# Committee Deliberation and Gender Differences in Influence

Jonas Radbruch[†] Amelie Schiprowski[§]

*May 24, 2023*

**Abstract**

This paper provides empirical evidence on the aggregation of information in committees. We analyze unique data from the decision-making process of hiring committees within a large private company. In the hiring process, committee members first conduct independent one-to-one interviews and give individual recommendations before deliberating on a collective hiring decision. We find that committees' final hiring decisions are systematically less aligned with the initial recommendations of women than with those of men, even though women and men are equally qualified and experienced. This disparity in influence is strongest when recommendations exhibit high disagreement and when a single woman deliberates with two men. The estimated distribution of influence reveals that almost all men are more influential than the median woman. We offer suggestive evidence that these findings have implications for the effectiveness of gender quotas.

**Keywords:** Committee Decision-Making, Gender Differences, 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 Szajbély provided excellent research assistance. Support by the German Research Foundation (DFG) through EXC 2126/1– 390838866, CRC TR 224 (Project A05, Schiprowski) and CRC TR 190 (Project Z, 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. Disagreement is inherent to this process, as individuals can hold conflicting views. Therefore, most committees deliberate before agreeing or voting on a final decision. During deliberation, the behavior of committee members can vary along various dimensions. For example, individuals may differ in their propensity to contribute their opinions and persuade others, in their reaction to others' information, and in how they handle conflicts. As deliberation is often a critical part of the decision-making process, such differences in behavior can lead to differences in influence.

In this paper, we use novel data on private-sector hiring committees to study the aggregation of individual information into collective decisions. We place a particular focus on identifying gender differences in individuals' influence, which are relevant for two main reasons. First, several experimental findings show that women and men differ in behaviors that matter for committee deliberation, such as the contribution of ideas or the choice to take leadership in group decision-making (e.g., Coffman, 2014; Born et al., 2022). Second, there have been controversial policy discussions regarding the gender diversity of high-level decision-making bodies, resulting in multiple efforts to achieve equal representation through gender quotas (e.g., Azmat & Boring, 2020). While quotas are a direct way to increase the share of women, their effectiveness ultimately hinges on the distribution of influence within the committee.

In the hiring process that we study, we can trace how a committee member's initial assessment of a candidate translates into the committee's final decision. Every candidate is first assessed independently by three different committee members, based on one-to-one interviews. After all interviews have been conducted, committee members reveal their individual recommendations on whether to hire the candidate in a committee meeting. Given that the three recommendations are independently made under different circumstances, they frequently exhibit disagreement. A joint decision is then reached through deliberation, without a fixed cutoff or voting rule. Exploiting the information on initial recommendations and the final decisions, we study three main questions: First, how do hiring committees aggregate in-

dividual recommendations? Second, does the translation of individual recommendations into decisions differ systematically between men and women in the committee? Third, do gender differences in influence among committee members affect the gender balance of new hires?

We first provide evidence on the aggregation of individual recommendations into hiring decisions. As expected, we observe that committees are more likely to make job offers when candidates have received more positive recommendations. However, decisions cannot be summarized by a deterministic cut-off or majority voting rule. In particular, about 35% of the cases leave substantial leeway to the committee, given their combination of initial recommendations. In line with the decision-level evidence, we observe that individual recommendations are less aligned with committee decisions when disagreement –and thus the scope for deliberation– is large.

We then study how the translation of individual recommendations into committee decisions differs between male and female committee members. A first descriptive analysis suggests systematic differences: positive recommendations by female members are less likely to result in a job offer than those of male members, and vice versa. These patterns arise despite the fact that men and women have on average the same level of experience and seniority, differ little in their evaluation behavior, and do not disagree more with their fellow committee members.

To analyze the observed gender differences in alignment more formally, we construct a binary indicator measuring whether an initial recommendation aligns with the committee's final decision. Exploiting the quasi-random assignment of committee members to candidates, we estimate how the probability of alignment varies between men and women. The raw gender gap in alignment amounts to about 4.9 percentage points (8% relative to the male average). In the full specification – which controls for candidate and committee member characteristics, leniency, and the extent of disagreement – it decreases only slightly to about 4.2 percentage points. The gender gap is sizable compared to other determinants of alignment. For example, it corresponds to the average effect of a 1.7 standard deviation ($\approx$ 3 years) change in hiring experience. Focusing on cases with greater scope for deliberation, the gender gap widens to about 8 percentage points (15%).

In the next step, we study the role of the committee's gender composition. In particular, we seek to understand how the gender gap in recommendation alignment interacts with the minority status of women in the committees. In our setup, the discussion on a given candidate mostly takes place among the three committee members who interviewed that candidate. We exploit variation in the gender composition of these three committee members to compare the gender gap in alignment under a male versus a female majority. We find strong evidence that candidates who are discussed by two men and one woman – which happens to be the most frequent case – are driving the gender gap in alignment. This also holds when controlling for the extent of disagreement. In contrast, no gender gap in either direction emerges when cases are discussed by two women and one man.

Turning from the institutional to the individual determinants of influence, we exploit the fact that we observe the same individual in several committees to understand individual-specific heterogeneity in influence. We estimate key statistics of the distribution of committee members' influence, finding significant dispersion in individuals' average influence among both men and women. However, compared to the dispersion within gender, the gap between men and women is large. For instance, the average gender gap in influence is similar in magnitude to the interquartile range of influence among men. Additionally, the results suggest that the average gender gap is mainly driven by the lower part of the distribution, with almost all men being more influential than the median woman on the committee. Nonetheless, due to the dispersion, there are also some women who are quite influential compared to most men.

We next discuss potential mechanisms that could explain our results. We do not find any evidence that women are less influential because they produce less informative evaluations based on their interviews. Moreover, the influence gap does not appear to be a simple artifact of women's minority status, given that men do not have a lower influence when being in the minority. Instead, the discussion suggests that behavioral differences in the deliberation process lead to the gender gap in influence. According to previous experimental findings, women and men can differ in how they communicate their information and respond to others' information. Given our real-world data, we are unable to discriminate across the exact dimensions of deliberation behavior. However, regardless of the exact mechanism, the find-

ings overall suggest that behavioral factors which are presumably not related to the quality of information influence how the information is weighted in the committee.

Finally, we examine the potential implications of our findings for the gender balance of new hires. A common objective of female representation in hiring committees is to improve the chances of female candidates. In our setup, female candidates at the interview stage are on average 17% less likely to be hired than their male counterparts. Whether the presence of women in committees can alter the gender balance of new hires depends both on the members' individual evaluation behavior and on the aggregation of evaluations at the committee level. We use our setup to provide evidence on these two levels. At the individual level, the data support the presumption that women are significantly more favorable towards female candidates. More precisely, the evaluations of female committee members exhibit on average a 50% smaller gender gap than those of their male colleagues. However, the committee's hiring behavior does not change in favor of female candidates when one man is replaced by a woman. Only when women become the 2:1 majority, decisions change drastically, and the gender gap in hiring outcomes even closes. This strong non-linearity suggests that the committee's aggregation process plays a crucial role for the impact of committee diversification on gender equality in outcomes.

The results in this paper document a systematic gender gap in influence on high-stakes committee decisions. This contributes to an emerging body of literature on gender differences 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, 2022), and less likely to become group leaders (Alan et al., 2019; Coffman et al., 2021; Born et al., 2022) or contribute ideas (Coffman, 2014; Isaksson, 2019; Hardt et al., 2022).[1] These differences have been shown to be particularly stark in male-typed decision environments. Based on a field experiment with accounting students, Stoddard et al. (2023) provide

[1] More generally, there exists broad evidence on gender differences in other dimensions of behavior, which are indirectly related to committee deliberation. For example, women have been shown to 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, 2021), engage less in self-promotion (e.g., Exley & Kessler, 2022), be more likely to accept unpopular tasks (e.g., Babcock et al., 2017), and ask for lower wages (e.g., Roussille, 2022) than men.

detailed evidence that lone women have a lower influence in their team compared to women in majority-women teams.[2] Overall, the existing evidence provides important insights on behavioral differences between men and women, as well as the effect of a group's gender composition and its underlying mechanisms. Our results show how these phenomena can affect high-stakes committee decisions in the labor market.

This paper also provides new insights into barriers to the effectiveness of gender quotas. Previous studies suggest that gender quotas in hiring and promotion committees do not improve the success of female candidates (e.g., Bagues & Esteve-Volart, 2010; Bagues et al., 2017; Williams & Ceci, 2015; Deschamps, 2022). Moreover, it has been shown that gender board quotas do not alter the labor market outcomes of the female workforce (e.g. Bertrand et al., 2018; Maida & Weber, 2022). Generally speaking, we add to this literature by showing that systematic gender differences in influence can limit the effectiveness of quotas. More specifically, we distinguish the effect of female representation at the level of the individual evaluation from its aggregate impact at the committee level. Our results make the case that women can indeed be more favorable towards female candidates. However, they also show that the introduction of 'lone women' may not be sufficient to change the outcomes of committee deliberation, given the behavior of the members. As it is often not an option 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 of committee members in the deliberation process.

More broadly, our findings add 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 &

[2] Additional descriptive evidence from various contexts provides evidence on gender differences in communication behavior. For example, women have been shown to be less likely 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 avoid speaking up in class (Burztyn et al., 2017).

Stocken, 2008). Empirical evidence on committee decision-making is scarce. Goeree and Yariv (2011) examine how voting rules with and without communication perform in a laboratory setting, Chan (2021) studies the role of seniority for the aggregation of information in physician teams, and Iaryczower et al. (2018) structurally analyze the effectiveness of deliberation in US appellate courts.[3] We contribute with empirical insights into how committees aggregate information in the absence of formal voting rules. Our results show that the aggregation outcomes of committees are not only shaped by their institutional rules but also the behavior of their individual members. Specifically, we find that the weighting of individual information differs systematically between members in a way that is unlikely to be due to differences in expertise or the accuracy of information.

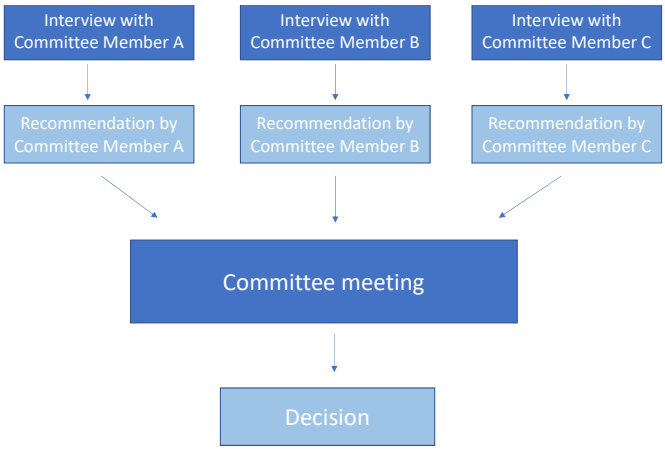
The remainder of this paper is structured as follows. Section 2 provides information on the institutional setting. Section 3 describes the data source. Section 4 provides evidence on the aggregation of information in hiring committees. In Section 5, we estimate the gender gap in influence on committee 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

We study the hiring process of a large consulting company that offers high-wage permanent positions to recent college graduates, as well as internships for college students. Starting wages for permanent positions are in the top 10% of the overall German wage distribution. The process is organized through interview days, which take place at different locations throughout the year. On every interview day, a committee of evaluators decides on the hiring outcomes of several candidates. The number of candidates and evaluators in the committee varies between interview days. The median interview day includes eight candidates and a committee of eight evaluators. There is no hiring quota at the level of the interview day, as the company

[3] More generally, a few empirical studies analyze how individual recommendations affect the decisions of third parties, such as journal editors (Card et al., 2019; Card & DellaVigna, 2019) or admission committees (Bai et al., 2022).

Figure 1: Hiring Decision Process (per Candidate)




hires on a rolling basis.

Candidates at the interview stage apply for a permanent position (67%) or an internship (33%) at the company's German-speaking branch. They have been pre-screened by the company's HR department on a paper basis. Evaluators in the committee are consultants at the company who have undergone professional interviewer training. The composition of the committee changes between interview days. While evaluators are assigned to a primary interview location, their participation depends on their availability and the required size of the committee on a given day. Moreover, it is common that evaluators help out in committees at other locations.

Figure 1 sketches the process through which hiring committees reach their decision on a given candidate during our data period. In a first step, three committee members conduct independent one-to-one interviews with the candidate and make an individual assessment. In the second step, they enter a committee meeting to deliberate on each candidate and reach hiring decisions. We now describe these two steps in more detail.

In the first step, every candidate goes through three one-to-one interviews with three different evaluators of the committee. At the end of our data period, the number of interviews per candidate was reduced to two (12% of the data). The HR department quasi-randomly assigns candidates to committee members. The assignment occurs conditional on candidate gender,

as the company tries to ensure that female candidates are interviewed by at least one woman.[4] Committee members usually interview and assess three candidates per day. They form their assessments individually, without any discussion or consultation with the other committee members. The main interview assessment outcome is a recommendation on whether to hire the candidate, which is expressed on a three-point scale (1=reject/2=possible offer/3=offer). This recommendation summarizes several sub-ratings, which assess the candidate's cognitive and non-cognitive skills. The cognitive rating is based on the candidate's performance in solving a business case and focuses on mathematical and structuring skills. The non-cognitive ratings measure different dimensions of leadership and teamwork skills.[5] The translation of sub-ratings into the recommendation is defined by the company.

After all evaluators have entered their recommendations and sub-ratings into the applicant tracking system, the committee meets to deliberate and decide on the hiring outcomes of all candidates, based on the recommendations. The committee consists of all evaluators who conducted interviews, as well as a representative of the HR department and the head of the interview team. There is no deterministic mapping of recommendations into committee decisions. As the vast majority (80%) of recommendations are not unanimous, there is substantial room for the committee's deliberation to influence the hiring decision. The discussion on a candidate mostly takes place between the evaluators who conducted interviews with the candidate. The HR representative sets the order in which evaluators express their assessments of a given candidate. As a general rule, positive recommendations are expressed first. Other committee members may also intervene in the discussion; for instance, by asking questions about the interviews. The deliberation results in a decision about a candidate's hiring outcome. In the case of internship applications, the committee needs to take a final decision. Applicants to a permanent position at the company can be forwarded to an additional interview with a senior manager, who double-checks the committee's assessment in case of doubt.[6]

[4] This is done to improve the workplace image conveyed to female candidates and to avoid the assessment of a female candidate only represents the views of male evaluators.

[5] Each of a candidate's three interviews measures a different set of non-cognitive skills.

[6] About 15% of candidates are forwarded to a final interview with a senior manager after the committee meeting. Our results are robust to excluding these candidates from the analysis (see Table C.6).

# 3 Data

## 3.1 Data Source

The data cover all interviews conducted between 2017 and 2021 at the company's German-speaking branch. The main unit of observation is the individual interview, corresponding to a unique committee member – candidate pair. Our estimation sample includes 8,117 interviews with 2,913 candidates, conducted by 359 different committee members on 429 interview days.[7] The data report the assessment outcome of each interview (hiring recommendation and sub-ratings), as well as the committee's final offer decision. Moreover, we observe a candidate's gender, study field, high-school GPA and aspired type of position (internship vs. permanent), the order of an interview, as well as the committee member's gender, managerial responsibility, and hiring experience (measured as the time since the first interview).

## 3.2 Descriptive Statistics

**Committee Member Characteristics and Evaluation Behavior** Table 1 shows summary statistics on the characteristics and evaluation behavior of committee members at the level of the interview (i.e., the committee member-candidate pair). Summary statistics on candidate characteristics are reported in Appendix Table A.1.

According to Panel A of Table 1, the average interview is conducted by a committee member with two years of hiring experience. About one-third of interviews are conducted by committee members holding a management position at the company, and about two-thirds of committee members' interviewing activity takes place in their assigned primary location. Notably, neither the average value of these characteristics nor their variance differs between male and female committee members.

Panel B provides summary statistics on the committee members' evaluation behavior. We observe that women give slightly higher recommendations on average. As shown in Appendix

[7] We excluded 726 interviews due to missing information on recommendations, committee member gender, or candidate gender. Moreover, we exclude 574 interviews because the candidate did not have the foreseen number of interviews.

Table 1: Summary Statistics on Committee Member Characteristics and Evaluation Behavior

| | **All** | | **Male CM** | | | **Female CM** | | | **Difference** | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Panel A: Characteristics* | Mean | SD | Mean | SD | N | Mean | SD | N | Diff. | p-value |
| Hiring Experience (in years) | 2.03 | 1.68 | 2.04 | 1.66 | 5517 | 2.00 | 1.71 | 2600 | -0.037 | 0.84 |
| Manager | 0.33 | 0.47 | 0.32 | 0.47 | 5517 | 0.34 | 0.47 | 2600 | 0.020 | 0.71 |
| Primary Interview Location | 0.70 | 0.46 | 0.70 | 0.46 | 5517 | 0.70 | 0.46 | 2600 | -0.005 | 0.86 |
| *Panel B: Evaluation Behavior* | Mean | SD | Mean | SD | N | Mean | SD | N | Diff. | p-value |
| Recommendation (1-3 scale) | 1.89 | 0.85 | 1.86 | 0.84 | 5517 | 1.94 | 0.85 | 2600 | 0.078 | 0.03 |
| Within-CM SD of Recommendations | 0.82 | 0.10 | 0.83 | 0.10 | 5515 | 0.82 | 0.10 | 2597 | -0.008 | 0.47 |
| Share of Disagreements | 0.57 | 0.39 | 0.57 | 0.39 | 5517 | 0.58 | 0.38 | 2600 | 0.011 | 0.20 |
| Avg. Disagreement in Points | 0.75 | 0.58 | 0.74 | 0.59 | 5517 | 0.76 | 0.57 | 2600 | 0.012 | 0.36 |

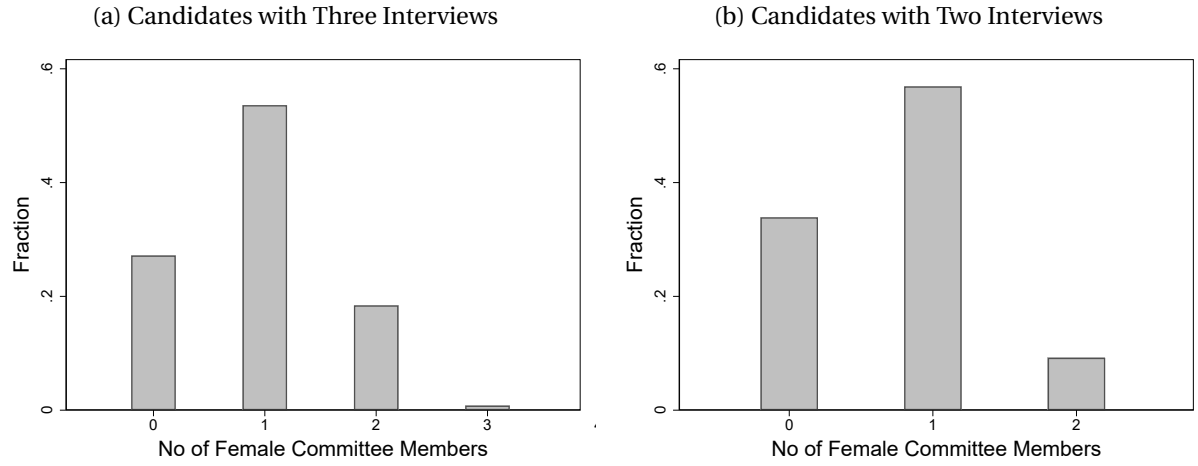
*Note:* CM= committee member. The unit of observation is the interview (i.e., the committee member-candidate pair). p-values are computed based on interviewer-level clustered standard errors. All characteristics are measured at the time of the interview. "Primary interview location" takes the value of one if the interview day takes place at the committee member's assigned primary location. "Share of disagreements" is the share of other committee members who give a different recommendation to the same candidate. "Avg. Disagreement in Points" is the average absolute difference in points between a committee member's own recommendation and the other members' recommendations on the same candidate.

Figure A.1, this reflects a small rightward shift of recommendations on the three-point scale. However, this small difference is too small to generate significant differences in the incidence or level of disagreement with the other committee members assessing the same candidate. Additionally, the within-individual varies of recommendations does not differ by gender, providing evidence that men and women do not systematically differ in their use of the three-point scale.

Appendix Table A.2 additionally reports that recommendations by men and women show the same correlation with a predicted measure of candidate quality, where the prediction is based on the candidate's high school GPA and CV screening scores. For a sub-sample of surveyed committee members, Appendix Figure A.2 shows that men and women express similar levels of confidence in their average recommendations. Overall, the data suggest that women and men in the hiring committees differ little in terms of their evaluation behavior.

**Number of Female Committee Members per Candidate** Figure 2 illustrates the variation in the number of female committee members per candidate. In cases with three interviews

Figure 2: Number of Female Committee Members per Candidate




*Note:* The unit of observation is the candidate. N=2,433 in panel (a), N=480 in panel (b).

(panel a), about 55% of candidates are interviewed by one woman and two men. About 25% are only interviewed by men and about 20% by two women and one man. It almost never occurs that a candidate is interviewed only by women. The distribution looks similar for candidates with two interviews (panel b). Appendix Figure A.3 additionally reports the distribution of the share of female committee members at the level of the interview day. At the average interview day, about 30% of committee members are female.

## 3.3 Assignment of Candidates to Committee Members

In Table 2, we assess whether male and female committee members are assigned to the same types of candidates. The only criterion officially taken into account for the assignment is gender, as the company tries to ensure that every female candidate is interviewed by at least one woman. In line with the assignment rule, we observe that female candidates are about 17-19% more likely to be evaluated by a female committee member. However, conditional on candidate gender, no other characteristic is significantly related to the gender of the committee member. This holds when both using the overall sample variation (column 1), and restricting it to within-committee (i.e., interview day) variation (column 2). Appendix Table A.3 reports the

Table 2: Assignment of Candidates to Female Committee Members

| | P(Female CM) | |
|---|---|---|
| | (1) | (2) |
| Female | 0.165*** | 0.185*** |
| | (0.010) | (0.010) |
| Field: Business | 0.003 | -0.007 |
| | (0.013) | (0.014) |
| Field: STEM | 0.006 | -0.009 |
| | (0.014) | (0.014) |
| Internship Application | -0.008 | 0.014 |
| | (0.013) | (0.014) |
| High GPA | 0.015 | -0.000 |
| | (0.010) | (0.009) |
| High GPA in Math | -0.010 | -0.013 |
| | (0.012) | (0.011) |
| High CV Score | 0.007 | 0.003 |
| | (0.010) | (0.010) |
| Committee FE | No | Yes |
| p-value joint significance (excl. gender) | 0.72 | 0.73 |
| Outcome Mean | 0.32 | 0.32 |
| N | 8117 | 8117 |

*Note:* The unit of observation is the candidate-committee member pair (i.e., the individual interview). CM=committee member. "High"= above median. Standard errors are clustered at the committee level (N=429). $^{*}$ $p < 0.10$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$.

results of a similar assessment at the candidate level. It supports the notion that a candidate's characteristics do not predict how many female committee members he or she gets assigned to.

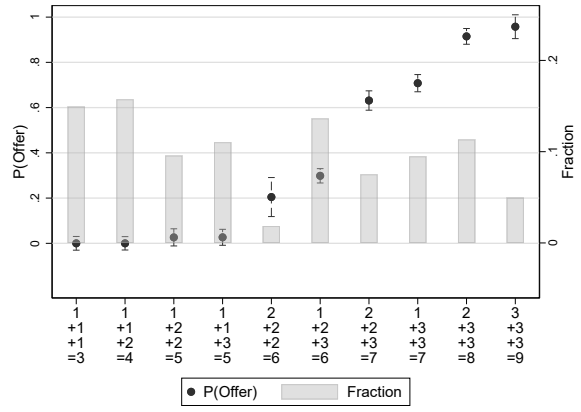
# 4 Aggregation of Information

In this section, we describe the aggregation of individual recommendations into committee decisions. First, we explore how committees decide conditional on the combination of recommendations on a given candidate. Second, we document the translation of individual recommendations into decisions.

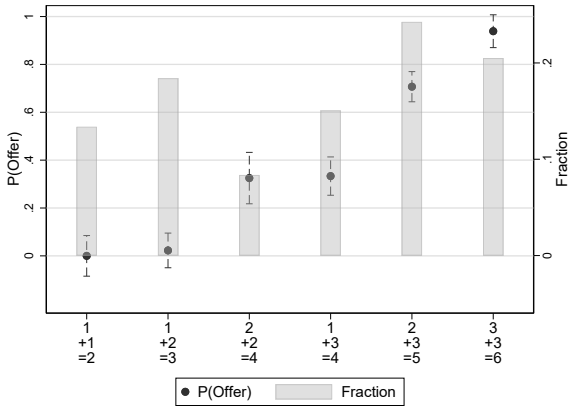
It is important to note that all committee members follow the same objective to hire the

Figure 3: Aggregation of Information in the Committee

(a) Decisions with Three Recommendations




(b) Decisions with Two Recommendations



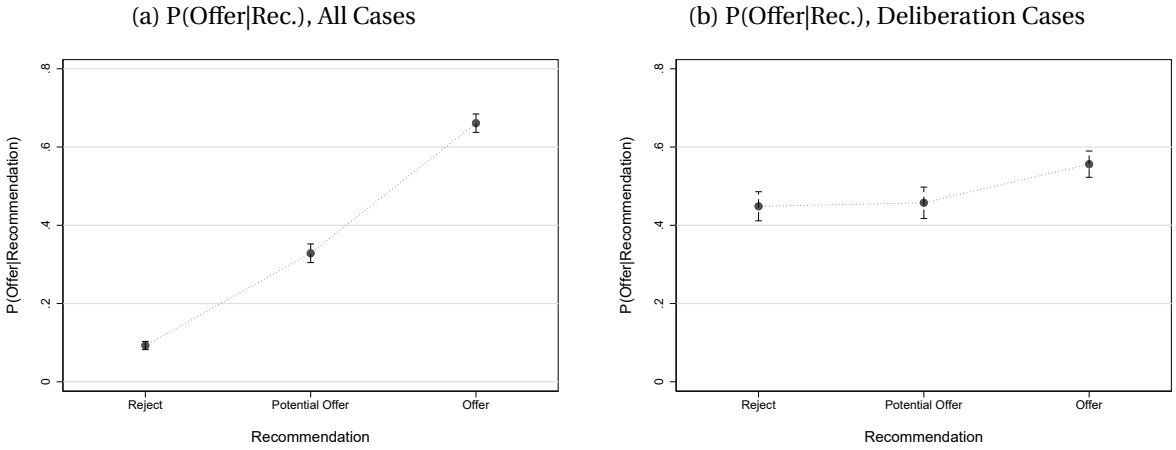

*Note:* The unit of observation is the candidate. N=2,433 in panel (a), N=480 in panel (b). The x-axis shows the combination of initial recommendations per candidate, where 1='reject'; 2='potential offer'; 3='offer'. The left y-axis shows the probability that a given combination results in a positive offer decision by the committee. The right y-axis shows the fraction of a given combination in the data. Dashed lines show 95% confidence intervals, with standard errors clustered at the committee level (N=429).

best candidates, who fulfill the company's pre-defined selection criteria. Nevertheless, they might have different assessments of the same candidate because they have observed the same candidate independently in different interviews. This can lead to variations in assessments; for example, due to circumstantial factors or the subjective nature of interviewing. Moreover, each interview assesses a slightly different set of non-cognitive skills. These factors can lead to disagreement in individual recommendations, even in the absence of disagreement about the right type of candidate.

**Decision-Level Evidence** Figure 3 shows the likelihood that the committee decides to make an offer, given a specific combination of individual offer recommendations.[8] Additionally, the figure depicts the fraction of a given combination in the data. Overall, we observe that the likelihood of receiving a job offer strongly increases with the sum of recommendations in points. Nonetheless, in line with the institutional setup, the increase is not deterministic: the offer

[8] Appendix Figure A.4 shows that the share of candidates who receive an offer widely differs between interview days, ranging from 0 to 1.

Figure 4: Translation of Recommendations into Offers




*Note:* The unit of observation is the interview recommendation (i.e., candidate-committee member pair). N=8,117 in panel (a); N=2,787 in panel (b). Dashed lines show 95% confidence intervals, with standard errors clustered at the committee level (N=429).

probability does not jump from zero to one upon reaching a certain threshold or a majority of positive recommendations. Instead, the pattern suggests a categorization into three groups: in around 50% of cases, candidates have very low support ($\leq 5$ points in panel (a); $\leq 3$ points in panel (b)), which results in a very small offer probability. Around 15% have very strong support ($\geq 8$ points in panel (a); 6 points in panel (b)) and receive an offer with a probability close to one. In the remaining 35% of cases, the decision is less clear-cut (6 or 7 points in panel (a); 4 or 5 points in panel (b)). Here, the likelihood of receiving an offer increases with the number of overall points, but in a non-deterministic way, suggesting an important role of committee deliberation. We refer to these decisions as 'deliberation cases'. Appendix Figure B.5 shows that a similar pattern emerges when looking at the average sub-ratings per candidate instead of the recommendations.

**Recommendation-Level Evidence** Moving to the level of the individual recommendation, Figure 4 shows the probability that a given recommendation translates into an offer. In line with the committee-level evidence, better recommendations are associated with a higher offer probability. The probability that a 'reject' recommendation translates into an offer is 9%,

compared to 66% for a clear 'offer' recommendation. When focusing on deliberation cases in panel (b), the individual recommendation is much less predictive of the offer decision. A 'reject' recommendation still translates into an offer with a probability of 45%. This probability increases only slightly to 56% for clear 'offer' recommendations.

Taken together, the patterns reveal that the individual recommendations are strong but imperfect predictors of the committees' hiring decisions. Moreover, the decision process cannot be summarized by voting or cut-off rules that give equal weight to each recommendation.[9] Thereby, the evidence suggests room for committee deliberation in determining final decisions.

# 5 Gender Differences in Influence

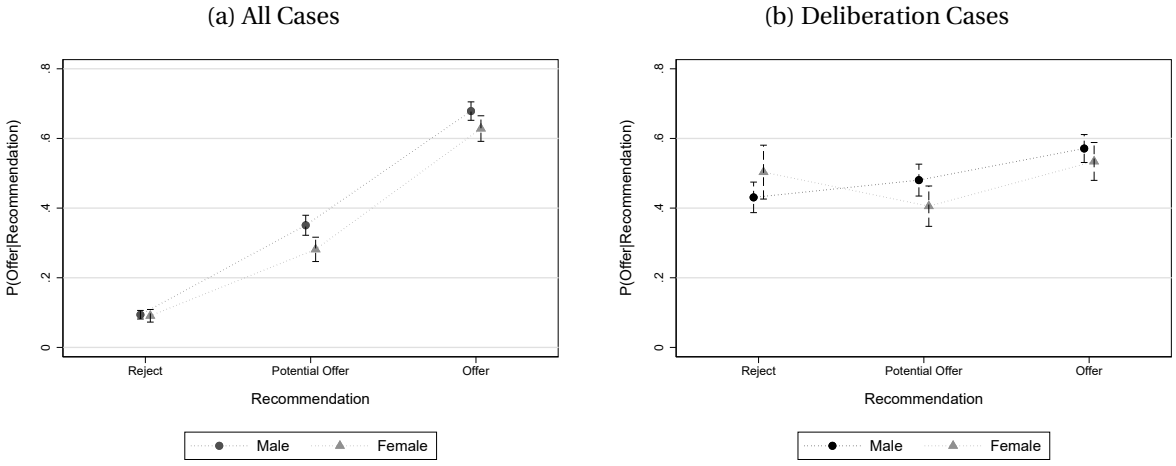
In this section, we analyze whether systematic gender differences exist in the influence of committee members' recommendations on the final decision outcome. We first show descriptive evidence on how recommendations by men and women translate into the committee's offer decision. We then proceed with a more formal analysis of gender differences in the probability that a recommendation aligns with the committee's decision.

## 5.1 Descriptive Evidence

In the first step, we provide descriptive evidence on gender differences in the translation of individual recommendations into committee decisions. Figure 5 shows the probability that a candidate receives an offer, conditional on the recommendations made by men and women. Panel (a) is based on the sample of all cases. It shows that the relationship between recommendations and offer decisions is weaker among female committee members. In particular, positive recommendations by women are five to seven percentage points less likely to translate into an offer. Appendix Table C.4 shows that the slope of the relationship – as estimated by a

[9] In principle, one might be able to explain the observed patterns with a cut-off rule with precision-weighted recommendations. However, this would require that committee members can quantify and communicate the true precision of their interview signal. According to conversations with the company's HR department, this is not relevant in practice.

Figure 5: Translation of Recommendation into Offers, by Gender of the Committee Member




*Note:* The unit of observation is the interview recommendation (i.e., candidate-committee member pair). N=8,117 in panel (a); N=2,787 in panel (b). 95% confidence intervals, standard errors clustered at the committee level (N=429).

linear regression – significantly differs by gender, being about 8% flatter for female committee members.

Panel (b) focuses on deliberation cases, where the committee decision is ex-ante ambiguous given the initial recommendations (see section 4). Again, positive recommendations by women are less likely to result in an offer. Moreover, deliberation cases also display an alignment gap for negative assessments: a reject recommendation is about 7 percentage points more likely to translate into an offer when it was made by a woman.

Overall, the raw data suggest that recommendations made by women are on average less aligned with the committee's decision. Appendix Figure C.6 shows that this also holds when considering the average interview rating instead of the recommendations. Moreover, Appendix Figure C.7 illustrates that the gender difference in translation is slightly larger for the non-cognitive rating. This is intuitive, as every committee member assesses the same cognitive, but different non-cognitive skills. As a result, the rating of non-cognitive skills is more prone to disagreement.

## 5.2 Empirical Strategy

To analyze the gender difference in influence more formally, we create an indicator of alignment between individual recommendations and committee decisions. The indicator equals one if a reject recommendation translates into a rejection and if a (possible) offer recommendation translates into a job offer.[10] We use the following framework to estimate the gender gap in the alignment of initial recommendations with committee decisions:

$$A_{i,j,c} = \beta_1 F_j + \delta X_{i,j} + \omega_c + \epsilon_{i,j,c} \tag{1}$$

The binary outcome $A_{i,j,c}$ equals one if the decision made by committee $c$ on candidate $i$'s outcome is in line with the committee member's $j$'s initial recommendation, as outlined above. The indicator $F_j$ equals one if the committee member who made the recommendation is female. The coefficient of interest, $\beta_1$, measures the gender gap in the probability of alignment.

The vector $X_{i,j}$ includes control variables for different factors that might drive gender differences in the alignment of recommendations with decisions. We subsequently add several sets of control variables. First, we control for committee member and candidate characteristics (see summary statistics in Table A.1 for candidates and Panel A of Table 1 for committee members). We then add controls for the amount of initial (dis-)agreement on a given candidate. More precisely, we include indicators for every possible combination of initial recommendations on the candidate (see bins plotted in Figure 3), as well as indicators for a committee member's individual level of disagreement with the other committee members.[11] Finally, we control for committee member leniency using her leave-one-out mean recommendation.

The vector $\omega_c$ includes committee fixed effects, which capture potential differences in decision-making between committees. As committee fixed effects coincide with interview day

[10] In robustness checks, we exclude 'potential offer' recommendations, which arguably express uncertainty in the assessment.

[11] More precisely, we sum up the pair-wise absolute differences between a member's own recommendation and the other two members' recommendations. For example, if a committee member gave a recommendation of 1 (=reject) and the other two committee members recommended 3 (=offer), the measure takes the value $|3-1|+|3-1|=4$.

Table 3: Gender Difference in Influence

| | Alignment Probability | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Female | -0.049*** | -0.054*** | -0.046*** | -0.042*** | -0.023** |
| | (0.011) | (0.012) | (0.010) | (0.010) | (0.010) |
| Female x Deliberation Case | | | | | -0.055** |
| | | | | | (0.022) |
| Hiring Experience (std.) | | 0.030*** | 0.024*** | 0.025*** | 0.025*** |
| | | (0.007) | (0.005) | (0.005) | (0.005) |
| Manager | | -0.018 | -0.016 | -0.009 | -0.009 |
| | | (0.013) | (0.011) | (0.012) | (0.012) |
| Primary Interview Location | | -0.005 | -0.006 | -0.007 | -0.007 |
| | | (0.013) | (0.011) | (0.011) | (0.011) |
| Committee FE | No | Yes | Yes | Yes | Yes |
| Basic Controls | No | Yes | Yes | Yes | Yes |
| Control for Disagreement | No | No | Yes | Yes | Yes |
| Control for Leniency | No | No | No | Yes | Yes |
| Outcome Mean | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 |
| N | 8117 | 8117 | 8117 | 8117 | 8117 |

*Note:* Estimates are based on equation 1. Basic controls include committee member and candidate characteristics reported in Panel A of Table 1 and Table A.1, as well as indicators for the order of the interview. Controls for disagreement are indicators for the combination of recommendations (see bins in Figure 3), as well as the sum of the pair-wise absolute differences between a member's own recommendation and the other two members' recommendations on a given candidate. The control for leniency is the committee member's leave-one-out recommendation. Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

fixed effects, they also account for the composition of the candidate pool and other potential time- and location-specific effects. Standard errors are clustered at the committee level (N=429).

## 5.3 Results

Table 3 reports the estimated gender gap in the probability that an individual recommendation aligns with the committee decision. Column (1) shows the raw gap, which amounts to about 5 percentage points (8% relative to the male average). In column (2), we add committee fixed effects, as well as controls for candidate and committee member characteristics. Column (3) adds detailed controls for the amount of (dis-)agreement in evaluators' recommen-

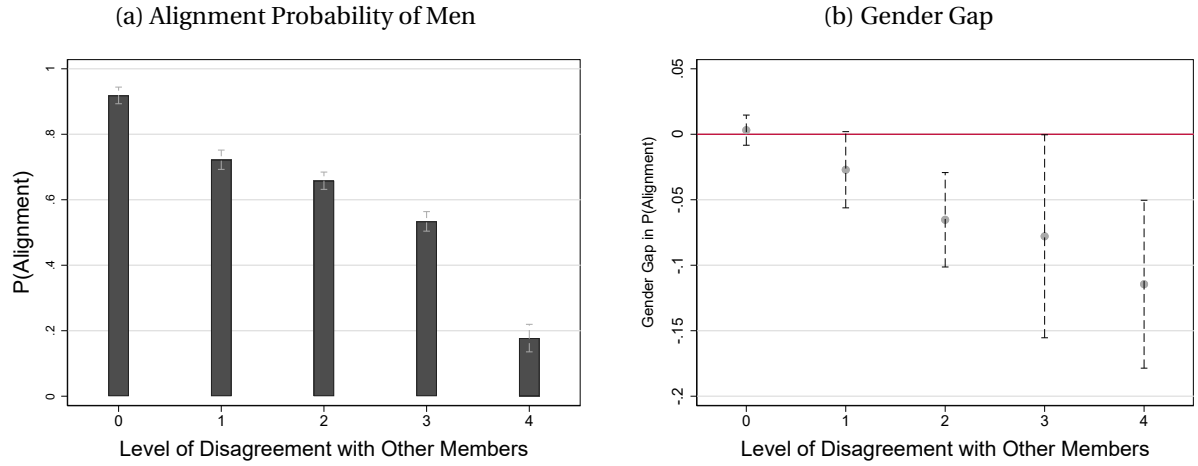
dations, as described above. In column (4), we additionally control for the leniency of a committee member through her leave-one-out mean recommendation. We thereby account for the fact that women are slightly more lenient in their recommendations, which might affect their alignment probability. Overall, the different control variables lead to small changes in the point coefficients, but do not significantly affect the estimated gap, whose point estimate still amounts to about 4.2 percentage points in the full specification.

The results also show that hiring experience – measured by the number of years since the first interview day – is the only other characteristic that shows a significant relationship with the probability of alignment. In terms of relative magnitudes, the size of the gender gap is comparable to a 1.7 standard deviation ($\approx$ 3 years) decrease in hiring experience. The alignment probability does not depend on the committee member's manager status or on whether the interview day occurs at the committee member's primary location.

As shown in Section 4, the scope for deliberation in the committee's decision process largely depends on the combination of initial recommendations. In particular, the scope for deliberation is large whenever a candidate's initial recommendations do not express a consistent assessment (see Figure 3). Column (5) shows that deliberation cases are the main driver of the gender gap in alignment. In these cases, the recommendations of women are about 8 percentage points less likely to align with the committee decision than those of men (15% relative to the male average in deliberation cases). We also note a minor alignment gap of about 2 percentage points in cases with little room for deliberation. This is attributable to cases in which committees decide against a clear majority of recommendations. For instance, in all the (rare) cases in which committees decide to hire a candidate with one 'offer' and two 'reject' recommendations, the single 'offer' recommendation was made by a man.

We further investigate the role of disagreement in Figure 6. In particular, we test whether the alignment gap increases when an individual committee member more strongly disagrees with their colleagues. For this purpose, we interact the female indicator with the individual-specific level of disagreement, which sums up the pair-wise absolute differences between an individual's own recommendations and the other two recommendations on the same candi-

Figure 6: Gender Gap in Influence, by Level of Disagreement




*Note:* The analysis only includes candidates with three recommendations. The level of disagreement is measured as the sum of the pair-wise absolute differences between a member's own recommendation and the other two members' recommendations on a given candidate. Estimates in panel (b) are based on equation 1, using the full set of control variables. Standard errors are clustered at the committee level (N=429). $^{*}$ $p < 0.10$,$^{**}$ $p < 0.05$,$^{***}$ $p < 0.01$.

date.[12] Panel (a) of Figure 6 shows the baseline probability of alignment for male committee members. Intuitively, the probability of alignment decreases with the level of disagreement. Panel (b) illustrates the corresponding gender gap. Somewhat mechanically, there is no gender gap in alignment when all committee members agree in their recommendations. The gap emerges among members with one point of disagreement and increases up to about 11.5 percentage points for the maximum level of disagreement (65% compared to male baseline in this category). Appendix Table C.5 reports results from the corresponding linear specification, showing that the gender gap in alignment increases on average by about 2.8 percentage points for every one-point increase in the disagreement measure.

Appendix C reports the results from several robustness checks and additional analyses. First, Appendix Table C.6 shows that the main estimates are robust to the exclusion of 'potential offer' recommendations, candidates who have an additional final interview after the committee's decision, and candidates with only two recommendations. Moreover, the results

[12] We restrict the analysis to cases with three recommendations per candidate, such that this measure is comparable across committee members.

Table 4: Gender Composition and Gender Difference in Influence

| | All Cases | | Deliberation Cases | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Male Majority × Female | -0.054*** | -0.056*** | -0.070** | -0.092*** |
| | (0.015) | (0.013) | (0.027) | (0.025) |
| Parity × Female | -0.048 | -0.031 | -0.046 | -0.034 |
| | (0.032) | (0.032) | (0.042) | (0.043) |
| Female Majority × Female | 0.015 | 0.003 | 0.045 | 0.011 |
| | (0.029) | (0.026) | (0.060) | (0.060) |
| Committee FE | No | Yes | No | Yes |
| Basic Controls | No | Yes | No | Yes |
| Disagreement Controls | No | Yes | No | Yes |
| Leniency Control | No | Yes | No | Yes |
| Outcome Mean | 0.68 | 0.68 | 0.52 | 0.52 |
| p-value: Male Maj.= Female Maj. | 0.03 | 0.04 | 0.08 | 0.11 |
| N | 8117 | 8117 | 2787 | 2787 |

*Note:* Male majority= 2 male : 1 female CM per candidate. Parity= 1 male : 1 female CM per candidate. Female majority = 2 female : 1 male CM per candidate. Estimates are based on equation 1. Basic controls include committee member and candidate characteristics reported in Panel A of Table 1 and Table A.1, as well as indicators for the order of the interview. Controls for disagreement are indicators for the combination of recommendations (see bins in Figure 3), as well as the total absolute difference in points between a member's recommendation and the other members' recommendations. Standard errors are clustered at the committee level (N=429). $^{*}p<0.10,^{**}p<0.05,^{***}p<0.01$.

are robust to the inclusion of candidate fixed effects. Table C.7 reports that the gender gap in alignment does not vary systematically with committee member or candidate characteristics, with the exception that it is slightly stronger when committees decide on female candidates. Notably, seniority or experience does not significantly mitigate the gap.

## 5.4 Role of the Gender Composition

An important feature of our institutional setup is the majority of men on the committee, which is representative of many high-level decision-making bodies. In the modal case (≈ 50%), a candidate is assessed and subsequently also discussed by two men and one woman (see Figure 2). We now investigate the extent to which the gender gap in alignment varies with the gender composition by comparing decisions reached under a 2:1 male majority, a 2:1 female

majority, and parity (1:1).[13]

Table 4 reports the results. In columns 1 (raw gap) and 2 (full specification), the estimation includes all cases. We observe that the gender composition has a significant impact on the gender difference in influence. The gap is mostly driven by situations where two men and one woman discuss a candidate. In turn, no gender gap in either direction emerges in the reverse case, where women are in a 2:1 majority. Cases of parity are not observed sufficiently often to allow for precise conclusions. However, the negative point estimate suggests that these cases exhibit a gender gap between the two extremes. Columns 3 and 4 show the same pattern when focusing only on deliberation cases. Among these cases, the gap increases to approximately 9.2 percentage points (17%) in situations of a male majority. Taken together, the results demonstrate that the gender composition of committees matters for the influence of their members.[14]

## 5.5 The Distribution of Influence

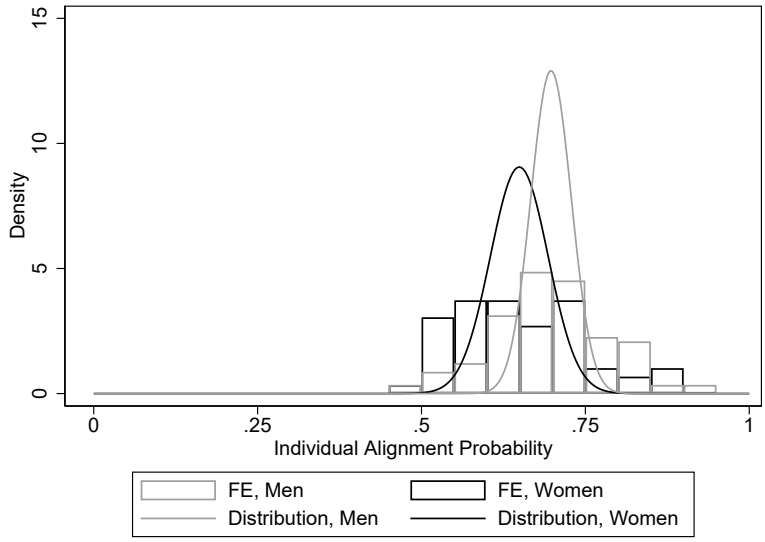
The results presented so far have documented systematic average gender differences in the alignment of initial recommendations with committee decisions. We now seek to understand the extent to which the alignment of recommendations varies across individual committee members. The goal is to compare the average gender gap in influence to the general dispersion of influence between individuals and to quantify the gender gap along the distribution of influence.

To estimate the distribution of influence, we exploit the fact that we observe the same individual in several committees, interacting with different sets of colleagues. We start by esti-

[13] Note that the full hiring committee usually includes more than only the three evaluators who assess a given candidate. Arguably, the gender composition of the rest of the committee could also play a role. We focus on variation in the composition of the three evaluators for two main reasons: first, it results from a quasi-random assignment process; and second, according to information provided by the company, the discussion is usually dominated by the three evaluators who interviewed a given candidate.

[14] Note that the institutional setup makes it unlikely that initial recommendations react strategically to the gender of the other committee members evaluating the same candidate. While the composition of the whole committee on an interview day is salient, it is not salient who are the other two evaluators assessing a given candidate. In line with this, Appendix Figure C.8 shows that the recommendations of female and male evaluators do not react to the gender composition at the candidate level.

Figure 7: Distribution of Influence: Illustration




*Note:* The figure shows a histogram of the "raw" individual fixed effects and the estimated parametric normal distributions, based on the estimates reported in Table 5.

mating a committee member's fixed effect on the probability of alignment, as defined in section 5.[15] Based on the estimated fixed effects, we estimate the mean and the standard deviation of the underlying distributions for male and female committee members. When estimating the standard deviation, we take into account the fact that the dispersion of the estimated fixed effects overstates the true dispersion of alignment, due to statistical noise.[16]

Figure 7 plots the individual-specific average alignment probabilities, i.e., the fixed effects added to the sample mean, as well as the parametric normal distributions for men and women, based on the respective mean and standard deviation. Table 5 reports the corresponding key statistics. We make the following main observations. First, the estimates suggest significant between-person heterogeneity in influence, among both men and women. A Chi-squared test

[15] We restrict the analysis to committee members with at least fifteen recommendations in the data. This limits the sample to 6,855 recommendations made by 174 committee members. Appendix Figures C.9 (a) and (b) provide robustness checks setting the minimum number of interviews to 10 or 20, leading to similar pictures.

[16] More precisely, we estimate the fixed effects $\hat{\beta}_i$ and the corresponding standard error $s_i$. We estimate the mean of the two normal distributions by taking the average fixed effect for male (m) and female (f) committee members, respectively, resulting in $\hat{\mu}_f$ and $\hat{\mu}_m$. We follow Aaronson et al. (2007) and estimate the standard deviation of the normal distribution by calculating the variance of the fixed effects and correcting for the estimation error using the expression $\hat{\sigma}^2_{\beta,g} = \frac{1}{J_f}\sum_{j=1}^{J_g}[(\hat{\beta}_j - \hat{\mu}_{\beta,g})^2 - s_j^2]$ for $g = m, f$.

Table 5: Distribution of Influence: Key Statistics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | $\sigma_{fe}$ | $\sigma_{adj.}$ | p10 | p25 | p50 | p75 | p90 | $\chi^2$ | p-value |
| Male | 0.698 | 0.089 | 0.031 | 0.658 | 0.677 | 0.697 | 0.718 | 0.737 | 136.905 | 0.071 |
| Female | 0.649 | 0.098 | 0.044 | 0.593 | 0.619 | 0.649 | 0.679 | 0.705 | 92.284 | 0.003 |

*Note:* The table reports the properties of the estimated fixed effects of alignment (added to the respective sample average) for male and female committee members with at least fifteen recommendations in the data. Column (1) displays the mean, column (2) the unadjusted variance, and column (3) the adjusted variance, taking into account the standard error of the estimates. We follow Aaronson et al. (2007) and calculate $\sigma_{adj.}$ using $\hat{\sigma}^2_{\beta,g} = \frac{1}{J_g}\sum_{j=1}^{J_g}[(\hat{\beta}_j - \hat{\mu}_{\beta,g})^2 - s_j^2]$ for each gender $g = m, f$. The percentiles in columns (4)-(8) are calculated using a normal distribution with the respective mean and the square of the adjusted standard deviation. Column (9) shows the $\chi^2$ test statistic for a test of heterogeneity and column (10) the corresponding p-value. The test statistic is $\sum_{j=1}^{J_g}(\hat{\beta}_j - \hat{\mu}_{\beta,g})^2/s_j^2$.

rejects the hypothesis of zero dispersion for both distributions. Second, the size of the average gender gap in influence compares to the interquartile range of influence among male committee members, where a man at the $25^{th}$ percentile can align about 68% of his initial recommendations, compared to 72% for a man at the $75^{th}$ percentile. Third, the estimates suggest that the distribution of women is more dispersed than that of men. Given that women are on average less influential, this implies the gender gap is more pronounced at the lower than the upper part of the distribution. For example, the gap amounts to about 7 percentage points at the $10^{th}$ percentile and decreases to about 3 percentage points at the $90^{th}$ percentile. Finally, the shape of the distributions suggests that almost all men are more influential than the median woman.

Appendix Figure C.10 provides evidence from a placebo check regarding the distribution of influence. We use the predicted offer probability of assigned candidates as a placebo outcome, which should not differ systematically between individual committee members.[17] Indeed, the estimated fixed effects, as well as the corresponding parametric distributions show no significant heterogeneity for this outcome.

[17]The offer probability is predicted based on a candidate's CV screening scores, and high-school GPA (overall and math). When estimating the coefficients of these variables, we leave out the candidate's interview day.

## 5.6 Discussion of Mechanism

The results have shown that recommendations made by women are systematically less aligned with committee decisions than those made by men. Given that woman in our sample are neither less trained nor less experienced in conducting interviews, it seems implausible that the gender gap arises because women produce less informative evaluations based on their interviews.[18] A further argument against systematic gender differences in the accuracy of evaluations is the observation that recommendations by men and women display the same correlation with pre-determined measures of candidate quality (see Appendix Table A.2). Moreover, survey responses of committee members reveal similar levels of confidence in the quality of recommendations by men and women (see Appendix Figure A.2). It is also unlikely that the gender gap in alignment is a simple artifact of being in the (gender) minority per se. While one might argue that committee members of the minority gender will generally have a harder time influencing committee decisions, we find no gender differences in influence when women are in the majority (see Table 4). Additionally, the gender gap in influence persists when controlling for disagreement with the other members' recommendations (see Table 3).

Given the empirical findings, our preferred explanation is that the gender gap in alignment arises due to gender differences in deliberation behavior during the committee meeting.[19] In the following, we briefly discuss examples of behavioral gender differences that have been identified by experimental research and might be able to explain our real-world evidence.

First, male and female committee members could differ in their discussion behavior. In line with laboratory evidence, women could be less inclined to contribute and defend their opinions about a given candidate (e.g., Coffman, 2014; Hardt et al., 2022), or take leadership in the discussion (e.g., Born et al., 2022; Alan et al., 2019). These types of behavioral differences could be exacerbated by the minority status of women (e.g., Stoddard et al., 2023), or the desire

[18] One could argue that committee members can produce better evaluations within their own gender, i.e., they are better able to assess candidates of the same gender. However, if anything, we observe that the gender influence gap is larger for decisions on female candidates (see Table C.7).

[19] A further argument supporting this view is that recommendations by male and female evaluators are equally weighted when aggregated by an independent third person in the context of refereeing (Card et al., 2019). This shows that in the absence of deliberation, a gender gap in influence does not necessarily emerge.

to conform with gender norms (e.g., Bertrand et al., 2015; Burztyn et al., 2017). Thereby, they could also explain why the gender gap in influence differs with the gender composition of the committee.[20]

Second, women and men might respond differently to their colleagues' information on a given candidate. Experimental evidence shows that individuals generally tend to under-react to information collected by others (e.g., Conlon et al., 2022b). While men and women show similar patterns of under-reaction when information is provided by strangers, men under-react significantly more in the household context (Conlon et al., 2022a).[21] In the context of hiring committees, men might listen or respond less to the evaluations of their colleagues compared to women. Thus, they might be less willing to deviate from their initial assessments after hearing their colleagues' arguments.

Overall, the deliberation behavior of male and female committee members – in their roles as either communicators or receivers of information – could explain the observed gender differences in influence. Given our observational real-world data, we are unable to pin down the exact dimension of deliberation behavior that is driving the results. However, irrespective of the exact mechanism, the results suggest that behavioral factors, which are presumably unrelated to the quality of information, affect how information is weighted in committees.

# 6 Women in Committees and the Gender Balance of Hires

Many organizations use explicit or implicit gender quotas in hiring committees with the goal of improving the gender balance of their hires. This goal is also pursued by the organization that we study, where female candidates at the interview stage are about 17% less likely to receive a job offer. However, results from previous studies point towards zero or even negative effects of female quotas on the success of female candidates (e.g., Bagues et al., 2017; Deschamps,

[20] Similarly, one could also argue that male and female committee members differ in their preferences for competition or in how much they value to 'win' the deliberation and that these differences in preferences invoke different behaviors.

[21] More specifically, Conlon et al. (2022a) provide evidence that men – but not women – under-react to information collected by their spouse. One could argue that the relationship between colleagues shares features with the household context and is therefore prone to similar differences.

2022). One potential reason could be that women do not actually hold a more favorable view towards female candidates, which is a necessary prerequisite for a positive impact of women on committees. A second reason might be related to the results shown in the previous section: even if women assess women more favorably, their influence in the committee might not be sufficiently strong to provoke systematic changes in the final hiring decision, especially when committees are dominated by men.[22]

Given the structure of our data, we can separately analyze the committee members' individual evaluation behavior and the committee's final decision outcome. In the following, we first provide evidence on whether female committee members exhibit a smaller gender gap in their individual recommendations. We then examine how the gender gap in final offer rates varies with the representation of women at the committee level. The analysis relies on the quasi-random assignment of committee members to candidates on a given interview day, conditional on candidate's gender. In particular, Appendix Table A.3 shows that candidate characteristics do not predict the gender composition of the committee members –neither among male nor among female candidates.

**Individual Evaluation Behavior** Table 6 reports how the gender gap in the likelihood to receive a positive evaluation offer recommendation ($\geq 2$ points) differs between male and female committee members. In columns (1) to (4), we move from a specification without control variables to a specification with committee fixed effects and control variables for interviewer and candidate characteristics. All specifications reveal that male committee members are about 11 to 12 percentage points ($\approx 18\%$) less likely to give a positive recommendation to a female than to a male candidate. They also show that the gap is 5 to 6 percentage points (i.e., 50%) lower among female committee members. This also holds when introducing candidate fixed effects (column 5). Appendix Table D.8 and Table D.9 show that this result is robust to using different evaluation outcomes, such as the 3-point scale hiring recommendation or the average sub-rating of different ability dimensions. Overall, the results highlight the fact that –

[22] A further possible reason is that men might become less supportive of female candidates when women enter the committee (see suggestive evidence by Bagues et al., 2017; Deschamps, 2022).

Table 6: Gender Evaluation Gap Among Male and Female Committee Members

| | P(Positive Offer Recommendation) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Female Candidate | -0.108*** | -0.113*** | -0.115*** | -0.115*** | |
| | (0.014) | (0.015) | (0.015) | (0.015) | |
| Female Candidate x Female CM | 0.049** | 0.047** | 0.053** | 0.055** | 0.055** |
| | (0.023) | (0.023) | (0.023) | (0.023) | (0.024) |
| Committee FE | No | Yes | Yes | Yes | No |
| Candidate Characteristics | No | No | Yes | Yes | No |
| CM Characteristics | 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 | 8117 | 8117 | 8117 | 8117 | 8117 |

*Note:* CM=committee member. Control variables on committee member and candidate characteristics are reported in Panel A of Table 1 and Table A.1, respectively. Standard errors are clustered at the committee level (N=429). * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

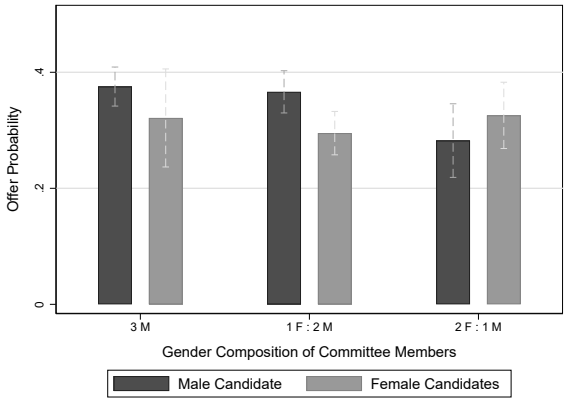
at the individual level – female committee members give significantly higher assessments to female candidates compared to their male colleagues.

**Committee Decisions** The evidence on individual recommendations supports the presumption that female committee members might be more favorable towards female candidates. While this is a necessary prerequisite for a positive impact of quotas on the gender balance of hires, it is unclear whether this difference in evaluations persists up to the final committee decision. In particular, the aggregation process might offset the individual-level differences for two main reasons: first, there is a mechanical barrier created by the fact that women are in the minority in most cases; and second, section 5 has documented a gender gap in influence, which is particularly pronounced when a committee is composed of one woman and two men (see Table 4). While these two factors cannot be distinguished empirically, we assess their joint relevance by comparing the gender gap in offers between decisions discussed by: *(i)* three men; *(ii)* one woman and two men; and *(iii)* two women and one man.[23]

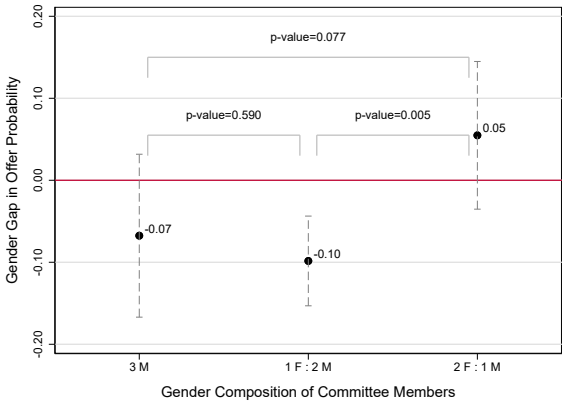
[23] Note that this analysis is subject to the caveat that relatively few female candidates (N=111) are evaluated only by men and few male candidates (N=199) are interviewed by two women.

Figure 8: Gender Composition of the Committee and Gender Balance in Hiring Decisions

(a) Raw Averages




(b) Gender Gap (with Controls)




*Note:* The unit of observation is the candidate. In panel (b), regressions include committee fixed effects and control for candidate characteristics reported in Table A.1. 95% confidence intervals, standard errors clustered at the committee level (N=429).

Figure 8 depicts the results. Panel (a) shows the raw offer probabilities of male and female candidates depending on the gender composition. Panel (b) shows the corresponding gender gaps, controlling for committee fixed effects and candidate characteristics. In cases where only male evaluators assessed a candidate, male candidates have a 39% probability of receiving a job offer, whereas the same probability is about 10 percentage points (i.e., 25%) lower for female candidates. When one male evaluator is replaced by a woman, this gap remains unaffected, in line with previous evidence (e.g., Bagues et al., 2017). In other words, a lone woman does not lead to a measurable change in the gender composition of new hires. The picture looks strikingly different when women are in the 2:1 majority. In these cases, the gender gap in offers closes (or even slightly reverses), which is driven by both a reduced offer rate for male candidates and an increased offer rate for female candidates. This strong non-linearity suggests that the committee's aggregation process plays a crucial role for the impact of committee diversification on hiring outcomes.

In Appendix Figure D.11, we replicate the same analysis using a candidate's predicted offer probability (based on pre-determined characteristics) as a placebo outcome. We observe no difference in the levels, nor in the gender gap, of this variable, suggesting that the differences in

hiring outcomes are not driven by any systematic selection of candidates to committee compositions.

# 7 Conclusion

Committee decision-making is ubiquitous in firms and organizations. This paper documents systematic gender differences in individuals' influence on the deliberation outcome of committees. Our results show that committee decisions are systematically less aligned with the initial recommendations of women than with those of men. This gender gap is particularly pronounced when deliberation occurs under a male majority, whereas cases deliberated under a female majority do not exhibit any gender gap in alignment. As women and men in our data are on average equally trained and experienced, and exhibit similar evaluation behavior, the findings point towards gender differences in deliberation behavior as the main explanation. Taken together, our findings reveal that previous experimental findings on gender differences in group behavior carry relevance for real-world decision processes with high stakes. At the same time, they raise new questions for future research on the exact dimensions of verbal and non-verbal communication behavior (e.g., tone of voice, body language, ...) that lead women to have a lower influence.

The results of this paper help to understand the process of opinion aggregation in groups and inform policies that try to increase diversity in high-level decision-making bodies. Systematic gender differences in influence can harm the effectiveness of such policies, which matters particularly when quotas introduce "lone women" on committees. Moreover, differences in influence that are unrelated to the quality of information can prevent committees from efficiently aggregating all available information.

Given the prevalence of group work and group decision-making in the labor market, awareness about potential gender differences in influence is important. However, the design of suitable policy responses is not trivial. While the avoidance of women's minority status seems a straightforward recommendation, it is not always feasible, particularly in environments where women represent a low share of the workforce. Moreover, such policies lead to inequalities in

the allocation of committee work, and can thereby exacerbate existing gender disparities in the allocation of non-promotable tasks (Babcock et al., 2017). Instead, organizations may want to reflect on procedural approaches that alter the nature of group discussions and encourage behaviors that allow everyone to have an equal voice in the decision-making process. Whether and how organizational changes can promote equality in influence in group decision-making is a question for future research.

# References


Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135.

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.

Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics, 130*(2), 571–614.

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.

Burztyn, L., Fujiwara, T., & Pallais, A. (2017). Acting wife: Marriage market incentives and labor market investments. *American Economic Review, 107*(11), 3288–3319.

Card, D., & DellaVigna, S. (2019). What do editors maximize? Evidence from four economics journals. *Review of Economics and Statistics, 102*, 195–217.

Card, D., DellaVigna, S., Funk, P., & Iriberri, N. (2019). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics, 135*(1), 269–327.

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.

Conlon, J., Mani, M., Rao, G., Ridley, M., & Schilbach, F. (2022a). Learning in the household. *mimeo*.

Conlon, J., Mani, M., Rao, G., Ridley, M., & Schilbach, F. (2022b). Not learning from others. *mimeo*.

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. (2022). Gender quotas in hiring committees: A boon or a bane for women? *mimeo*.

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. (2022). Overriding in teams: The role of beliefs, social image, and gender. *Management Science*, *forthcoming*.

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. (2022). Who does the talking here? The impact of gender composition on team interactions. *mimeo*.

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.

Isaksson, S. (2019). It takes two: Gender differences in group work. *mimeo*.

Jacobi, T., & Schweers, D. (2017). Justice, interrupted: The effect of gender, ideology, and seniority at supreme court oral arguments. *Va. L. Rev.*, *103*, 1379.

Koffi, M. (2021). Innovative ideas and gender inequality. *mimeo*.

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.

Roussille, N. (2022). The central role of the ask gap in gender pay inequality. *mimeo*.

Sarsons, H., Gërxhani, K., Reuben, E., & Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, *129*(1), 101–147.

Stoddard, O., Karpowitz, C., & Preece, J. (2023). Strength in numbers? Gender Composition, leadership, and women's influence in teams. *mimeo*.

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)

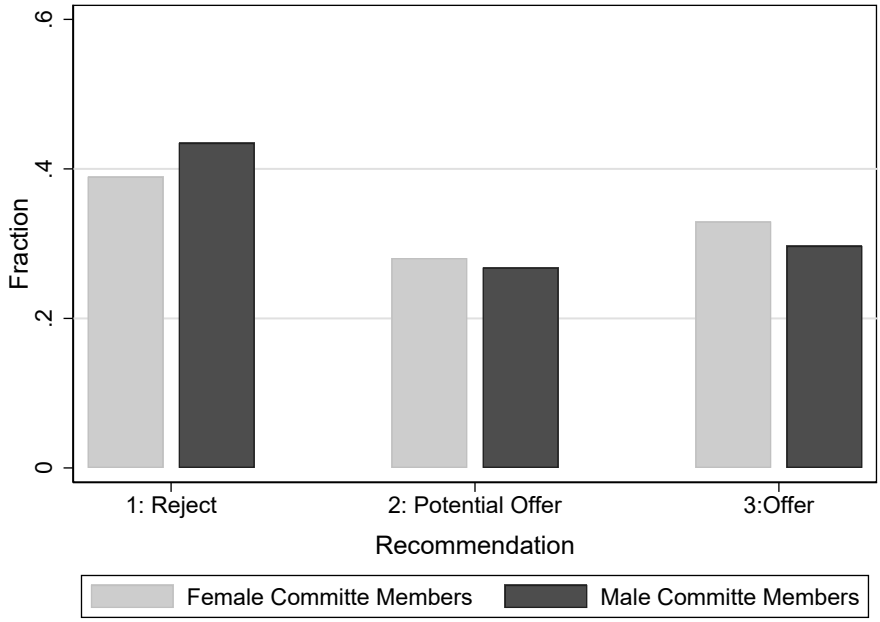
## A Additional Material: Data

Table A.1: Summary Statistics on Candidate Characteristics

| | N | Mean | SD |
|---|---|---|---|
| Female | 2913 | 0.40 | 0.49 |
| Study Field: Chemistry | 2913 | 0.02 | 0.14 |
| Study Field: IT | 2913 | 0.01 | 0.10 |
| Study Field: Engineering | 2913 | 0.10 | 0.30 |
| Study Field: Law | 2913 | 0.03 | 0.16 |
| Study Field: Life Sciences | 2913 | 0.06 | 0.24 |
| Study Field: Math/Physics | 2913 | 0.03 | 0.18 |
| Study Field: Other Nat. Science | 2913 | 0.02 | 0.14 |
| Study Field: Soc. Sciences/Humanities | 2913 | 0.05 | 0.22 |
| Study Field: Economics/Business | 2913 | 0.58 | 0.49 |
| Study Field: Unknown | 2913 | 0.02 | 0.14 |
| Internship Application | 2913 | 0.33 | 0.47 |
| High GPA (overall) | 2913 | 0.36 | 0.48 |
| High GPA (math) | 2913 | 0.23 | 0.42 |
| High CV Score | 2913 | 0.63 | 0.48 |

*Note:* The unit of observation is the candidate. "High"= above median.

Figure A.1: Recommendations by Male and Female Committee Members



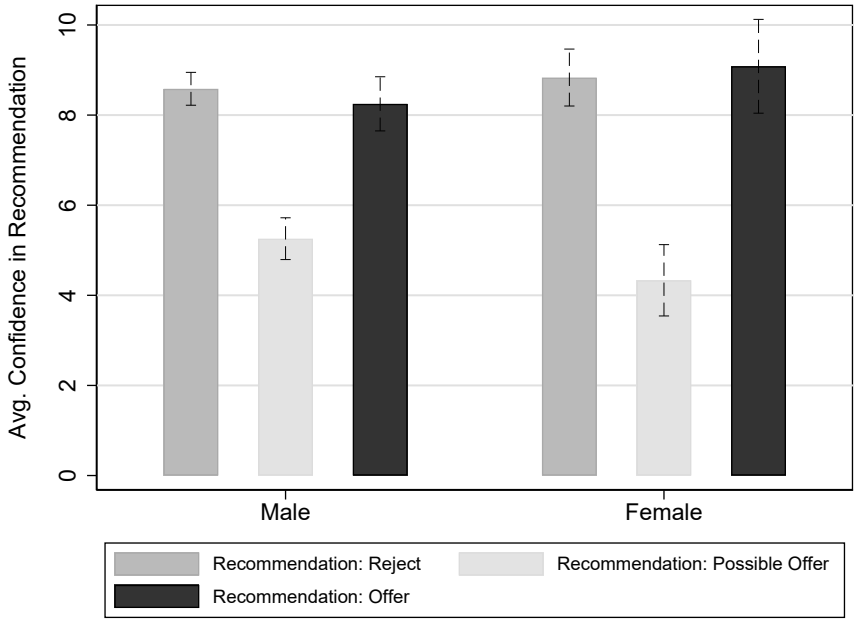

*Note:* The unit of observation is the interview recommendation (N=8,117).

Table A.2: Correlation between Predicted Offer Probability and Recommendations

| | Offer Recommendation | |
|---|---|---|
| | (1) | (2) |
| Predicted P(Offer) | 1.378*** | 1.318*** |
| | (0.175) | (0.188) |
| Female Committee Member | 0.085 | 0.062 |
| | (0.088) | (0.086) |
| Predicted P(Offer) × Female Committee Member | -0.028 | -0.012 |
| | (0.259) | (0.254) |
| Committee FE | No | Yes |
| Outcome Mean | 1.88 | 1.88 |
| N | 7103 | 7103 |

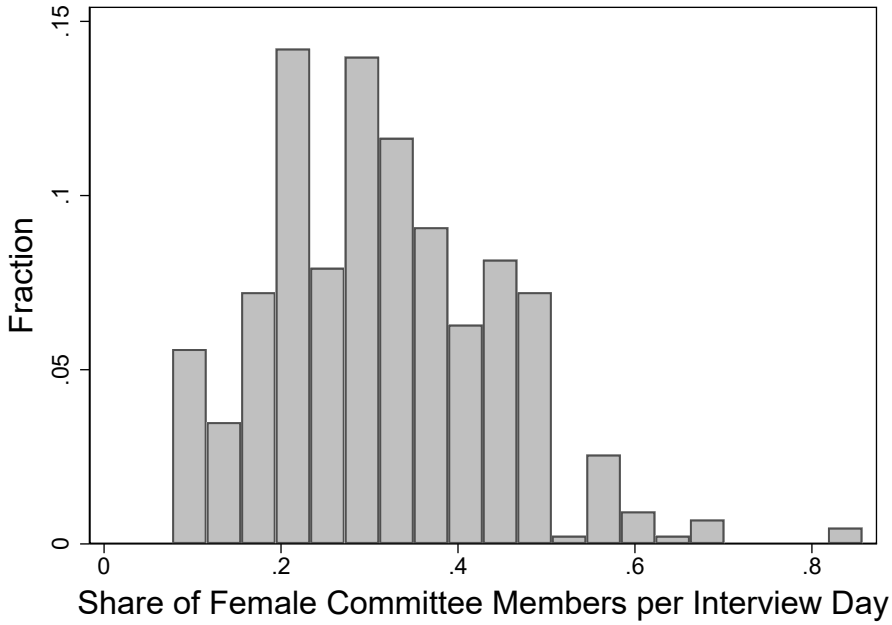
*Note:* The offer probability is predicted based on a candidate's CV screening scores, and high-school GPA (overall and maths). When estimating the coefficients of these variables, we leave out the candidate's own interview day. The number of observations is lower than in the main analysis due to missing values in the CV screening scores. Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Figure A.2: Committee Members' Confidence in Recommendations



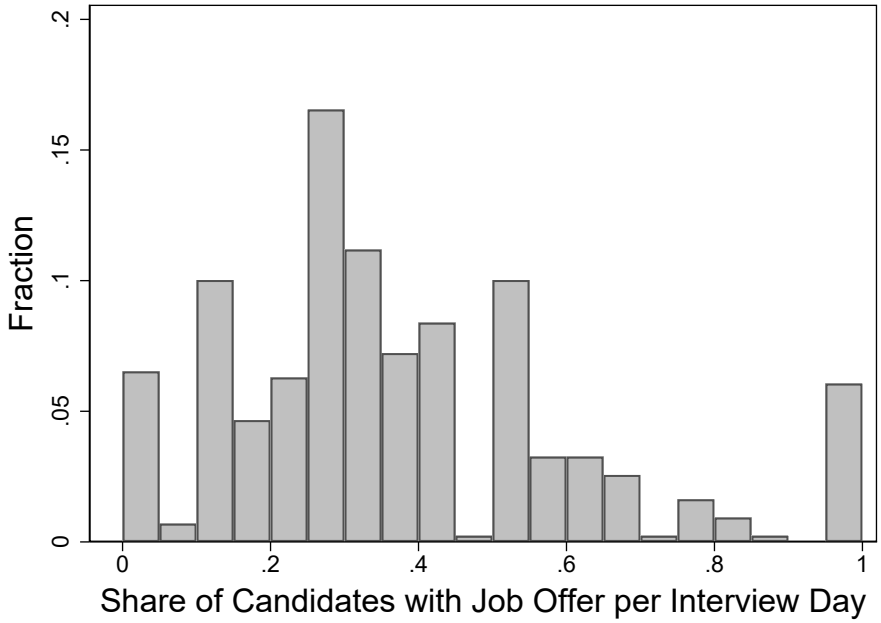

*Note:* The figure shows responses to questions on committee members' average confidence in their recommendations ('When recommending a rejection/potential offer/offer, how confident are you on average that the candidate should (not) be hired?'). Confidence is expressed on a 1-10 scale. The questions 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%.

Figure A.3: Share of Female Committee Members per Interview Day




*Note:* The unit of observation is the interview day (N=429).

Figure A.4: Share of Candidates with a Job Offer per Interview Day




*Note:* The unit of observation is the interview day (N=429).

Table A.3: Assignment of Candidates to Female Committee Members: Candidate-Level

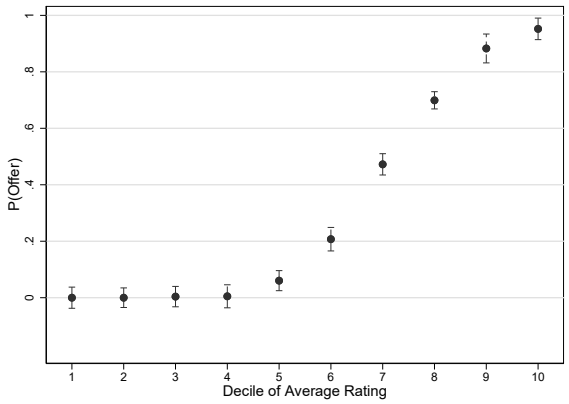
| | Number of Female CM | | | P(Majority of Female CM) | | |
|---|---|---|---|---|---|---|
| | (1) All | (2) Male Cand. | (3) Fem. Cand. | (4) All | (5) Male Cand. | (6) Fem. Cand. |
| Female | 0.740*** | | | 0.152*** | | |
| | (0.044) | | | (0.018) | | |
| Field: Business | -0.044 | 0.005 | -0.141 | 0.002 | 0.019 | -0.036 |
| | (0.061) | (0.086) | (0.088) | (0.023) | (0.031) | (0.038) |
| Field: STEM | -0.029 | 0.007 | -0.154 | -0.002 | -0.003 | -0.042 |
| | (0.063) | (0.088) | (0.105) | (0.024) | (0.031) | (0.043) |
| Internship Application | 0.061 | -0.025 | 0.051 | 0.029 | -0.011 | 0.039 |
| | (0.063) | (0.087) | (0.099) | (0.025) | (0.030) | (0.045) |
| High GPA | 0.026 | 0.001 | 0.067 | 0.016 | 0.014 | 0.025 |
| | (0.041) | (0.059) | (0.064) | (0.016) | (0.022) | (0.028) |
| High GPA in Math | -0.066 | -0.071 | 0.061 | -0.023 | -0.040* | 0.042 |
| | (0.050) | (0.062) | (0.092) | (0.020) | (0.024) | (0.039) |
| High CV Score | -0.007 | -0.012 | 0.070 | -0.027 | -0.021 | 0.005 |
| | (0.043) | (0.060) | (0.064) | (0.017) | (0.020) | (0.028) |
| Committee FE | Yes | Yes | Yes | Yes | Yes | Yes |
| p-value joint significance (excl. gender) | 0.77 | 0.90 | 0.36 | 0.49 | 0.38 | 0.41 |
| Outcome Mean | 1.18 | 0.91 | 1.58 | 0.17 | 0.12 | 0.26 |
| N | 2913 | 1750 | 1163 | 2913 | 1750 | 1163 |

*Note:* CM=Committee Member. The unit of observation is the candidate. Standard errors are clustered at the committee level (N=429). $^{*}p<0.10,^{**}p<0.05,^{***}p<0.01$.

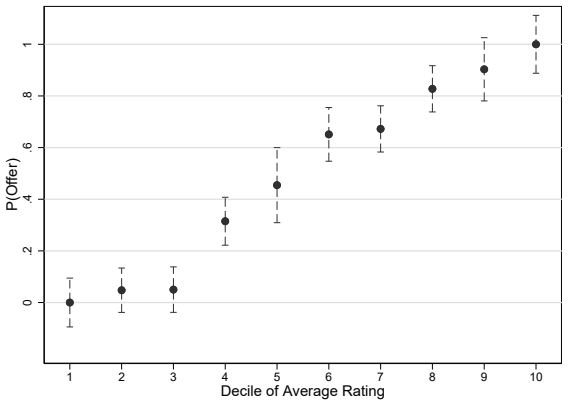
# B Additional Material: Aggregation of Information

Figure B.5: Aggregation of Information in the Committee

(a) Decisions w/ 3 Recommendations




(b) Decisions w/ 2 Recommendations




*Note:* The unit of observation is the candidate. N=2,433 in panel (a), N=480 in panel (b). The x-axis shows deciles of the candidate's average rating from all interviews. The left y-axis shows the probability that a given combination results in a positive offer decision by the committee. Dashed lines show 95% confidence intervals, with standard errors clustered at the committee level (N=429).

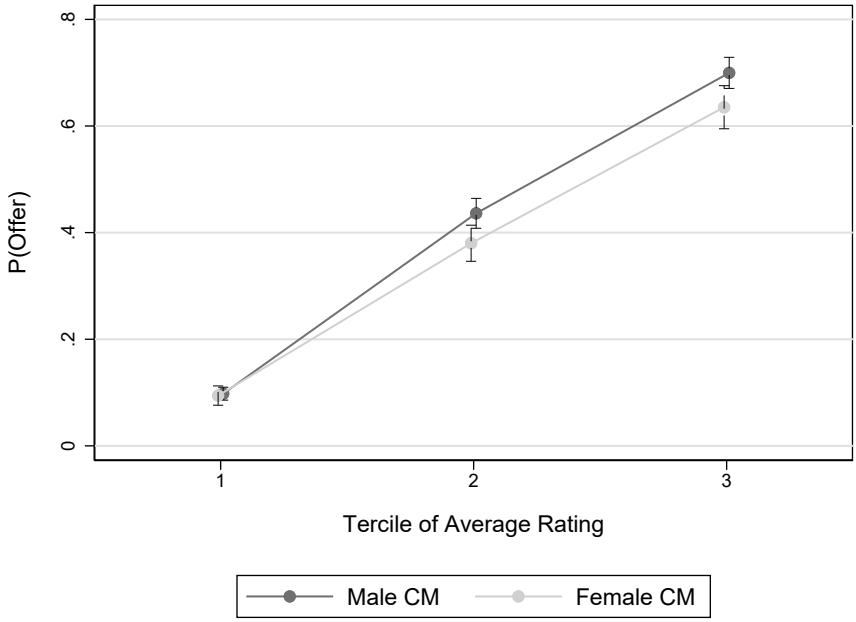
# C Additional Material: Gender Differences in Influence

Table C.4: Translation of Recommendations into an Offer Decision: Linear Regressions

| | P(Offer) | |
|---|---|---|
| | (1) | (2) |
| Recommendation (scale 1-3) | 0.295*** | 0.295*** |
| | (0.007) | (0.007) |
| Female | 0.007 | |
| | (0.020) | |
| Recommendation x Female | -0.023** | -0.026** |
| | (0.012) | (0.012) |
| Committee Member FE | No | Yes |
| Outcome Mean | 0.33 | 0.33 |
| N | 8117 | 8117 |

*Note:* Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

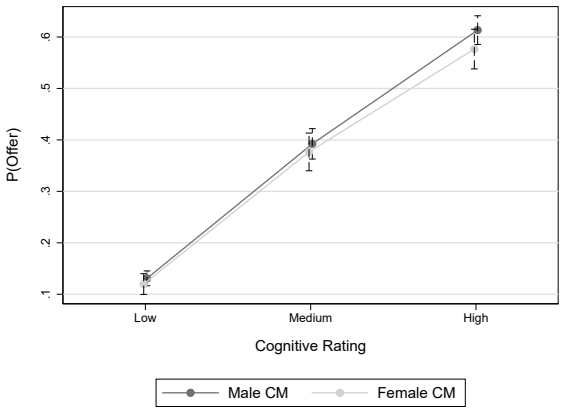
Figure C.6: Translation of Average Interview Ratings into Offer Decisions



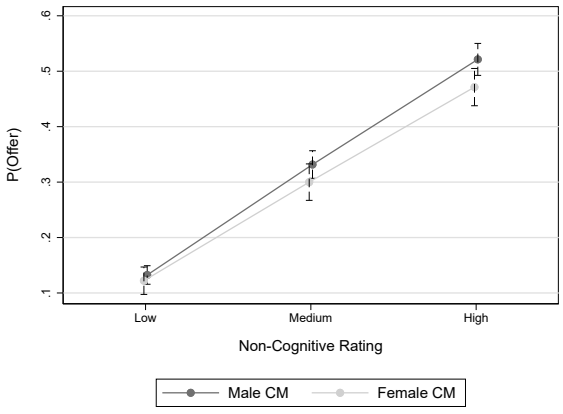

*Note:* CM = Committee Member. N=8,117. 95% confidence intervals, standard errors clustered at the committee level (N=429).

Figure C.7: Translation of Cognitive and Non-Cognitive Ratings into Offer Decisions

(a) Cognitive Rating




(b) Non-Cognitive Rating




*Note:* CM = Committee Member. N=8,117. Low: < 2 points; Medium: 2 points; High: > 2 points. 95% confidence intervals, standard errors clustered at the committee level (N=429).

Table C.5: Gender Difference in Influence, Depending on Disagreement

| | Alignment Probability | |
|---|---|---|
| | (1) | (2) |
| Female | -0.004 | -0.000 |
| | (0.013) | (0.013) |
| Female × Level of Disagreement (in Points) | -0.026*** | -0.028*** |
| | (0.008) | (0.009) |
| Committee FE | No | Yes |
| Basic Controls | No | Yes |
| Control for Disagreement | No | Yes |
| Control for Leniency | No | Yes |
| Outcome Mean | 0.68 | 0.68 |
| N | 7159 | 7159 |

*Note:* Estimates are based on equation 1. The analysis only includes candidates with three recommendations. Level of disagreement = sum of the pair-wise absolute differences between a member's own recommendations and the other two members' recommendations on the candidate. Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table C.6: Robustness of Main Estimates

| | Alignment Probability | |
|---|---|---|
| | (1) | (2) |
| *Panel A: Baseline* | | |
| Female | -0.042*** | -0.023** |
| | (0.010) | (0.010) |
| Female x Deliberation Case | | -0.055** |
| | | (0.022) |
| Outcome Mean | 0.68 | 0.68 |
| N | 8117 | 8117 |
| *Panel B: Excluding potential offer recommendations* | | |
| Female | -0.026** | -0.001 |
| | (0.010) | (0.006) |
| Female x Deliberation Case | | -0.084*** |
| | | (0.028) |
| Outcome Mean | 0.81 | 0.81 |
| N | 5908 | 5908 |
| *Panel C: Excluding candidates with final interviews* | | |
| Female | -0.031*** | -0.019* |
| | (0.010) | (0.011) |
| Female x Deliberation Case | | -0.046* |
| | | (0.025) |
| Outcome Mean | 0.69 | 0.69 |
| N | 6858 | 6858 |
| *Panel D: Excluding candidates with two recommendations* | | |
| Female | -0.042*** | -0.022** |
| | (0.011) | (0.011) |
| Female x Deliberation Case | | -0.060** |
| | | (0.025) |
| Outcome Mean | 0.68 | 0.68 |
| N | 7159 | 7159 |
| *Panel E: Estimation with candidate FE* | | |
| Female | -0.039*** | -0.024** |
| | (0.011) | (0.012) |
| Female x Deliberation Case | | -0.042* |
| | | (0.025) |
| Outcome Mean | 0.68 | 0.68 |
| N | 8117 | 8117 |

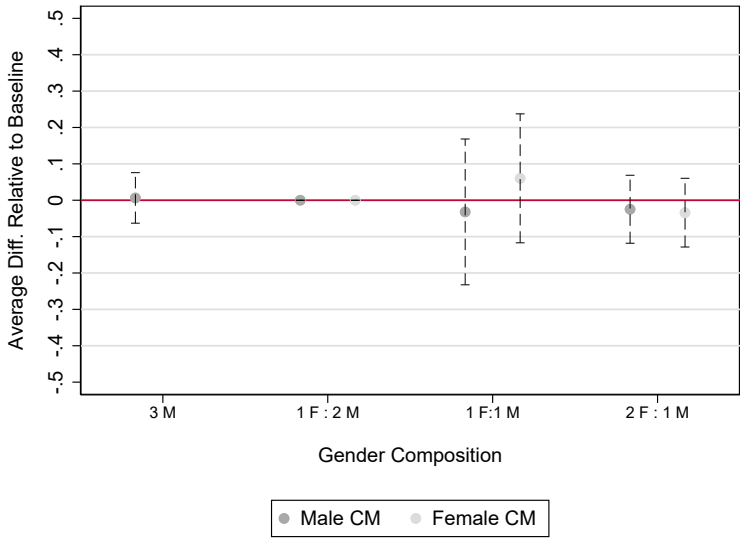
*Note:* Estimates are based on equation 1. All regressions include committee fixed effects and the full set of control variables described in section 5.2 (candidate & committee member characteristics, interview order, leniency, disagreement). Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table C.7: Heterogeneity by Committee Member and Candidate Characteristics

| | Alignment Probability | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Female | -0.055*** | -0.033** | -0.066*** | -0.023* | -0.055*** |
| | (0.012) | (0.014) | (0.019) | (0.014) | (0.012) |
| Female × Manager | 0.026 | | | | |
| | (0.020) | | | | |
| Female × Above Median Experience | | -0.028 | | | |
| | | (0.020) | | | |
| Female × Primary Interview Location | | | 0.028 | | |
| | | | (0.021) | | |
| Female × Female Candidate | | | | -0.052*** | |
| | | | | (0.020) | |
| Female × Internship Application | | | | | 0.029 |
| | | | | | (0.019) |
| Outcome Mean | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 |
| N | 8117 | 8117 | 8117 | 8117 | 8117 |

*Note:* Estimates based on equation 1. All regressions include committee fixed effects and the full set of control variables described in section 5.2 (candidate & committee member characteristics, interview order, leniency, disagreement). Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

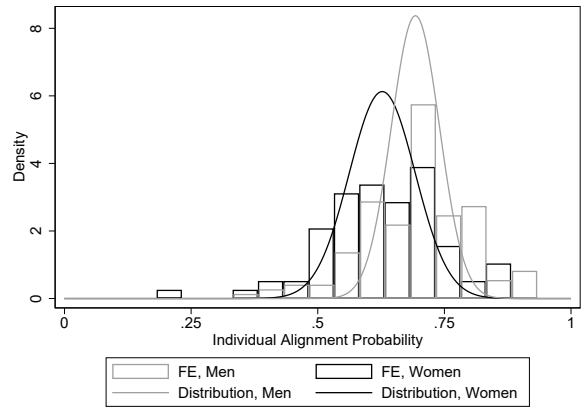
Figure C.8: Do Initial Recommendations React to the Gender Composition?



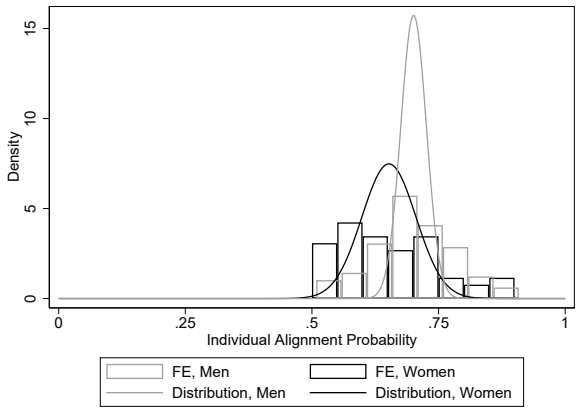

*Note:* CM = Committee member. The underlying regression tests whether committee members' recommendations react to the gender composition of the other two committee members assessing the same candidate. Regressions include committee fixed effects, as well as controls for committee member characteristics, candidate characteristics, and the order of the interview. 95% confidence intervals, with standard errors clustered at the committee level (N=429).

Figure C.9: Individual Fixed Effects and Distribution of Influence by Gender: Robustness

(a) At least 10 interviews per committee member

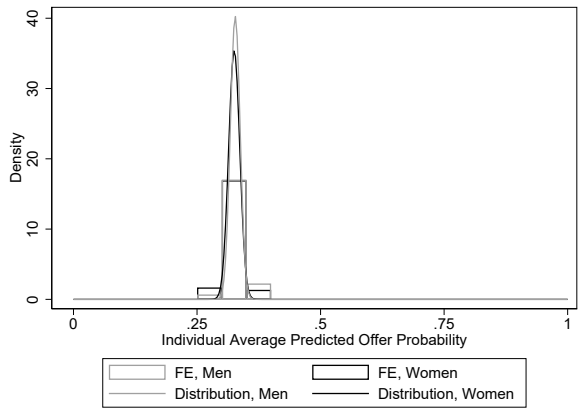



(b) At least 20 interviews per committee member




*Note:* The figure shows a histogram of the individual fixed effects and the estimated parametric normal distributions in analog to Figure 7, but with varying minimum cutoffs for the number of interviews conducted by each committee member during the data period.

Figure C.10: Individual Fixed Effects and Distribution of Influence by Gender: Placebo Outcome




*Note:* The figure replicates Figure 7, using the candidate's predicted offer probability as a placebo outcome. The prediction is based on a candidate's CV screening scores, high-school GPA, and math GPA. When estimating the coefficients of these variables, we leave out the candidate's own interview day.

# D Additional Material: Women in Committees and the Gender Balance of Hires

Table D.8: Gender Evaluation Gap Among Male and Female Committee Members

| | Offer Recommendation (3-point scale) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Female Candidate | -0.181*** | -0.199*** | -0.200*** | -0.199*** | |
| | (0.024) | (0.026) | (0.026) | (0.026) | |
| Female Candidate x Female CM | 0.067* | 0.068* | 0.079* | 0.082** | 0.060 |
| | (0.041) | (0.041) | (0.041) | (0.041) | (0.042) |
| Committee FE | No | Yes | Yes | Yes | No |
| Candidate Characteristics | No | No | Yes | Yes | No |
| CM Characteristics | No | No | No | Yes | Yes |
| Candidate FE | No | No | No | No | Yes |
| Outcome Mean | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 |
| N | 8117 | 8117 | 8117 | 8117 | 8117 |

*Note:* CM=committee member. Control variables on committee member and candidate characteristics are reported in Panel A of Table 1 and Table A.1, respectively. Standard errors are clustered at the committee level (N=429). * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

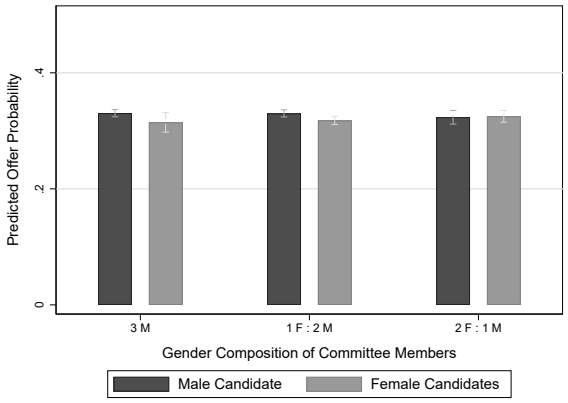
Table D.9: Gender Evaluation Gap Among Male and Female Committee Members

| | Avg. Sub-Rating (3-point scale) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Female Candidate | -0.099*** | -0.108*** | -0.108*** | -0.107*** | |
| | (0.011) | (0.012) | (0.012) | (0.012) | |
| Female Candidate x Female CM | 0.036** | 0.035** | 0.041** | 0.043*** | 0.034** |
| | (0.016) | (0.016) | (0.016) | (0.016) | (0.016) |
| Committee FE | No | Yes | Yes | Yes | No |
| Candidate Characteristics | No | No | Yes | Yes | No |
| CM Characteristics | No | No | No | Yes | Yes |
| Candidate FE | No | No | No | No | Yes |
| Outcome Mean | 1.94 | 1.94 | 1.94 | 1.94 | 1.94 |
| N | 8117 | 8117 | 8117 | 8117 | 8117 |

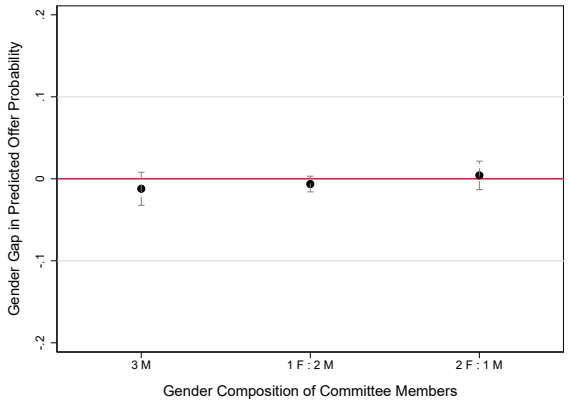
*Note:* CM=committee member. Control variables on committee member and candidate characteristics are reported in Panel A of Table 1 and Table A.1, respectively. Standard errors are clustered at the committee level (N=429). $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Figure D.11: Placebo Checks: Gender Composition of the Committee and Gender Balance in Hiring Decisions

(a) Raw Averages




(b) Gender Gap (with Controls)




*Note:* The unit of observation is the candidate. The figure replicates Figure 8, using a candidate's predicted offer probability as a placebo outcome. The prediction is based on a candidate's CV screening scores, high-school GPA, and math GPA. When estimating the coefficients of these variables, we leave out the candidate's own interview day. In panel (b), regressions control for committee fixed effects. 95% confidence intervals, standard errors clustered at the committee level (N=429).