

Figure B.4: Distribution of Predicted Offer Probability by Candidate Gender



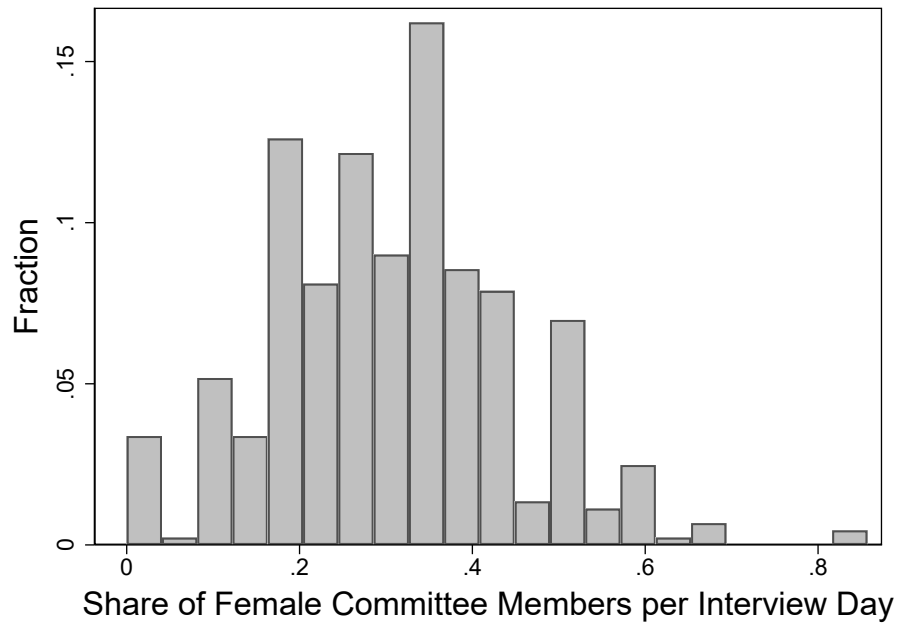
Note: The unit of observation is the candidate ($N=1,800$ for male candidates and $N=1,185$ for female candidates). The figure plots the distribution of candidates' (out-of-sample) predicted offer probabilities, estimated from high-school GPA, mathematics grade, and CV score (see Section 3.1.3 for details on the prediction). The average predicted offer probability is 0.34, with a standard deviation of 0.06.

Table B.2: Share of Committee Members' Job Titles

	No management resp.				With management resp.			<i>N</i>
	1	2	3	4	5	6	7	
Male CM	18.9	1.4	0.6	27.1	15.0	33.6	3.4	354
Female CM	20.4	0.0	1.1	30.1	15.1	31.7	1.6	186
All	19.4	0.9	0.7	28.1	15.0	33.0	2.8	540
<i>p</i> -value	0.675	0.103	0.511	0.463	0.980	0.656	0.232	
Joint $\chi^2(6) = 5.10$, <i>p</i> -value = 0.531								

Note: The table reports the share (in percent) of committee members' anonymized job titles, separately for male and female committee members, with one observation per unique committee member–job–title pair. Job titles with ID 1–4 carry no managerial responsibility and job titles with ID 5–7 carry managerial responsibility. Within the managerial and non-managerial categories, job IDs do not imply a seniority ranking. The row labeled “*p*-value” reports, for each job title, the *p*-value of a χ^2 -test of whether the share in that title differs between male and female committee members. The bottom row reports the overall χ^2 -test of equality of the job-title distribution across committee-member gender.

Figure B.5: Share of Female Committee Members per Interview Day



Note: The unit of observation is the interview day/hiring committee (N=444). The average share is 0.30, with a standard deviation of 0.14.

Table B.3: Balance of Candidate Characteristics Across Male and Female Committee Members

	P(Interview w/ Female CM)		Number of Female CM per Cand.	
	(1) Female Cand.	(2) Male Cand.	(3) Female Cand.	(4) Male Cand.
Internship Application	0.012 (0.020)	-0.003 (0.020)	0.031 (0.055)	-0.014 (0.057)
CV Score (std.)	-0.002 (0.007)	0.002 (0.005)	-0.009 (0.020)	0.004 (0.014)
High School GPA (Std.)	-0.006 (0.008)	0.011 (0.007)	-0.016 (0.023)	0.032 (0.020)
High School GPA: Top Grade	0.024 (0.020)	-0.006 (0.020)	0.068 (0.056)	-0.008 (0.055)
High School Math Grade (Std.)	0.013 (0.008)	-0.004 (0.006)	0.033 (0.023)	-0.012 (0.018)
Study Field: Engineering/IT	-0.030 (0.024)	-0.008 (0.017)	-0.076 (0.067)	-0.023 (0.048)
Study Field: Maths/Physical Sciences/Chemistry	0.028 (0.028)	0.010 (0.021)	0.071 (0.075)	0.028 (0.057)
Study Field: Life Sciences	-0.015 (0.025)	-0.009 (0.022)	-0.046 (0.073)	-0.015 (0.065)
Study Field: Law	-0.001 (0.039)	0.029 (0.028)	0.005 (0.109)	0.089 (0.076)
Study Field: Social Sciences/Humanities	0.026 (0.024)	0.006 (0.032)	0.075 (0.064)	0.025 (0.090)
Study Field: Unknown/Others	-0.025 (0.027)	-0.027 (0.034)	-0.080 (0.080)	-0.076 (0.099)
Hiring Committee FE	Yes	Yes	Yes	Yes
p-value joint significance (excl. gender)	0.49	0.87	0.45	0.84
Outcome Mean	0.41	0.25	1.14	0.69
N	3292	5030	1185	1800

Note: In columns (1) and (2), the dependent variable is an indicator for whether the interview is conducted by a female committee member (CM), defined at the interview level. In columns (3) and (4), the dependent variable is the number of female committee members assigned to the candidate, defined at the candidate level. The omitted study field is economics/business. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Assignment of Non-Cognitive Skill Sets to Male vs. Female Committee Members

	P(Skill Set 1)		P(Skill Set 2)		P(Skill Set 3)	
	(1)	(2)	(3)	(4)	(5)	(6)
Female CM	0.008 (0.006)	0.003 (0.006)	-0.011* (0.006)	-0.001 (0.005)	0.003 (0.006)	-0.002 (0.006)
Hiring Committee FE	No	Yes	No	Yes	No	Yes
Outcome Mean	0.35	0.35	0.30	0.30	0.35	0.35
N	8322	8322	8322	8322	8322	8322

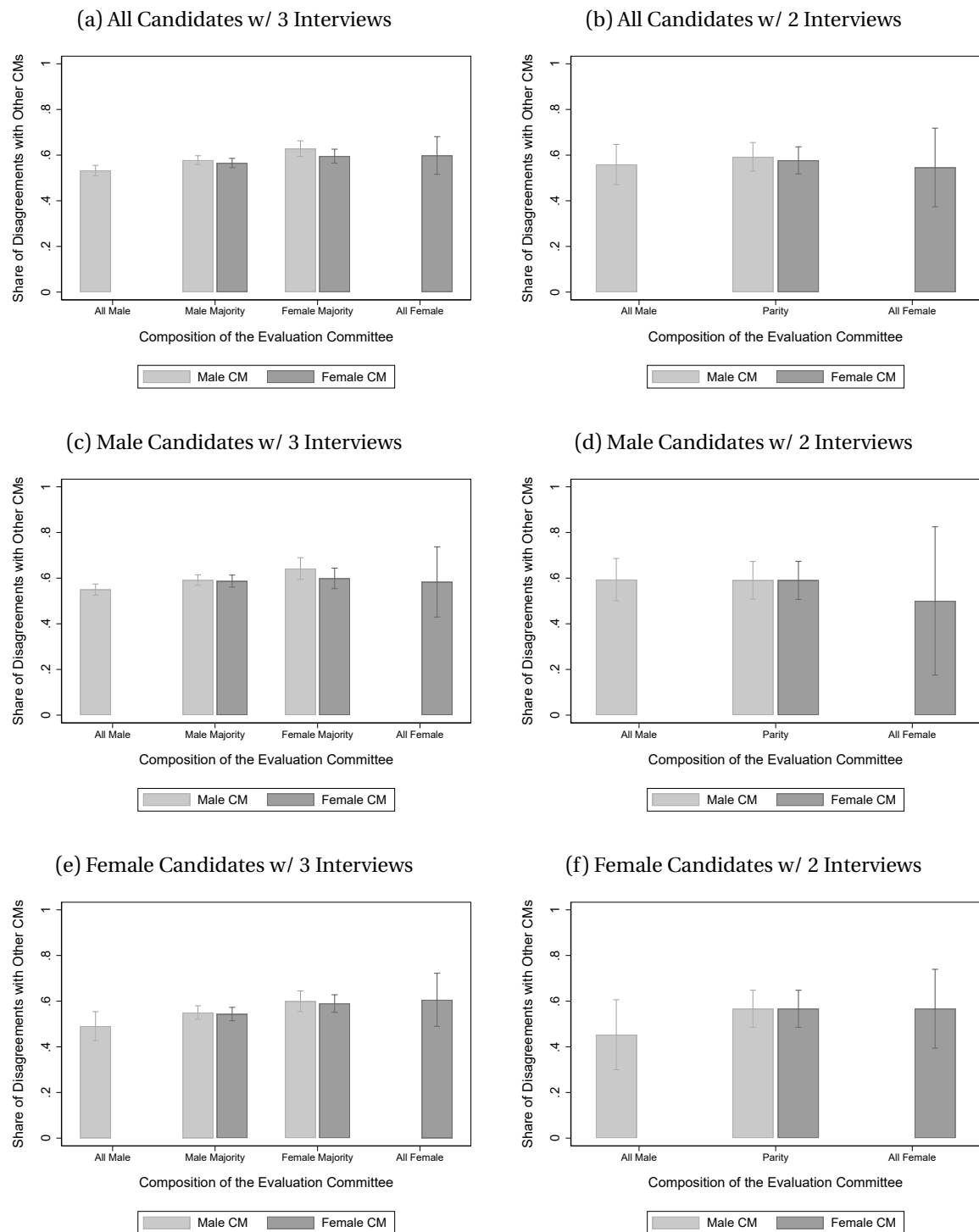
Note: CM = committee member. The unit of observation is the interview. The dependent variable is an indicator of whether one of three different non-cognitive skill sets was assessed in an interview. The content of the skill sets cannot be named due to confidentiality. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Relationship between Candidate Gender and Committee Member Characteristics

	P(Female Candidate)	
	(1)	(2)
Experience (Std.)	0.007 (0.006)	0.001 (0.006)
Management Responsibility	-0.055*** (0.016)	-0.005 (0.014)
On Leave	-0.000 (0.014)	0.011 (0.013)
Int Day at Main Location	0.007 (0.014)	-0.011 (0.012)
Hiring Committee FE	No	Yes
p-value joint significance	0.00	0.58
Outcome Mean	0.40	0.40
N	8322	8322

Note: The unit of observation is the interview. The outcome is the probability that the interviewed candidate is female. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

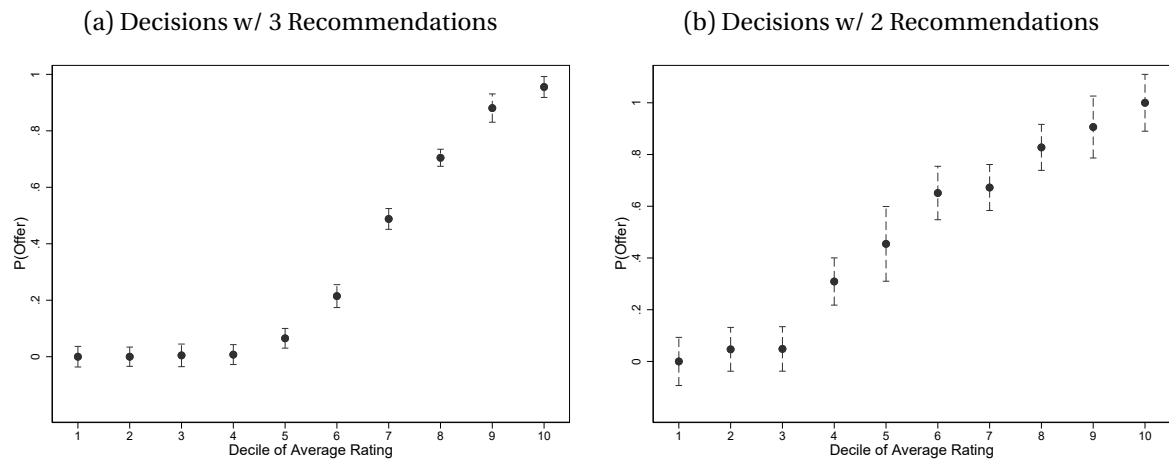
Figure B.6: Disagreement and Composition of the Evaluation Committee



Note: The unit of observation is the interview. Male Majority= candidate is evaluated by 1 female & 2 male committee members. Female Majority= candidate is evaluated by 1 male & 2 female committee members. “Share of disagreements” is the share of other committee members giving a different recommendation on the same candidate. 95% confidence intervals, with standard errors clustered at the level of the hiring committee (N=444).

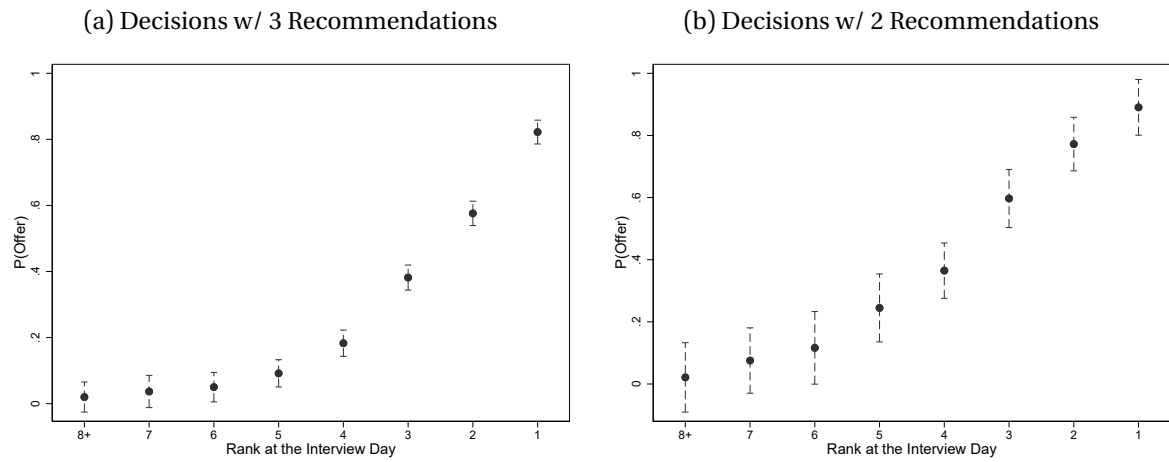
C Additional Material: Aggregation of Information

Figure C.1: Offer Probability Depending on Average Interview Rating



Note: The unit of observation is the candidate. N=2,500 in panel (a), N=485 in panel (b). The x-axis shows deciles of the candidate's average rating from all interviews. The y-axis shows the probability that the committee decides to make an offer. Dashed lines show 95% confidence intervals.

Figure C.2: Offer Probability Depending on Candidate's Rank at the Interview Day



Note: The unit of observation is the candidate. N=2,500 in panel (a), N=485 in panel (b). The x-axis shows the rank of the candidate relative to the other candidates present at the interview day (based on average pre-deliberation interview ratings). The left y-axis shows the probability that the committee decides to make an offer. Dashed lines show 95% confidence intervals.

Table C.1: Variation in Offers Explained by Recommendations, Candidate Characteristics and Ratings

	All Cases				Deliberation Cases			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R-Squared	0.60	0.60	0.60	0.63	0.16	0.19	0.20	0.28
Hiring Committee FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Candidate Characteristics	No	No	Yes	Yes	No	No	Yes	Yes
Sub-Ratings	No	No	No	Yes	No	No	No	Yes
N	2985	2985	2985	2985	1045	1045	1045	1045

Note: The dependent variable is an indicator equal to one if the committee made a job offer. The table reports R-squared values from OLS regressions on indicators for the combination of initial recommendations (bins in Figure 3). We subsequently add hiring committee fixed effects (column 2), candidate characteristics (column 3), and the candidate's average cognitive and non-cognitive rating as well as their interaction (column 4). Columns (5)–(8) replicate columns (1)–(4) for the subsample of deliberation cases.

D Additional Material: Gender Differences in Influence

Table D.1: Gender Differences in the Translation of Recommendations & Ratings into Committee Decisions

	P(Offer)			
	(1)	(2)	(3)	(4)
Female	0.007 (0.017)			
Recommendation (scale 1-3)	0.296*** (0.006)	0.290*** (0.006)	0.184*** (0.008)	
Recommendation x Female CM	-0.024** (0.011)	-0.026** (0.011)	-0.023** (0.010)	
Avg. Cognitive Rating			0.024*** (0.006)	0.117*** (0.006)
Cognitive Rating (Std.) x Female CM				0.004 (0.010)
Avg. Non-Cognitive Rating			0.015*** (0.005)	0.087*** (0.005)
Non-Cognitive Rating (Std.) x Female CM				-0.022** (0.009)
Committee Member FE	No	Yes	Yes	Yes
Candidate Characteristics	No	Yes	Yes	Yes
Co-Evaluator Sub-Ratings	No	No	Yes	Yes
Outcome Mean	0.33	0.33	0.33	0.33
N	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the candidate receives an offer. All regressions are estimated using OLS. Candidate characteristics include all variables reported in Appendix Table B.1 and indicators for the set of non-cognitive skills assessed in the interview. Co-evaluator sub-ratings include the average cognitive and non-cognitive ratings given by the other evaluators to the same candidate. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: Alignment of Recommendations with Decisions: Role of Other CM Characteristics

	Alignment Probability					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.051*** (0.011)	-0.053*** (0.012)	-0.047*** (0.012)	-0.035*** (0.011)	-0.028*** (0.008)	-0.005 (0.006)
Female x Delib. Case						-0.066*** (0.020)
Experience (Std.)		0.030*** (0.006)	0.030*** (0.006)	0.029*** (0.005)	0.012*** (0.004)	0.002 (0.004)
Experience (Std.) x Delib. Case						0.026*** (0.010)
Manager		0.005 (0.016)	0.012 (0.016)	0.018 (0.015)	-0.009 (0.011)	0.006 (0.007)
Manager x Delib. Case						-0.040 (0.027)
Int Day at Main Location		-0.004 (0.013)	-0.006 (0.013)	-0.013 (0.012)	0.004 (0.008)	0.005 (0.006)
Int Day at Main Location x Delib. Case						-0.002 (0.021)
On Leave		0.041*** (0.015)	0.039*** (0.015)	0.040*** (0.013)	0.011 (0.009)	0.005 (0.007)
On Leave x Delib. Case						0.019 (0.025)
Hiring Committee FE	No	Yes	Yes	Yes	Yes	Yes
Candidate & CM Characteristics	No	Yes	Yes	Yes	Yes	Yes
Leniency	No	No	Yes	Yes	Yes	Yes
Case Type	No	No	No	Yes	Yes	Yes
Case Type x Disagreement	No	No	No	No	Yes	Yes
R-Squared	0.00	0.01	0.01	0.22	0.57	0.57
Outcome Mean	0.68	0.68	0.68	0.68	0.68	0.68
N	8322	8322	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated by OLS based on equation (1), with varying sets of control variables as indicated at the bottom of the table. The controls for candidate and committee member (CM) characteristics include all variables reported in Table 2 and Appendix Table B.1, as well as indicators for the set of non-cognitive skills assessed in the interview. The leniency control is the committee member's leave-one-out mean recommendation on the 1–3 scale. Case type controls include indicators for each combination of initial recommendations (the bins shown in Figure 3). Disagreement is measured as the average pairwise absolute difference between a committee member's own recommendation and the recommendations of the other two members for a given candidate. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Alignment under Deliberation vs. Hypothetical Cutoff Rule

	Alignment Probability under Different Decision Rules					
	Deliberation		Cutoff: Avg. Ratings >2		Cutoff: Avg. Recomm. ≥ 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.028*** (0.008)	-0.005 (0.006)	-0.006 (0.008)	-0.004 (0.006)	0.004 (0.007)	-0.005 (0.004)
Female x Delib. Case		-0.066*** (0.020)		-0.008 (0.018)		0.024 (0.019)
R-Squared	0.57	0.57	0.57	0.57	0.60	0.60
Outcome Mean	0.68	0.68	0.72	0.72	0.71	0.71
N	8322	8322	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision (actual or hypothetical) aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1), including the full set of controls. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). Columns (1) and (2) replicate columns (5) and (6) of Table 6. In columns (3) to (6), we assign hypothetical offers to candidates whose average sub-rating is above 2 points (columns 3 to 4) or whose average recommendation is at least a 'potential offer' (columns 5 to 6). We then compute the alignment of recommendations with the hypothetical offer decision. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: Heterogeneity by Committee Member and Candidate Characteristics

	Alignment Probability				
	(1)	(2)	(3)	(4)	(5)
Female	-0.061*** (0.014)	-0.053*** (0.012)	-0.080*** (0.023)	-0.031* (0.016)	-0.059*** (0.015)
Female × Manager	0.023 (0.024)				
Female × Experience (Std.)		-0.011 (0.012)			
Female × Int Day at Main Location			0.038 (0.027)		
Female × Female Cand.				-0.046** (0.023)	
Female × Internship Application					0.020 (0.023)
Outcome Mean	0.68	0.68	0.68	0.68	0.68
N	8322	8322	8322	8322	8322

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1), including hiring committee fixed effects and controls for candidate as well as committee member characteristics. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: Robustness of Main Estimates: Alternative Control Variables

	Alignment Probability	
	(1)	(2)
<i>Panel A: Baseline results from Table 6</i>		
Female	-0.028*** (0.008)	-0.005 (0.006)
Female x Delib. Case		-0.066*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel B: Baseline + evaluation committee (i.e., candidate) FE</i>		
Female	-0.023*** (0.008)	-0.006 (0.005)
Female x Delib. Case		-0.052** (0.022)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel C: Baseline + candidate char. interacted with leniency & case type</i>		
Female	-0.027*** (0.008)	-0.004 (0.006)
Female x Delib. Case		-0.069*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel D: Baseline + job title indicators</i>		
Female	-0.028*** (0.008)	-0.006 (0.006)
Female x Delib. Case		-0.066*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel E: Baseline + quintiles of cognitive & non-cognitive sub-ratings</i>		
Female	-0.023*** (0.008)	-0.002 (0.006)
Female x Delib. Case		-0.062*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel F: Baseline + gender composition of other committee members</i>		
Female	-0.032*** (0.010)	-0.010 (0.008)
Female x Delib. Case		-0.066*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322

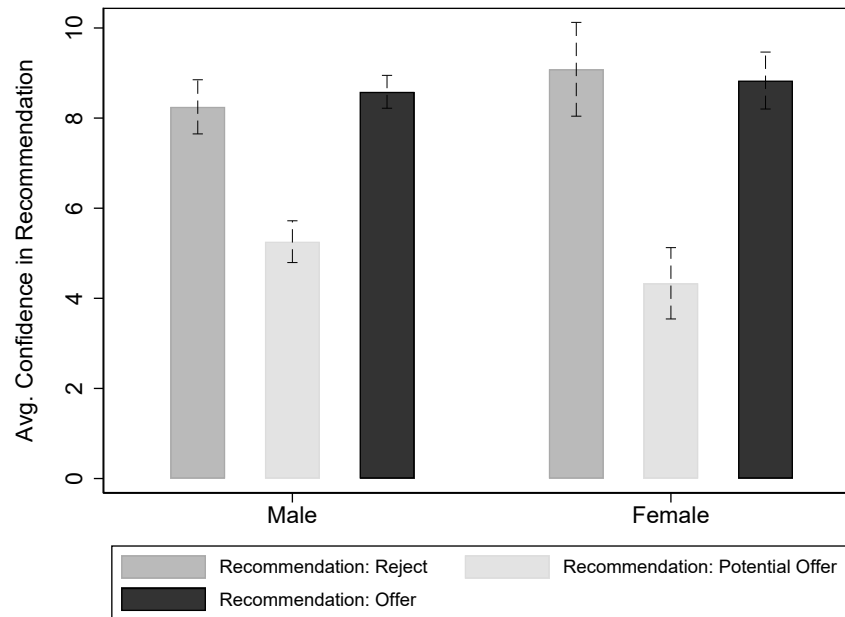
Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1). Regressions include hiring committee fixed effects and the full set of control variables corresponding to columns (5) and (6) of Table 6. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). In Panel D, the job level indicators correspond to those reported in Table B.2. In Panel E, controls include indicators for quintiles of the cognitive and non-cognitive skill ratings, as well as their interactions. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Robustness of Main Estimates: Alternative Samples

	Alignment Probability	
	(1)	(2)
<i>Panel A: Baseline results from Table 6</i>		
Female	-0.028*** (0.008)	-0.005 (0.006)
Female x Delib. Case		-0.066*** (0.020)
Outcome Mean	0.68	0.68
N	8322	8322
<i>Panel B: Excluding interviews with potential offer recommendations</i>		
Female	-0.020** (0.010)	-0.001 (0.005)
Female x Delib. Case		-0.066** (0.028)
Outcome Mean	0.81	0.81
N	6063	6063
<i>Panel C: Excluding candidates with final interviews</i>		
Female	-0.016** (0.007)	0.000 (0.005)
Female x Delib. Case		-0.063*** (0.023)
Outcome Mean	0.69	0.69
N	7027	7027
<i>Panel D: Excluding candidates with two interviews</i>		
Female	-0.024*** (0.008)	-0.002 (0.005)
Female x Delib. Case		-0.070*** (0.023)
Outcome Mean	0.68	0.68
N	7354	7354
<i>Panel E: Excluding interview days with one female CM</i>		
Female	-0.025*** (0.009)	-0.006 (0.006)
Female x Delib. Case		-0.054*** (0.021)
Outcome Mean	0.68	0.68
N	7393	7393

Note: The dependent variable is an indicator equal to one if the committee's final hiring decision aligns with the committee member's initial recommendation. All regressions are estimated using OLS based on equation (1). Regressions include hiring committee fixed effects and the full set of control variables corresponding to columns (5) and (6) of Table 6. 'Delib. Case' is an indicator for deliberation cases, describing candidates whose initial recommendations leave scope for deliberation to determine the hiring decision (see Section 4). Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure D.1: Committee Members' Confidence in Recommendations



Note: The figure shows responses to questions on committee members' average confidence in their recommendations ('When recommending a rejection/potential offer/offer, how confident are you on average that the candidate should (not) be hired?'). Confidence is expressed on a 0–10 scale. The questions were part of a survey on committee members' hiring experiences conducted in 2021. The survey includes 48 committee members (36 M, 12 F). The participation rate among committee members who were involved in hiring activities during the survey period was 22%. Dashed lines indicate 95% confidence intervals.

E Additional Material: Committee Composition and the Gender Balance of New Hires

Table E.1: Gender Evaluation Gap Among Male and Female Committee Members: Placebo Check

	Predicted Offer Probability			
	(1)	(2)	(3)	(4)
Female Candidate	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Female CM		0.002 (0.002)	0.000 (0.001)	0.000 (0.001)
Female Candidate x Female CM		-0.001 (0.002)	-0.000 (0.002)	-0.000 (0.002)
Hiring Committee FE	No	No	Yes	Yes
CM Characteristics	No	No	No	Yes
Outcome Mean	0.34	0.34	0.34	0.34
N	8322	8322	8322	8322

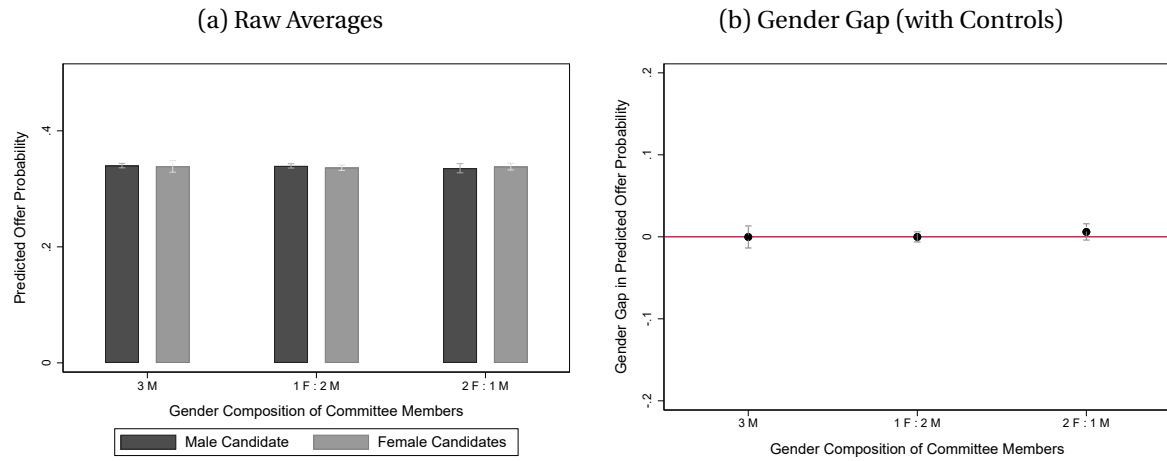
Note: The dependent variable is the candidate's (out-of-sample) predicted offer probability, estimated from high-school GPA, mathematics grade, and CV score (see Section 3.1.3 for details). CM = committee member. Controls include the committee member characteristics reported in Table 2. Candidate characteristics are not included because they are part of the outcome variable. All regressions are estimated using OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2: Translation of Positive Recommendations into Offers

	P(Offer)					
	All		Female Candidates		Male Candidates	
	(1)	(2)	(3)	(4)	(5)	(6)
Female CM	-0.004 (0.011)	0.002 (0.011)	0.022 (0.015)	0.016 (0.015)	-0.025* (0.015)	-0.013 (0.017)
Positive Rec.	0.440*** (0.012)	0.376*** (0.012)	0.446*** (0.019)	0.330*** (0.019)	0.433*** (0.015)	0.350*** (0.015)
Positive Rec. x Female CM	-0.053*** (0.019)	-0.051*** (0.018)	-0.083*** (0.028)	-0.060** (0.026)	-0.021 (0.025)	-0.026 (0.025)
Hiring Committee FE	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Outcome Mean	0.33	0.33	0.30	0.30	0.36	0.36
N	8322	8322	3292	3292	5030	5030

Note: The dependent variable is an indicator equal to one if the committee decided to extend a job offer to the candidate. CM = committee member. Control variables include committee member and candidate characteristics as reported in Table 2 and Appendix Table B.1, as well as indicators for the set of non-cognitive skills assessed in the interview. All regressions are estimated by OLS. Standard errors, clustered at the level of the hiring committee (N=444), are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure E.1: Placebo Check: Gender Composition of the Committee and Gender Balance in Hiring Decisions



Note: The unit of observation is the candidate. The figure replicates Figure 6, using a candidate's predicted offer probability (constructed as described in Section 3.1.3) as a placebo outcome. In panel (b), regressions control for hiring committee fixed effects. Candidate characteristics are not included because they are part of the outcome variable. 95% confidence intervals, with standard errors clustered at the level of the hiring committee (N=444).