

“Potential” and the Gender Promotion Gap*

Alan Benson
Univ. of Minnesota

Danielle Li
MIT & NBER

Kelly Shue
Yale & NBER

October 15, 2021

Please see [here](#) for latest version

Abstract

We show that widely-used subjective assessments of employee “potential” contribute to gender gaps in promotion and pay. Using data on 30,000 management-track employees from a large retail chain, we find that women receive substantially lower potential ratings despite receiving higher job performance ratings. Differences in potential ratings account for 30-50% of the gender promotion gap. Women’s lower potential ratings do not appear to be based on accurate forecasts of future performance: women outperform male colleagues with the same potential ratings, both on average and on the margin of promotion. Yet, even when women outperform their previously forecasted potential, their subsequent potential ratings remain low, suggesting that firms persistently underestimate the potential of their female employees.

JEL Classifications: M51 (Firm employment decisions; promotions); J71 (Discrimination); J31 (Wage level and structure; wage differentials)

Keywords: Promotions, performance evaluations, glass ceiling, gender bias, leadership, compensation, wage differentials, role congruity theory, Peter Principle

*Alan Benson: bensona@umn.edu. Danielle Li: d_li@mit.edu. Kelly Shue: kelly.shue@yale.edu. We thank Yi Li for excellent research assistance and the International Center for Finance at the Yale School of Management for their support. We thank seminar audiences at University of Alberta, CUHK, INSEAD, Minnesota, MIT, NBER Personnel Economics, SIOE, Southern Methodist University, and Yale for helpful comments.

1 Introduction

When making promotion decisions, firms must form predictions about the future performance of each employee: if given the opportunity, would someone make a good manager? To guide this assessment, firms may use information about a worker’s job performance, but such information is imperfect both because people can change over time and because management roles often require a different skillset (Peter and Hull, 1969; Benson, Li and Shue, 2019). Ultimately, then, firms are left to make inferences about a worker’s “potential.” Yet because potential can never be directly observed, these assessments can be highly subjective, leaving room for bias.

Promotion decisions that rely on assessments of potential may negatively impact the careers of women in particular. Research on role congruity theory (Eagly and Karau, 2002) presents broad evidence that people may have trouble imagining women in leadership positions because the qualities stereotypically associated with effective leaders—e.g. competitiveness, ambition, and an orientation toward execution—are incongruous with the qualities stereotypically associated with women.¹ Research by Player et al. (2019), for instance, finds that experimental subjects assess greater leadership potential and forecast higher future performance for male applicants based on otherwise identical resumes. Indeed, the simple fact that women are less frequently observed in managerial roles may make it more difficult to overcome these stereotypes (Blau and Kahn, 2017). Finally, women may engage less in self-promotion, networking, or negotiation (Cullen and Perez-Truglia, 2020; Fang and Huang, 2017; Roussille, 2020; Babcock and Laschever, 2009; Biasi and Sarsons, 2020)²—choices which may translate into disadvantages when it comes to subjective assessments that are influenced by favoritism and politicking (Prendergast and Topel, 1993).

In this paper, we show that subjective assessments of worker potential contribute to gender gaps in promotion, and to an inefficient allocation of talent across roles. We study promotions among 29,809 management-track workers within a large North American retail chain. Our firm uses a popular assessment and succession planning tool known as a “Nine Box” grid, in which supervisors rate subordinates on two dimensions: their current performance and their future potential. These dimensions take three values (1-low, 2-medium, and 3-high), creating a 3-by-3 matrix with nine cells. Whereas performance ratings are intentionally backward looking and often based on demonstrable achievements, potential ratings are forecasts of a worker’s future performance and contributions to the firm, making them fundamentally more subjective. Beyond their use at our firm, Nine Box and

¹See also Bursztyn, Fujiwara and Pallais (2017); Koenig et al. (2011); Proudfoot, Kay and Koval (2015); Correll et al. (2020); and Kaplan, Klebanov and Sorensen (2012).

²These choices could be due to lower access to such activities and expectations of lower benefits from engaging in such activities (e.g., Abraham (2020)).

similar assessments of potential are nearly ubiquitous in large organizations, where they often play a major role in determining promotions, developmental opportunities, and compensation.³

Our paper has three main findings. First, women receive lower potential ratings despite earning higher performance ratings. This gap in potential ratings accounts for up to half of the overall gender gap in promotions. Second, potential ratings appear to be biased against women. Among workers with the same current performance and potential ratings, women receive higher performance ratings in the future. This is true for our full sample of workers, as well as for the subset who are promoted into new roles. Our analysis of promoted workers shows that women receive higher ratings of their future performance as managers both on average, and at the margin of promotion. Finally, we identify a key trade-off between information and equity. We show that potential ratings, although biased, are nevertheless informative about future performance. Rather than ignoring them, firms may want to invest in solutions that reduce bias in potential ratings, in order to retain their information content.

We motivate our analysis by documenting a substantial decline in female representation as workers climb the career ladder. In our firm, women constitute 56% of entry level field workers, but only 48% of department managers, 35% of store managers, and 14% of district managers. These patterns are consistent with the “glass ceiling” effect, whereby gendered barriers to promotion intensify and yield diminishing shares of women in senior jobs (Blau and Kahn, 2017). Because salaries are closely tied to job levels, gender differences in job levels account for approximately 70% of the overall gender wage gap in our data. This result is consistent with Petersen and Saporta (2004), which finds that the gender wage gap in the United States largely arises from the assignment of jobs rather than wage discrimination within jobs.

Consistent with these aggregate differences, we find a robust gender gap in promotions at the individual level: women are 14% less likely to be promoted relative to the sample average. This baseline gender gap in promotions cannot be explained by differences in past performance: women receive higher performance ratings on average and are 7.3% more likely earn the top performance rating, relative to the base rate.

We show that the gender gap in promotions is better explained by differences in forecasts of potential. Women in our firm receive 8.3% lower potential ratings than men on average, despite their higher performance ratings, and despite the fact that performance and potential ratings are positively correlated overall. At our firm, potential ratings strongly predict promotions. An increase from “medium” to “high” potential ratings corresponds to a 75% increase in the likelihood of promotion relative to the base rate, compared to only a 27% increase for the equivalent increase

³Cappelli and Keller (2014) and Church et al. (2015) discuss the history of Nine Box, its widespread adoption, and other assessments of employee potential.

in performance ratings. Taken together, we find that gender differences in potential ratings can explain up to 50% of the overall gender promotion gap.

Practitioners describe potential as an individual’s ability to contribute to the firm in the future, either through improved performance and greater responsibilities in her original job role or through leadership in a new managerial role (Cappelli and Keller, 2014; Groysberg and Nohria, 2011; Silzer and Church, 2009; Yarnall and Lucy, 2015). Women’s lower potential ratings, therefore, may be justified if they have lower future performance. We show that this is not the case: relative to men with the same current period Nine Box ratings for performance and potential, women earn higher performance scores in their next evaluation. That is, to receive the same potential ratings as their male colleagues, women’s realized future performance must be higher on average.

This result holds whether we consider the worker’s future performance in the same role or after promotion to a higher-level role. Using an IV approach based on variation in the availability of promotion opportunities over time and across roles, we find that marginally promoted women have higher future performance ratings, relative to marginally promoted men with the same initial potential ratings. This finding is indicative of misallocation: our firm could improve the quality of its managers by favoring women in circumstances where men and women receive the same Nine Box ratings. Equivalently, it could increase the potential ratings of women on the margin of promotion.

We do not find evidence that managers update their evaluations of women’s potential upon observing that women outperform men with similar prior potential scores. Even in such cases, we show that women continue receiving lower potential ratings in their next evaluation. Indeed, we find that the gender gap in potential ratings widens at higher levels of the organizational hierarchy. This fact implies that the gender gap in promotion rates grows not just from the compounding of a constant gender gap in potential scores, but that its growth accelerates as women advance in their careers.

One may be concerned that performance ratings do not fully capture a worker’s ability to contribute to the firm in the future. In particular, if women are more likely to leave the firm, then their potential contributions may be lower even if they outperform men while still employed. We next consider whether women’s lower potential ratings can be accounted for by higher rates of attrition or leave-taking (Tô, 2018; England et al., 2016). Leaving aside the legality and ethics of such considerations, we find evidence suggesting that such beliefs are not justified by the data. Women in our data are actually significantly less likely to leave the firm than men, and while they are more likely to take leaves of absence, the absolute levels of leave among women are too low to explain the large gap in potential ratings and promotions.

Interestingly, we find suggestive evidence of a different story: men receive higher potential ratings precisely because they are more likely to leave the firm. Specifically, we show that men are

far more likely to leave the firm when they are passed over for promotions, or when they receive lower potential ratings, consistent with evidence from [Blackaby, Booth and Frank \(2005\)](#). Our firm is aware of this, accurately assessing men to be at a greater “risk of loss.” Rather than viewing attrition risk as a sign that a worker is less likely to contribute to the firm in the future, our firm appears to reward at-risk workers with higher potential ratings, pay, and promotions. Taken together, it appears that the firm grants higher potential scores to men who are less likely to perform well in the future and more likely to leave the firm altogether.

A remaining possibility is that women have lower potential ratings because they are less likely to accept promotions or otherwise seek out more challenging roles ([Fernandez and Mors, 2008](#); [Azmat, Cuñat and Henry, 2020](#)). While we do not directly observe offers of promotion or internal applications, we provide suggestive evidence that our results are not explained by this possibility. Specifically, we continue to find gender gaps in subsamples where promotions are less likely to conflict with childcare priorities: among older workers who are less likely to have young children at home, and for promotions that do not involve a change in geographic location.

Having established that promotion practices based on assessments of potential likely systematically disadvantage women, we consider two sets of potential remedies: changes in the managers that provide potential ratings and changes to potential ratings themselves.

First, we ask whether firms can help female employees by changing their manager assignment. It is commonly suggested that women’s outcomes can be improved by assigning them to either female or “star” managers, under the logic that female managers may be more effective in mentoring female subordinates and high performing managers may be better able to evaluate their subordinates in an unbiased manner. Our data contain information on the demographics and own Nine Box ratings of each worker’s manager, allowing us to evaluate how manager characteristics correlate with their subordinate job outcomes. In these analyses, however, we caution that we cannot distinguish between treatment and selection effects; that is, subordinate outcomes may differ across managers because managers are assigned to different types of subordinates or because managers differ in how they evaluate their subordinates.

We find that gender gaps in potential, pay, and promotions are smaller under female managers. Female managers, however, are associated with lower overall potential ratings, pay, and promotion rates for their subordinates of both genders, suggesting that female managers are either assigned to weaker subordinates or less effective in advocating for them to receive stronger ratings. Conversely, we find that gender gaps in potential, pay, and promotions are larger under more highly-rated managers, but that the overall levels of these outcomes are also higher under more highly-rated managers. Taken together, the opposing level and interaction effects in both cases imply that

female subordinates are not clearly better off working under either female managers or highly-rated managers.

Second, we ask whether firms can improve promotion outcomes by varying how potential scores are used or assigned. We consider the following counterfactuals: ignoring potential scores and gender altogether, and “de-biasing” potential scores by making gender-specific adjustments. We show that ignoring potential and gender would nearly eliminate the gender promotion gap, but also decrease the average future performance of workers who are selected to be promoted. This result reflects the fact that, despite their bias, potential ratings contain useful information about future performance.

We also consider a policy that retains information on both potential and gender by increasing the potential ratings of women who receive the highest performance rating. We find that this approach eliminates the gender promotion gap while also increasing the predicted future performance of promoted workers. While this simple policy may be challenging to implement in practice (for instance, managers may shade female potential ratings down in anticipation of this gender-specific bonus), it suggests that firms stand to gain from finding ways to de-bias their otherwise informative assessments of potential.

Taken together, our results show that firms face difficult trade-offs related to the objectivity, informativeness, and equity of worker evaluations. Promotions based solely on past performance may be less subjective but are not the most informative.⁴ This is consistent with research on the “Peter Principle,” which has shown that performance in one’s current role is a highly imperfect predictor of one’s potential to succeed as a manager (e.g., [Benson, Li and Shue, 2019](#)). Reducing bias in assessments of potential could improve both the equity and efficacy of promotion decisions.

2 Background

2.1 Setting

Our data come from the U.S. operations of a large retailer from February 2009 to October 2015. Over this period, the firm employed over one million workers, nearly one percent of the US labor force, primarily in entry-level hourly roles (e.g. cashiers, sales, customer support, and material handling). Our analysis focuses on the firm’s core salaried, full-time employees, nearly all of whom are evaluated using Nine Box ratings. Our main sample therefore consists of 29,809 management-track workers, spread across the firm’s core retail operations and corporate headquarters.

⁴Unlike future performance, past performance is partly observable by managers, making it in principle less subjective. Nevertheless, evaluations of past performance have also been shown to be biased against women, see e.g., [Sarsons \(2017b\)](#), [Sarsons et al. \(2021\)](#), and [Cziraki and Robertson \(2021\)](#). If performance ratings are indeed biased against women, it is all the more striking that women in our data earn higher performance ratings than men and still earn lower potential ratings.

Employees in our firm’s corporate headquarters perform a variety of professional functions, the largest of which are in information technology, supply chain management, finance, human resources, and real estate management. Career ladders follow a traditional system of pay grades nested within bands. 40% of corporate workers with Nine Box scores are categorized as individual contributors, 40% are managers, and 20% are directors and executives. Although workers receive regular raises, large raises ultimately require workers to be promoted.

Employees in our firm’s direct retail operations work at one of over 4,000 establishments. Most employees work in a store that is led by a head store manager and a team of department managers. Store managers report to a senior manager who covers all stores of a given format in one of the country’s 37 districts. Store managers are primarily responsible for leadership and management activities, including analyzing data, formulating a strategy, and inspiring others to successfully execute that strategy. Store managers are assessed on their ability to achieve performance goals but are otherwise given wide latitude in how to achieve them. Department managers are responsible for efficiently executing the strategy set out by the store manager. This includes customer-facing duties and supervisory duties such as the hiring and coaching of entry-level staff.

2.2 Nine Box evaluation

Employees in our data are evaluated using a Nine Box grid, a widespread talent assessment and succession planning tool that instructs supervisors to categorize their subordinates into one of nine boxes representing the interaction of a three-level rating (high, medium, low) on two dimensions: the worker’s prior job performance and worker’s future potential.

The performance dimension of Nine Box scores is thought of as a backward-looking assessment of workers’ achievements in their current roles. For instance, store managers may be evaluated on whether their departments met sales targets, replenishment managers may be evaluated on meeting inventory level and delivery targets, and loss prevention managers may be evaluated on inventory lost to theft or damage. In contrast, the potential dimension of Nine Box is a forward-looking assessment. While there is little formal definition for how to define potential, practitioner guides focus on a worker’s capacity for growth within the same role or within the same organization in a different role ([Silzer and Church, 2009](#)).

Nine Box ratings are often used for succession planning and the allocation of training, development, and promotion opportunities, and can also be tied to compensation.⁵ In principle, Nine Box allows organizations to distinguish star individual contributors from the best candidates for promotion, a distinction that may be particularly relevant in technical fields like science, engineering,

⁵For instance, Microsoft has also used performance ratings to distribute cash bonuses and potential ratings to allocate promotions ([Bartlett, 2001](#)).

law, and academia where the skills required to perform and manage a job are quite different ([Baker, Jensen and Murphy, 1988](#)). Otherwise, promoting on prior performance alone can yield substantial mismatch between a worker’s skills and their role ([Benson, Li and Shue, 2019](#)).

Critics, however, argue that Nine Box is less transparent, objective, and consistent than the formal psychometric and skills evaluations that they replaced. In their review of talent management practices, [Cappelli and Keller \(2014, page 315\)](#) summarize:

“The conceptual idea behind assessing potential has been to identify abilities, given knowledge and skills that presumably can be learned through the development process....however, employers appear to have fallen back on the basic approach of simply asking supervisors to make an assessment of potential, an approach built in to performance appraisals through the nine-box grid, again made famous by GE. It is a matrix in which performance is assessed on one axis and potential on the other. However, the lack of a definition for what constitutes potential, both within firms and within the academic literature ([Groysberg and Nohria, 2011](#); [Silzer and Church, 2009](#)), gives us little reason to believe that this process should produce valid information, despite its widespread use.”

Interviews conducted by [Yarnall and Lucy \(2015\)](#) found that even raters themselves believe Nine Box potential ratings to be highly subjective. In particular, because supervisors are often provided with limited guidance, weak criteria, and little or no concrete evidence, Nine Box ratings may be prone to well-documented rater biases. Raters, for instance, may refer to prior years’ potential ratings (anchoring bias), first impressions (primacy bias), last impressions (recency bias), and ratings of other dimensions (halo bias) (for a review, see [Kahneman, 2011](#)).

Research points to several channels through which biases in subjective evaluations may work against women in particular. First, the psychology literature on role congruity theory, first proposed by [Eagly and Karau \(2002\)](#), has argued that people may have a hard time imagining women as qualified for leadership positions because of a mismatch between traditional female stereotypes and traditional leadership stereotypes. This tendency may be exacerbated by the fact that women are less represented in positions of leadership. Second, subjective evaluations may be influenced by politicking and favoritism ([Prendergast and Topel, 1993](#)). This could generate gender based disparities if women have less access to networking opportunities ([Cullen and Perez-Truglia, 2019](#)), benefit less from connections ([Fang and Huang, 2017](#)), or engage in less self promotion.⁶ Finally, self-interested managers may manipulate potential scores to keep their best subordinates ([Friebel](#)

⁶[Azmat, Cuñat and Henry \(2020\)](#) find that female lawyers report lower partnership aspirations. In our setting, low aspirations may translate into women engaging in less self promotion, which could, in turn, negatively impact their potential ratings. We note that possible gender differences in aspirations are compatible with a view in which promotions are biased against women. [Azmat, Cuñat and Henry \(2020\)](#) model aspirations as endogenously determined

and Raith, 2013). Haegele (2021) find that such “talent hoarding” leads to disproportionately lower promotion rates for women, possibly because female subordinates have a stronger distaste for confrontation with their managers.

Despite these concerns, Nine Box remains a highly popular method of identifying candidates for developmental opportunities and promotion, both because it’s easy to implement on its own and integrated into the leading HR software packages. As an article in HR Magazine points out: “What’s not to like about the Nine Box grid? It’s free, easy to use, and ubiquitous.”⁷

At our firm, Nine Box ratings are assigned annually in a two-step process. First, managers provide initial ratings for their salaried subordinates. Second, there are a series of district- and headquarter-level calibration meetings during which scores may be adjusted to ensure that similar standards are being applied. Although we are not aware of any systematic surveys that examine exactly how organizations use Nine Box, its implementation in HR software and our conversations with practitioners suggest our firm is typical in its two dimensional rating, the labels it ascribes to each of the nine boxes, the reliance on immediate supervisors rather than objective data or psychological tests for initial ratings, and its use of calibration meetings to mitigate bias.

2.3 Data and summary statistics

We obtain data on Nine Box ratings, promotions, and various demographic characteristics for 29,809 management-track workers employed between 2011 and 2015. These represent the near universe of full-time, salaried, management-track employees at our sample firm during this period. Our data includes workers employed in the firm’s corporate headquarters, as well as workers employed across 4,101 retail locations.

Our main data is at the worker-month level. We code a worker as being promoted if we observe an upward change in job levels in the next month. Examples of promotions in our data include moving from department manager to store manager, or moving from web developer to lead web developer.

Nine Box assessments are finalized and recorded in the fourth quarter of each fiscal year, which ends in January. The exact month in which each worker is assessed a new Nine Box rating varies from year to year. We set a worker’s Nine Box rating in each year-month equal to the new updated

by workplace gender discrimination (see also Brands and Fernandez-Mateo (2017)). Further, we find in our retail setting that women receive higher performance ratings than men, suggesting that any gender differences in aspirations do not translate into lower effort provision by women.

⁷We are not aware of any systematic studies of Nine Box’s adoption. However, Nine Box is integrated in the major human capital management software packages including Workday, SAP, PeopleSoft, Cezanne, Trakstar, Pipefy, emPerform. These services facilitate Nine Box reporting and its translation into development and succession planning. Nine Box Excel templates are also freely available online. Practitioners we spoke to at Accenture, CitiGroup, Bristol Myers Squibb, Honeywell, 3M, Ecolab, and General Mills confirmed it is widely known and used, including at their organizations.

rating if the rating is updated in the current month, and equal to her next updated rating if the rating is not updated in the current month.

Figure 1 shows the labels used by our data provider to describe each box within the Nine Box system. Our data provider reserves the upper left box, representing low performance and high potential for new hires. Because this rating is mechanically assigned based on tenure, we drop these observations from our analysis sample.

In addition to Nine Box ratings, we observe the following individual-level information: gender, race, ethnicity, tenure in the firm, compensation, job role, subordinates, and manager. For those employed in retail operations, we also observe identifiers for store location and the store’s overall financial performance in that year.

We determine promotions using data on standardized job titles and annual salary. Most job titles are clearly hierarchical, e.g., a typical career ladder in retail operations can be ordered as assistant department manager, department manager, assistant store manager, store manager, assistant district manager, district manager, vice president, senior vice president, etc. In other cases, the ranking is less clear (e.g., coordinator versus supervisor). We rank job titles by average compensation and classify a worker as having received a promotion if she experiences a change in job title that is associated with an increase in average compensation associated with that job title or experiences a change in job title that is associated with a personal raise in salary exceeding 5%.

Table 1 Panel A provides an overview of our sample coverage in terms of workers, time period, and promotion events. Panel B provides summary statistics associated with our sample and key variables. 41% of employees in our sample are female and the average annualized promotion rate is 11.9% (equal to the monthly promotion rate \times 12). Panel C provides pairwise correlations between some of our key variables. As a simple preview of our more detailed empirical results, it is evident from this panel that being female is positively correlated with performance ratings and negatively correlated with promotion, annual salary, and potential ratings.

3 Potential and the Gender Promotion Gap

In this section, we present several results aimed at documenting how potential ratings can help explain the gender promotion gap. We begin by describing promotion rates, both in the raw data and controlling for various worker characteristics. We then document the gender gap in potential and show that it can explain a substantial portion of the overall gender promotion gap.

3.1 Gender and Promotion

In our firm, as in many others, the share of women progressively decreases as one ascends the career ladder, as illustrated in Figure 2. In the left panel, we focus on workers in retail operations, for which there exists a clear ordering of job titles. In stores, women make up 56% of entry level workers (such as cashiers, merchandisers, backroom associates, and salespeople), 48% of department managers, 35% of store managers, and only 14% of district managers. In the right panel, we examine female representation by pay decile (sorted within fiscal year) for all workers with Nine Box ratings within the whole organization. We see a similar pattern of decreasing female representation as one advances in pay deciles. 49% of workers in the bottom pay decile who receive Nine Box ratings are women, compared with 29% at the top.

Declining female representation toward the top of the organizational hierarchy is suggestive of a gender gap in promotions to higher level job roles. We explore whether women are less likely to be promoted using the following regression:

$$\text{Promotion}_{it} = a_1 \text{Female}_i + a_2 X_{it} + \delta_y + \varepsilon_{it}. \quad (1)$$

In Equation (1), the level of observation is at the worker-year-month level, where i indexes individuals and t index time measured in months. The sample consists of all full-time workers with Nine Box ratings (these workers are considered management track and exclude entry level workers such as cashiers). The main outcome of interest is Promotion_{it} , an indicator for whether a worker is promoted in the next month, but we also consider other outcomes such as compensation. Monthly promotion rates are low, so we convert it to an annualized percent by multiplying it by 1,200 (12 months \times 100 percent). The key independent variable is an indicator for whether the worker is female. In all specifications, we control for year fixed effects δ_y to account for time trends. In some specifications, we also controls for a worker’s Nine Box performance and/or potential rating, log age, log tenure, race fixed effects, and location fixed effects. Without these control variables, the coefficient on Female_i measures the overall, unconditional gender gap. With these control variables, the coefficient on Female_i measures the unexplained gender gap after accounting for gender differences in control variables X_{it} . Standard errors are clustered by worker to account for account for correlated errors within worker over time.

Table 2 documents a substantial and robust gender gap in promotion rates. Column 1 presents the overall gender gap in our data. The coefficient, -1.6, on the female indicator implies that women are 1.6 percentage points less likely to be promoted, or 13.5% less likely to be promoted related to the base promotion rate 11.9%. Because this difference in promotion could be due to differences in performance, we control for the worker’s Nine Box performance ratings fixed effects in Column 2

(the omitted category is a performance rating of 1). We find that higher performance ratings are strongly predictive of promotion. More importantly, controlling for worker performance actually increases the gender gap in promotions. As we shall see in future analysis, this occurs because female workers receive higher performance ratings. Once we condition on workers who receive the same performance ratings, we observe a larger female disadvantage in promotions.

In Column 3, we show that part of the gender gap in promotions can be explained by differences in correlated demographic variables. As shown in Table 1 Panel B, women tend to be older and have longer tenure within the firm, and to be Black or Hispanic; these demographic variables are also associated with lower promotion rates. However, even after controlling for these demographic characteristics, women are 1.08 percentage points less likely to be promoted each year (or 9.03% less likely to be promoted relative to the base rate).⁸

Women may also be assigned to different store and administrative locations than men. In Column 4, we control for location fixed effects, and find the gender promotion gap remains approximately constant in size if we compare workers in the same location. To offer a comprehensive view of the data, we present both the “overall gender promotion gap,” which refers to the raw gap in Column 1 and the “gender promotion gap, controlling for performance, demographics, and location,” which refers to the estimates in Column 4, after including performance, demographic, and store location controls. We will continue to estimate specifications that include these control variables throughout our analysis.

Table 3 documents how differences in promotion rates may lead to differences in compensation. Column 1 shows the overall gender wage gap in our data: the coefficient of -0.118 implies that women’s salaries are 12.5% lower than men’s. This gap shrinks dramatically to just 3.7% in Column 2, after we control for job level by year fixed effects. Thus, hierarchical differences in assigned job roles account for 70% of the gender wage gap. In Columns 3 and 4, we introduce additional controls for performance and potential ratings, as well as demographic variables and location fixed effects. While these variable do significantly predict compensation, they do not lead to large changes in our estimate of the gender wage gap. Instead, gender differences in job levels, which are determined by promotions, appear to be the main determinant of the gender wage gap.

3.2 Gender and Potential

We now consider why women have lower promotion rates. Table 4 documents how Nine Box performance and potential ratings differ for men and women. Panel A measures ratings on a 1, 2, and 3 scale while Panel B looks at differences in the probability of earning the top rating of 3.

⁸We do not observe significant interaction effects between gender and other demographic characteristics (the “double jeopardy” hypothesis) within our sample.

We find that women receive substantially higher performance ratings, both in the raw data and conditional on demographics and location. In particular, Column 1 of Panel B shows that women are 1.81 percentage points more likely to earn the top performance rating, an increase of 7% relative to the base probability of earning the top performance rating.

In contrast, women earn substantially lower potential ratings, both in the raw data and conditional on demographics and location. Column 3 of Panel B shows that women are 1.43 percentage points less likely to earn the top potential rating. This gap is quite large compared to the base probability of earning the top potential rating (4.46%)—women are 32% less likely to earn the top potential rating. The divergence in potential and performance ratings for women is all the more surprising because that the two ratings are positively correlated in the overall sample, as shown in Table 1 Panel B.⁹ This divergence suggests that potential scores may be biased against women, a question we evaluate in more detail in Section 4.

Figure 3 plots additional details about the gender difference in performance and potential scores. The left panel plots the distribution of performance and potential scores for men in our sample, while the right panel represents the differences in shares for women relative to men. Women are significantly less likely to earn low performance ratings and significantly more likely to earn high performance ratings. The opposite pattern occurs for potential ratings. Women are significantly more likely to earn the lowest potential rating and significantly less likely to earn the highest potential rating.¹⁰

In Table 5, we examine the extent to which the gender gap in potential ratings explains the gender gap in promotion. We replicate each column of Table 2, adding controls for the worker’s potential rating. By comparing the coefficient on the female indicator in each column of Table 5 with the corresponding coefficient in Table 2, we can estimate the fraction of the gender gap in promotion rates that is explained by gender differences in potential ratings.

We find that the coefficient on the female indicator shrinks substantially once we control for potential ratings. 53% of the overall gender gap in promotions can be explained by potential ratings. Potential ratings can also explain 48% of the promotion gap conditional on performance ratings, 46% of the promotion gap conditional on performance ratings and demographic characteristics, and 33% of the promotion gap conditional on the above variables and location assignment.

The high explanatory power of potential ratings for the gender promotion gap can be attributed to two forces. First, as seen previously, women are assigned lower potential ratings both unconditionally, and conditional on performance scores, demographics, and location assignment. Second, Table 5

⁹Note that a positive correlation of 0.088 between potential and performance ratings is considered substantial given that these are ordinal variables taking on integer values between 1 and 3.

¹⁰See Appendix Figure A1 for the raw frequency of observations by gender and the promotion rate within each of the nine boxes.

shows that potential ratings are strong predictors of promotion. In all specifications, we find that a one point increase in potential ratings corresponds to a greater jump in the probability of promotion than a comparable one point increase in the performance ratings. For example, Column 2 shows that a change in potential ratings from 2 to 3 corresponds to a 8.98 percentage point increase in the promotion rate, while a similar change in performance ratings from 2 to 3 corresponds to only a 3.24 percentage point increase in the promotion rate. The remaining unexplained gender promotion gap, as measured by the coefficient on the female indicator, in Table 5, may capture several omitted factors. First, women may be less likely to seek or accept advancement opportunities (Fernandez and Mors, 2008; Fernandez-Mateo and Fernandez, 2016). These gender differences in career aspirations may arise endogenously from other forms of gender bias, and should not necessarily be considered a distinct force. Recent studies have consistently found that stated aspirations are endogenous to perceived opportunities (see, e.g. Correll, 2004). Similarly, Azmat, Cuñat and Henry (2020) find that female lawyers who faced harassment and discrimination report lowered aspirations. Differences in training, particularly related to career development, could also be an omitted factor, though this too is likely endogenous. Nine Box potential ratings at our firm (and by convention) are used to allocate scarce internal developmental opportunities, and the prospect of being high-performing but “invisible” can reduce incentives for women and other minorities to invest in development themselves Milgrom and Oster (1987).

In addition to our main analysis of the relation between potential ratings and promotions in Table 5, we also provide a supplementary Blinder-Oaxaca three-fold decomposition in Appendix Table A1. The decomposition reports the portion of the overall gender gap in promotion rates that can be attributed to differences in endowments (differences in the level of potential ratings and other variables), differences in coefficients (differences in the return to potential ratings and other variables), and interactions between endowments and coefficients. The decomposition reveals an overall gender gap in promotion rates of 1.635 percentage points, of which 0.9 (or 55%) can be explained by gender differences in the endowments of potential ratings and -0.159 (or -9.7%) can be explained by differences in the endowments of performance ratings (this figure is negative because women earn higher performance ratings on average). We also see suggestive evidence that women have a lower return to potential scores when it comes to being promoted. However, these coefficient differences are estimated with greater noise.¹¹

¹¹We do not directly report the coefficients because the decomposition results for categorical predictors such as the potential ratings dummies is sensitive to the choice of the omitted base category. We follow Yun (2005) to estimate a decomposition using normalized effects.

4 Informativeness of Potential Assessments

So far, we have shown that low assessments of potential help explain why women are less likely to be promoted. In this section, we examine the information content of potential assessments and show that, despite containing useful information about a worker’s future performance, potential scores appear to be biased against women.¹²

4.1 Realized potential: Gender gaps in future performance ratings

We begin by assessing the predictive value of potential ratings. While the exact definition of “potential” is often debated even within organizations, most practitioners agree that potential ratings should forecast an individual’s ability to contribute to the firm in the future, either through improved performance and greater responsibilities in her original job role or through leadership in a new managerial role (Cappelli and Keller, 2014; Groysberg and Nohria, 2011; Silzer and Church, 2009; Yarnall and Lucy, 2015). Thus, effective potential ratings should predict actual future performance, particularly among the sample of workers who are promoted into management positions. We therefore think of a worker’s future performance ratings as a measure of his or her “realized potential.”

Gender differences in average future performance ratings

Table 6 Panel A shows that high current potential ratings predict higher measures of realized potential 12 months into the future. Our estimates in Column 1 indicate that, relative to those with a low potential rating (the median for the sample), workers with a high potential rating have a 0.17 point higher performance rating in the following fiscal year. Importantly, this positive and highly significant correlation holds even after conditioning on the worker’s current performance score. That is, potential ratings appear to contain real information about a worker’s future performance, beyond what can be forecast using information on prior performance alone. This correlation holds in both the full sample of workers, as well as within the subsample of employees who experience a promotion event (so that their future performance ratings reflect performance in a new role).

Since potential scores are predictive of future performance, one natural explanation for why women may receive lower potential scores is that they are likely to have worse future performance.¹³ If women’s lower potential scores are indeed justified, then we would expect that, controlling for

¹²This is consistent with Li (2017), which shows that ignoring advice from biased advisers would reduce the overall quality of investment decisions, because biased advice still contains useful signals of a project’s quality.

¹³Azmat and Ferrer (2017) find that differences in billable hours and new business origination explain about half of the gender gap in lawyers’ pay. Relatedly, Cook et al. (2018) find women Uber drivers earn less per hour than men despite identical pay contracts due to differences in experience and driving preferences. However, Sarsons (2017a) finds female surgeons are slightly higher ability than male surgeons.

current period potential ratings, men and women should have similar future performance ratings. Following the logic of a Becker outcomes test for discrimination, assessments of potential are biased against women if, for the same potential score, women have higher measures of future realized potential.

We examine this relationship in several ways. We begin by showing that women have higher realized potential on average, relative to male colleagues with the same current potential score, in both the full sample of workers and the subsample of workers who are promoted into new roles. Next, for the sample of promoted workers, we also show that this relationship holds at the margin. That is, we use an instrumental variables approach to show that a marginally promoted woman performs better in her new role, relative to a marginally promoted man with the same initial potential rating. This difference at the margin suggests that firms are misallocating promotion opportunities: for the same potential rating, women are held to a higher threshold of future performance in the promotions process.

Table 6 Panel A examines this possibility by relating current period potential scores with the next performance rating (measure 12 months in the future). Columns 1 and 2 focus on the full sample of workers, where “next period” performance can refer to either performance in the same role or in a different role. Column 1 controls for year fixed effects while Column 2 also controls for location fixed effects and demographics. In both cases, we find that, controlling for a worker’s current potential and performance scores, women receive higher future performance ratings than their male colleagues. That is, women systematically outperform forecasts of their potential. In Columns 3 and 4, we limit the sample to workers promoted in the current year-month and again regress future performance ratings on a female indicator and pre-promotion ratings. Since the sample of promoted workers is much smaller, and some locations only have one promotion event within our sample period, we exclude location fixed effects in this and all other analysis restricted to the promoted subsample. We again find a similarly-sized significant positive coefficient on the female indicator, implying that promoted women outperform promoted men, conditional on current potential and performance ratings and other observable control variables.

In supplementary analysis, we explore how the gender gap in potential ratings varies with county-level measures of labor market gender inequality. We view this analysis as providing additional evidence consistent with the view that potential ratings are negatively biased against women. Managers working in counties with low female representation in management-level positions, large gender wage gaps, and low female educational attainment, may face extra difficulties in imagining that women are qualified for higher level positions because they do not frequently observe women in these positions. Under the predictions of role congruity theory, managers in these counties may be especially likely to assign low potential ratings to female subordinates. In Appendix Table A2, we

show that the gender gap in potential ratings is indeed significantly larger in counties with lower female representation in management-level positions, larger gender wage gaps, and lower female educational attainment. Full details for the construction of the county-level measures of labor market gender inequality are provided in the Appendix.

Are firms promoting too few women? IV evidence

The findings in Table 6 are suggestive of bias against women: on average, promoted women outperform promoted men with the same pre-promotion potential ratings. This raises the question of whether firms can increase managerial quality by promoting more women on the margin. To explore this, we use an IV strategy (described shortly) to measure the realized future performance of women who are promoted on the margin, and compare it to that of marginally promoted men with the same potential scores. If marginally promoted women have higher realized potential then, when faced with workers with the same potential scores, firms can increase managerial quality by promoting more women. This analysis is equivalent to a Becker outcomes test for discrimination: we ask whether firms set different thresholds for promotion by gender.

Following [Benson, Li and Shue \(2019\)](#), which builds on earlier work by [Abadie \(2003\)](#) and [Arnold, Dobbie and Yang \(2018\)](#), we identify “marginally promoted” applicants using an instrument for promotions. The intuition is that, given a valid promotion instrument, the set of instrument compliers are, by definition, marginal: they are promoted if they receive a good draw of the instrument, but not otherwise. If, among workers with the same current period Nine Box ratings, female compliers have higher future performance than their male counterparts, then the firm has set a higher promotion threshold for women and would benefit (in terms of higher expected managerial performance) by promoting more women on the margin.

Similar to the approach in [Benson, Li and Shue \(2019\)](#), our promotion instrument captures the idea that workers employed during employment expansions are more likely to be promoted irrespective of their performance or potential. Specifically, we instrument worker i 's promotion outcome in year month t using Z_{it} , the average promotion rate for workers with the same job title in the same year t , leaving out all workers in worker i 's same office or store location.

A natural concern with this instrument is that employment expansion may be correlated with future Nine Box ratings: for example, instrument compliers promoted in expansions may face more favorable circumstances and may be credited with higher performance as a result. We address this concern in several ways. First, in our analysis, we measure a worker's future performance rating residualized for job title and year fixed effects. That is, we consider a worker's future performance relative to other workers with the same job in the same year: by construction, this measure of realized future performance is not related to job-time level changes, such as changing consumer

demand, that may play a role in shaping our aggregate promotion rate instrument. Another potential concern is reverse causality: if a given worker is particularly strong, the firm may chose to promote her, generating a higher promotion rate for that worker’s job title at that time. Using a jackknife approach and leaving out a worker’s own promotion status (and that of her colleagues) severs the correlation between our instrument and an individual worker’s quality. Finally, note that we are ultimately interested in the *difference* in future performance ratings for marginally promoted men and women. As such, we are concerned about biases in our instrument that differ for men and women; our analysis is unbiased as long as male and female IV compliers are promoted into similar economic conditions.

To compute the future ratings of compliers, we estimate the following regressions:

$$Y_{it} \times P_{it} = \alpha_0 + \alpha_1 P_{it} + X'_{it} \alpha + \varepsilon_{it} \text{ if male} \quad (2)$$

$$Y_{it} \times P_{it} = \beta_0 + \beta_1 P_{it} + X'_{it} \beta + \nu_{it} \text{ if female} \quad (3)$$

In Equations (2) and (3), $Y_{it} \times P_{it}$ is equal to applicant i ’s future ratings outcome Y_{it} if she is promoted ($P_{it} = 1$) and to zero otherwise. We include controls for worker’s current period Nine Box potential and performance ratings. The OLS coefficients $\hat{\alpha}_1^{OLS}$ and $\hat{\beta}_1^{OLS}$ estimates average future performance ratings among all promoted men and women, after controlling for other covariates, respectively. The IV estimates $\hat{\alpha}_1^{IV}$ and $\hat{\beta}_1^{IV}$, in contrast, represent future ratings among male and female compliers, respectively. This logic is analogous to the idea that IV estimates identify a LATE amongst compliers.¹⁴ Our analysis focuses on the difference between male and female workers who are promoted on the margin: $\hat{\alpha}_1^{IV} - \hat{\beta}_1^{IV}$.

Panel A of Figure 4 presents the results of our analysis, with the accompanying regressions reported in Appendix Table A3. We see that marginally promoted women have higher future performance ratings, relative to marginally promoted men with the same current period Nine Box ratings. Appendix Table A3 also reports that these results hold when controlling for demographics and location. To provide additional context, Panels A and B of Appendix Figure A2 compare the future performance ratings of the average and marginally promoted male and female worker, respectively. For both men and women, the average promoted worker performs better in the future than the marginal promoted worker. This is expected, as the quality of “always-takers” (those the firm promotes regardless of the value of the instrument) should exceed the quality of “compliers” (marginally promoted workers). Together, these results suggest that firms are more likely to promote higher quality workers relative to lower quality workers within gender, but that women appear to be held to a higher threshold.

¹⁴For a formal proof, see Benson, Li and Shue (2019).

This finding suggests that gender bias in potential ratings leads to misallocation in managerial opportunities. To see this explicitly, consider the following modification of the firm’s existing promotion policy P :

$$P(\tilde{X}) = \begin{cases} P(X)^{Z=1} & \text{if female,} \\ P(X)^{Z=0} & \text{if male.} \end{cases}$$

This new promotion policy modifies the firm’s existing practices by favoring women on the margin. Specifically, consider two workers, male and female, with the same Nine Box ratings X . The new policy \tilde{P} asks that women be evaluated for promotion as if the firm had many promotion opportunities available ($Z = 1$), and men be evaluated as if the firm had few such opportunities ($Z = 0$).¹⁵ By construction, the promotion decisions of this new policy differ only in its treatment of instrument compliers. In particular, women would have been promoted only when overall promotion rates were high would be promoted under this new policy, but men in the same situation would not be. The difference in future managerial performance between \tilde{P} and P is therefore given by the difference in the future performance of female and male compliers. As demonstrated in Panel A of Figure 4, this difference is positive: the firm can improve the quality of its managers by favoring women when selecting among employees with similar Nine Box ratings.

4.2 Updating biased beliefs: Gender gaps in future potential ratings

We next consider how firms update evaluations of women’s potential in response to their realized future performance. To do this, we replicate the analysis in the previous section, using 12-month ahead *potential* ratings as the outcome of interest.

Panel B of Table 6 shows that women continue to receive significantly lower future potential ratings compared to men with identical current performance and potential ratings, both in the full sample and in the sample of newly promoted workers. Panel B of Figure 4 shows that this pattern also holds in the sample of marginally promoted workers, with regression analogues reported in Appendix Tables A4. Both on average and on the margin, women continue to receive lower next period assessments of their potential, even after they have outperformed their male colleagues with the same prior period potential rating.

In Appendix Table A5, we show that our findings are robust to controlling for manager fixed effects. That is, among subordinates of the same manager with the same current potential ratings, women outperform men in the future but continue to earn lower future potential ratings.

Finally, we emphasize that performance and potential together comprise the firm’s Nine Box evaluation. Because of this performance and potential ratings are determined during the same

¹⁵For simplicity in exposition, we let Z be a binary instrument in this example (whether job level promotion rates are above or below median) though in practice we will use a continuous variable.

meetings, by the same managers. This means that, at the same time that women are given performance ratings indicating that they outperformed their previous year’s potential scores (relative to men with the same potential scores), women are still assessed as having lower potential going forward.

4.3 Compounding gender gaps

This evidence raises the possibility that persistent gender gaps in performance and potential evaluations may compound over time, leading to wider gaps in representation at top levels, consistent with Figure 2. Such a pattern would challenge the efficacy of corporate diversity programs aimed at expanding female leadership.

To examine variation over the career ladder, Appendix Table A6 explores how the gender gap in performance, potential ratings, pay, and promotion rates changes as one advances in the organizational hierarchy. We regress workers’ performance and potential ratings, as well as salary and promotion outcomes, on the interaction of a worker’s gender and their career ladder position as proxied by the worker’s decile in the firm’s pay distribution in the corresponding fiscal year, controlling for the individual effects of these variables. In all specifications, we control for fiscal year fixed effects, demographics, and location fixed effects.

Our results indicate that, while women continue to outperform men throughout the hierarchy (and this gap does not change significantly), the gender gap in potential ratings widens at higher levels of the firm’s hierarchy. This widening gender gap in potential occurs alongside a widening gender gap in promotions, which holds even after controlling for performance ratings. We also find that women continue to be paid less than their male colleagues, although this pay gap stays constant with respect to pay decile.

We note that a simple compounding effect implies that, given a constant gender gap in promotions, female representation will decline with each increasing rung in the career ladder. The fact that the gender gap in potential ratings and promotions grows with pay decile implies that the gap in female representation not only grows, but actually accelerates. Taken together, these results show that women face growing disadvantages as they advance up the career ladder.

5 Leaves of absence, retention, and risk of loss

One possible explanation for our results is that potential ratings reflect a manager’s expectation of a worker’s commitment to the firm. If managers expect that women’s careers are more likely to

be interrupted or cut short by family care duties, they make lower their assessments of women’s potential to contribute to the firm in the future.¹⁶

Our previous results in Section 4 focus on the relation between current potential ratings and future contributions to the firm as proxied by future performance ratings. This test does not account for the possibility that workers’ future contributions could differ as a result of attrition or leave of absences. In this section, we explore how the gender potential gap we document relates to these additional factors.

In Table 7 we examine attrition from the firm entirely. Column 1 demonstrates that, on average, women have lower attrition than men. Therefore, managers should not give women lower potential ratings due to concerns that women are more likely to exit the firm; the opposite is true in the data.

In Column 2, we consider whether workers are more likely to exit when they are “passed over” for a promotion, which we code as having occurred if another worker reporting to the same manager is promoted (moves to a higher position in the next month). We find that men who are passed over are 35-40% more likely to exit the firm, relative to the baseline rate; among women who are passed over (with the same Nine Box ratings), this figure is only 13-20%. In Columns 3-4, we repeat this exercise, but restrict the sample to workers who received the highest Nine Box performance rating score. Among this group, passed over workers have a much higher likelihood of attrition relative to baseline and this elevated rate is driven almost entirely by men. High performance males who are passed over are 40-50% more likely to leave, whereas women in the same position are at most only 10% more likely to leave.

The fact that men are at higher risk of attrition may impact how they are treated by the firm. In Table 8, we consider the relation between gender, perceptions of attrition risk, and potential scores. For three years of our data, we observe firm ratings for each employee’s “risk of loss,” a three point rating capturing a worker’s risk of leaving the firm. In Column 1, we show that risk of loss ratings are indeed predictive of future attrition: workers rated as being at high risk of loss are over 60% more likely to exit the firm, relative to those at low risk. In Column 2, we see that women receive substantially lower risk of loss ratings, relative to men with the same Nine Box performance and potential ratings. Finally, Columns 4 and 5 show how perceptions of attrition risk may help explain why men achieve better outcomes along a range of dimensions. In Column 3, we see that risk of loss ratings are positively and significantly related to a worker’s next potential rating (measured 12 months in the future), controlling for current performance and potential ratings. In Columns 4 and 5, we find that higher risk of loss ratings are also associated with significantly higher promotion

¹⁶We leave aside the important question of whether it is legal or ethical to consider future attrition or leave, particularly maternity leave, when forming Nine Box ratings or promotions decisions. In this paper, we focus on the empirical question of whether women in our sample are more likely to exit the firm or to take leave, and whether that information is correlated with potential ratings.

probability and compensation. Finally, we note that the coefficient on female remains large and negative throughout these regressions, implying that the gender gap cannot be completely explained by women’s lower risk of loss and potential ratings.

Taken together, our results suggest that firms anticipate men’s higher rates of attrition in their risk of loss assessments, and attempt to retain them by granting higher next period potential scores, promotions, and pay. This pattern suggests that firms essentially reward the threat of exit, rather than perceiving it as a negative signal of a worker’s commitment or ability to contribute to the firm in the future. Yet, as can be seen in Tables 6 and 7, this leads the firm to be more likely to promote men who, relative to their female peers with the same Nine Box ratings, tend to have lower future performance and higher rates of future attrition.

In Table 9, we consider differences in leaves of absence, defined as temporary time off of work that could be paid or unpaid. The most common reasons for leaves are related to family and child care, or personal or family related medical issues. Column 1 shows that women, indeed, are substantially more likely to take leaves of absence from the firm. The coefficient on the female indicator implies that women are 0.45 percentage points more likely to be on leave in the following month, 65% higher than the baseline of 0.70 percentage points. In Columns 2-4, we explore how this difference relates to women’s potential ratings. Column 2 reports the raw gender potential gap restricted to the slightly reduced sample for which we observe leaves of absence data. Column 3 shows that the gender gap in potential ratings remains similar in magnitude after controlling for the worker’s past leaves measured in number of months. Column 4 shows that the gender gap in potential ratings remains similar even after controlling for realizations of leaves in the future. In both cases, past and future leaves are negatively related to potential scores, but the gender gap in potential scores appears to exist separately from inferences about leave.

Of course, managers may assign female subordinates lower potential ratings because of expectations about future leaves (we would not be able to control for these expectations using only data on actual future leaves, as in Column 4). We can instead conduct the following thought experiment: how much extra future leave must managers believe women will take to explain the gender gap in potential scores? Column 3 implies that, based on the relationship between past leaves and potential ratings, managers would have to believe that women take on average four extra months of leave to justify a gender gap of 0.086 points in potential. These beliefs do not match the data: compared to men, women take an extra .05 months of leave per year relative to men. Even if the manager considers potential leaves over the next 10 years, women on average only take an additional half of month of leave relative to men. In other words, women take relatively more leave than men, but the absolute levels of leave among women are too low to explain the large gender gap in potential ratings.

In all analysis in this subsection, we have focused on raw gender differences, including those due to correlated demographics and differential location assignment. In Appendix Tables [A7-A9](#), we show the results are qualitatively similar after controlling for demographics and location assignment.

6 Potential policy responses

In this section, we provide suggestive evidence related to the potential efficacy of two types of HR policy responses: changes to the managers making Nine Box ratings decisions, and changes to Nine Box ratings practices themselves.

6.1 Heterogeneity by manager assignment

We begin by documenting how gender gaps in ratings, pay, and promotions vary across different types of managers. In particular, we focus on two manager characteristics: gender and the manager’s own performance and potential ratings. This analysis is motivated by the common suggestion that women would benefit from working under female managers, who may be less biased against other women and act as mentors and advocates for their female subordinates. [McGinn and Milkman \(2013\)](#), for instance, show that female managers serve as role models for their female subordinates, and enhance their career progression.¹⁷ Likewise, women may benefit from working under higher quality managers who may be better at assessing their subordinates’ true performance and potential, less concerned about competition from their own subordinates, and less likely to hoard their talented subordinates.

Throughout this analysis, we regress a worker’s performance and potential rating, pay, or promotion outcomes on gender, the manager characteristic of interest (gender or performance and potential rating), and the interaction between worker gender and manager characteristics. We control for year, store location, other worker demographics and, in some cases, worker Nine Box ratings. We caution, however, that we cannot distinguish between treatment or selection effects. That is, subordinate outcomes may differ across managers both because managers are assigned to different types of subordinates and because managers differ in how they assess or advocate for their subordinates. As a result, these results should be viewed as suggestive.

In [Table 10](#), we examine whether a subordinate’s rating depends on their gender and the gender of the manager who is rating them, and is motivated by studies that have found such interaction effects on termination and career advancement (e.g., [Egan, Matvos and Seru, 2017](#); [Cullen and Perez-Truglia, 2020](#)). In [Column 1](#), we find that female workers earn higher performance ratings

¹⁷Research on the “queen bee” syndrome shows that female managers can sometimes be tougher on their female subordinates, possibly due to a competition effect (see e.g., [Ellemers et al. \(2004\)](#)). Thus, it is not obvious that female subordinates would be better off working under a female manager.

than their male colleagues, and this outperformance does not vary with the manager’s gender. Other columns show that gender gaps in potential ratings, pay, and promotions are smaller (but not eliminated) under female managers. However, female managers are also associated with lower overall levels of ratings, pay, and promotion rates for all their subordinates, regardless of subordinate gender. This is evidenced by the negative coefficients on the manager female gender indicator. Taken together, the opposing level and interaction effects imply that it is not obvious that female employees would be better off working under a female manager. A female employee can expect a smaller gender *gap*, but not necessarily an increase in the absolute levels of ratings, pay, or promotion rates. Our findings of offsetting effects echoes related results in [Cardoso and Winter-Ebmer \(2010\)](#), who show that female-led firms are associated with lower gender wage gaps as well as lower levels of wages.

In [Table 11](#), we explore how worker outcomes vary with their manager’s performance and potential evaluations. We find that subordinates assigned to managers with higher performance and potential ratings are significantly more likely to receive higher performance and potential ratings, have higher salaries, and to be promoted. To the extent that these differences represent treatment effects, female workers would benefit from being assigned to more highly-rated managers. We find, however, that the interaction effects between worker gender and manager ratings are negative in most cases. This implies that although the level of worker outcomes is higher, gender gaps among subordinates assigned to more highly-rated managers are also larger. To the extent that highly-rated managers may see themselves as more “meritocratic,” this result aligns with [Castilla and Benard \(2010\)](#), which shows that ratings exhibit stronger gender gaps when merit is emphasized. On net, it is unclear whether women benefit from these assignments.

6.2 Counterfactual promotion policies

In this section, we consider the impact of counterfactual promotion policies on both equity, as measured by differences in promotion rates for men and women, and on efficiency, as measured by the future performance of promoted candidates.

We consider the following counterfactual policies. First, given that we have documented evidence of gender bias in potential scores, one potential remedy is to simply stop using them in promotion decisions. The first counterfactual promotion policy we consider is to remove information on gender and potential from promotion decisions. Another possibility is to continue using potential scores, but to first “adjust” them to account for gender bias. We consider two very simple ways of accomplishing this task: the first is to add one point to the potential scores of female candidates, so that women who are rated “low” are now rated “medium,” and those rated “medium” are now rated “high” potential. Since the maximum possible potential score is “high,” we leave the potential scores of

women who receive “high” potential scores unchanged. The second approach is a milder version of this correction, which only adds one to the potential ratings of women who are rated “high” in terms of performance.

To assess the impact of these promotion policies, we begin by estimating a regression of promotion on the female indicator, potential ratings dummies, performance ratings dummies, demographics, and year fixed effects. The coefficients from this regression represent the firm’s actual promotion policy given a set of ratings and other worker characteristics. We then predict a worker’s likelihood of promotion using the fitted value from this regression given the new inputs used by the proposed counterfactual policy.

That is, to evaluate the impact of blinding promotion to potential and gender information, we form estimates of promotion likelihood by setting the coefficients on gender and potential to zero. To evaluate the impact of policies in which we adjust potential ratings, we simply use our counterfactual adjusted potential scores as inputs into predicting promotion likelihood.

Once we have fitted measures of a worker’s likelihood of promotion under each policy, we assess the impact of each counterfactual policy on the expected promotion rates of men and women, as well as on estimates of the average future performance of promoted candidates. To form the average future performance of promoted candidates under each counterfactual policy, we use the weighted average of workers’ next period performance ratings, where the weights are the worker’s fitted likelihood of promotion under each counterfactual policy. That is, if a worker is more likely to be promoted under a given policy, we place more weight on this worker’s expected future performance. We can compute expected next performance ratings using the weighted average over either the full sample or the sample of true promotions. The full sample offers more complete coverage, while restricting the sample to true promotion events may offer a better measure of how future performance changes after a promotion event.

We report the results of this exercise in Table 12. Columns 1 and 2 compare promotion rates for men and women under each policy, while Columns 3 and 4 present the average future performance rating among promoted workers under each promotion policy. Focusing on the first two rows, we show that blinding promotion decisions to gender and potential ratings does indeed reduce the gender gap in promotions by 65%, from a 1.7 percentage point gap to a 0.6 percentage point gap. The gender gap does not disappear because we do not blind the firm to other demographics that are correlated with gender and associated with promotion rates. Yet, we also find that this reduction in the gender gap would come at the expense of reducing the expected average future performance of workers who are more likely to be promoted. Specifically, in Columns 3, 4, and 5, we report weighted averages of next period performance over the full sample of workers, the sample of workers who were not promoted, and the sample of workers who were promoted. respectively. In all cases,

we find that expected next performance ratings decline relative to the baseline promotion policy shown in the top row.

These results are consistent with Table 6, which shows that, despite being biased, potential ratings do contain useful information about future performance. A promotion policy that ignores potential scores altogether would be based on a coarser information set than one that is able to incorporate the information in potential scores in some way.

Next, we consider what happens when we retain information on potential scores, but apply adjustments aimed at “undoing” gender bias. We show that uniformly increasing the potential scores of all women leads to a substantial reversal of the gender promotion gap, so that now women are 40% more likely to be promoted than men. This approach, however, also leads to a small decrease in the expected quality of workers who are more likely to be promoted. Finally, in our third counterfactual, we show that a more targeted shift in potential scores, applying only to women who are rated highest in terms of performance, leads to an improvement in both equity and efficiency. In particular, this approach eliminates the gender promotion gap, while also increasing the estimated next period performance of promoted workers.

This analysis is subject to two limitations. First, while we observe next period performance for most workers regardless of promotion, next period performance holding the job role constant may differ from next period performance conditional on promotion. Therefore, we use the full sample in Column 3 and the true promoted sample in Column 4 to create weighted averages for the expected next performance rating. Our results in Column 4 are based only on the subset of workers who the firm saw fit to promote, and thus may offer a more realistic measure of how performance may change after promotion. However, if women are positively selected into the true promoted sample relative to men (e.g., due to discrimination against women in the promotion process), then a counterfactual promotion policy that increases the weights on women within the true promoted sample may overstate the efficiency gains of the counterfactual policy when applied to the entire sample.

Second, the “adjustment” policies that we evaluate may be circumvented if managers lower scores for female subordinates in anticipation of them receiving a gender-specific bonus. Given this, we regard our third counterfactual not a specific policy proposal, but as a demonstration that firms may be able to increase both the quality and equity of their promotion decisions by identifying ways to de-bias assessments of potential, rather than getting rid of potential assessments entirely.

7 Conclusion

In this paper, we provide evidence that subjective assessments of worker potential, widely used for career planning within firms, contribute to persistent gender gaps in promotion and pay. Despite being more likely to receive top performance ratings, women are less likely to be thought of as having high “potential.” These lower potential ratings can explain up 50% of the observed gender gap in promotions.

Women’s lower potential ratings may be justified if they accurately forecast worse performance in the future. We show that this is not the case. Rather, we find that potential assessments, while informative about future performance on net, are biased against women. That is, among employees with the same current performance and potential ratings, women outperform men on evaluations of their future performance. This is consistent with classic discrimination models in which, to receive the same potential rating, women are held to a higher bar in terms of their expected future performance. We show, further, that this bias in potential ratings does not appear to be self-correcting: even though women outperform their potential ratings, they continue to receive lower potential evaluations in the future. This persistence of lower potential ratings is true for women who continue in their current roles, as well as for women who are promoted and perform well in their new role.

Taken together, our results suggest that subjective assessments of potential are an ever present barrier to women’s advancement in their careers. In our data, the gender gap in promotions widens as one moves up the career ladder, as does the gender gap in potential scores. The failure of firms to update potential ratings to be in line with realized future performance suggests that inaccurate stereotypes and other types of biases may limit firms’ abilities to accurately forecast actual potential.

We also find that biased evaluations of potential are challenging to address. First, one cannot simply decrease the gender promotion gap by having more female managers. The presence of female managers attenuates the potential and promotions gap to some extent but, on net, female managers still give lower potential scores to women conditional on performance. This suggests that policies that seek to decrease the gap between assessed potential and future performance need to address broader organizational questions, rather than simply changing the gender of the evaluator.

Similarly, we show that assigning women to higher quality managers would not reduce gender bias. While managers who themselves receive higher performance and potential scores appear to be stronger advocates for their subordinates on net—they give them higher ratings and salaries, these benefits accrue almost entirely to male subordinates of high performing managers: gender gaps in performance, potential, and promotions expand under such managers.

Second, our results show that firms should not simply do away with potential ratings altogether. A growing literature now supports the long-held anecdotal belief that the best workers do not always make the best managers. When current performance is an imperfect indicator for future performance, it is reasonable for firms to look for other ways of assessing potential. In our data, assessments of potential are predictive of future performance beyond what can be predicted by current performance. This means that potential scores add value despite their gender bias.

Instead, our results show that there may be large gains from finding ways to de-bias assessments of potential, for instance by reducing reliance on stereotypes of who may be an effective leader. In recent years, firms have made various attempts to increase promotions and retention among women and minorities, from the use of bias-conscious algorithms in screening to training programs focused on conscious and unconscious bias. This paper suggests that these would be fruitful areas for further research.

References

- Abadie, Alberto.** 2003. “Semiparametric instrumental variable estimation of treatment response models.” *Journal of econometrics*, 113(2): 231–263.
- Abraham, Mabel.** 2020. “Gender-role Incongruity and Audience-based Gender Bias: An Examination of Networking among Entrepreneurs.” *Administrative Science Quarterly*, 65(1): 151–180.
- Arnold, David, Will Dobbie, and Crystal S Yang.** 2018. “Racial bias in bail decisions.” *The Quarterly Journal of Economics*, 133(4): 1885–1932.
- Azmat, Ghazala, and Rosa Ferrer.** 2017. “Gender gaps in performance: Evidence from young lawyers.” *Journal of Political Economy*, 125(5): 1306–1355.
- Azmat, Ghazala, Vicente Cuñat, and Emeric Henry.** 2020. “Gender promotion gaps: Career aspirations and workplace discrimination.” *CEPR Discussion Paper No. DP14311*.
- Babcock, Linda, and Sara Laschever.** 2009. *Women don't ask*. Princeton University Press.
- Baker, George P., Michael C. Jensen, and Kevin J. Murphy.** 1988. “Compensation and Incentives: Practice vs. Theory.” *The Journal of Finance*, 43(3): 593–616.
- Bartlett, Christopher.** 2001. “Microsoft: Competing on Talent (A).” Harvard Business School case study 9-300-001.
- Benson, Alan, Danielle Li, and Kelly Shue.** 2019. “Promotions and the Peter Principle*.” *The Quarterly Journal of Economics*, 134(4): 2085–2134.
- Biasi, Barbara, and Heather Sarsons.** 2020. “Flexible wages, bargaining, and the gender gap.” National Bureau of Economic Research.
- Blackaby, David, Alison L Booth, and Jeff Frank.** 2005. “Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market.” *The Economic Journal*, 115(501): F81–F107.
- Blau, Francine D, and Lawrence M Kahn.** 2017. “The gender wage gap: Extent, trends, and explanations.” *Journal of economic literature*, 55(3): 789–865.
- Brands, Raina A., and Isabel Fernandez-Mateo.** 2017. “Leaning Out: How Negative Recruitment Experiences Shape Women’s Decisions to Compete for Executive Roles.” *Administrative Science Quarterly*, 62(3): 405–442.

- Bureau, U.S. Census. 2019. <https://www.census.gov>, Accessed: 2021-05-26.
- Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. “Acting Wife’: Marriage Market Incentives and Labor Market Investments.” *American Economic Review*, 107(11): 3288–3319.
- Cappelli, Peter, and JR Keller. 2014. “Talent management: Conceptual approaches and practical challenges.” *Annu. Rev. Organ. Psychol. Organ. Behav.*, 1(1): 305–331.
- Cardoso, Ana Rute, and Rudolf Winter-Ebmer. 2010. “Female-led firms and gender wage policies.” *Industrial and Labor Relations Review*, 64(1): 143–163.
- Castilla, Emilio J., and Stephen Benard. 2010. “The Paradox of Meritocracy in Organizations.” *Administrative Science Quarterly*, 55(4): 543–676.
- Church, Allan H, Christopher T Rotolo, Nicole M Ginther, and Rebecca Levine. 2015. “How are top companies designing and managing their high-potential programs? A follow-up talent management benchmark study.” *Consulting Psychology Journal: Practice and Research*, 67(1): 17.
- Cook, Cody, Rebecca Diamond, Jonathan Hall, John A List, and Paul Oyer. 2018. “The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers.” National Bureau of Economic Research.
- Correll, Shelley J. 2004. “Constraints into preferences: Gender, status, and emerging career aspirations.” *American sociological review*, 69(1): 93–113.
- Correll, Shelley J, Katherine R Weisshaar, Alison T Wynn, and JoAnne Delfino Wehner. 2020. “Inside the black box of organizational life: The gendered language of performance assessment.” *American Sociological Review*, 85(6): 1022–1050.
- Cullen, Zoë B, and Ricardo Perez-Truglia. 2019. “The Old Boys’ Club: Schmoozing and the Gender Gap.” National Bureau of Economic Research.
- Cullen, Zoë B, and Ricardo Perez-Truglia. 2020. “The Old Boys’ Club: Schmoozing and the Gender Gap.” National Bureau of Economic Research.
- Cziraki, Peter, and Adriana Robertson. 2021. “Credentials Matter, but Only for Men: Evidence from the S&P 500.” *Available at SSRN 3894730*.
- Eagly, Alice H, and Steven J Karau. 2002. “Role congruity theory of prejudice toward female leaders.” *Psychological review*, 109(3): 573.

- Egan, Mark L, Gregor Matvos, and Amit Seru.** 2017. “When Harry fired Sally: The double standard in punishing misconduct.” National Bureau of Economic Research.
- Ellemers, Naomi, Henriette Van den Heuvel, Dick De Gilder, Anne Maass, and Alessandra Bonvini.** 2004. “The underrepresentation of women in science: Differential commitment or the queen bee syndrome?” *British Journal of Social Psychology*, 43(3): 315–338.
- England, Paula, Jonathan Bearak, Michelle J. Budig, and Melissa J. Hodges.** 2016. “Do Highly Paid, Highly Skilled Women Experience the Largest Motherhood Penalty?” *American Sociological Review*, 81(6): 1161–1189.
- Fang, Lily Hua, and Sterling Huang.** 2017. “Gender and connections among Wall Street analysts.” *The Review of Financial Studies*, 30(9): 3305–3335.
- Fernandez-Mateo, Isabel, and Roberto M Fernandez.** 2016. “Bending the pipeline? Executive search and gender inequality in hiring for top management jobs.” *Management Science*, 62(12): 3636–3655.
- Fernandez, Roberto M, and Marie Louise Mors.** 2008. “Competing for jobs: Labor queues and gender sorting in the hiring process.” *Social Science Research*, 37(4): 1061–1080.
- Friebel, Guido, and Michael Raith.** 2013. “Managers, training, and internal labor markets.” *Simon School Working Paper No. FR 13-31*.
- Groysberg, Boris, and Nitin Nohria.** 2011. “How to hang on to your high potentials.” *Harvard Business Review*, 77–83.
- Haeghele, Ingrid.** 2021. “Talent Hoarding in Organizations.” *Working paper*.
- Human Development Reports.** 2021. <http://hdr.undp.org/en/content/gender-inequality-index-gii>, Accessed: 2021-05-26.
- Kahneman, Daniel.** 2011. *Thinking, fast and slow*. Macmillan.
- Kaplan, Steven N, Mark M Klebanov, and Morten Sorensen.** 2012. “Which CEO characteristics and abilities matter?” *The Journal of Finance*, 67(3): 973–1007.
- Koenig, Anne M, Alice H Eagly, Abigail A Mitchell, and Tiina Ristikari.** 2011. “Are leader stereotypes masculine? A meta-analysis of three research paradigms.” *Psychological bulletin*, 137(4): 616.
- Li, Danielle.** 2017. “Expertise versus Bias in Evaluation: Evidence from the NIH.” *American Economic Journal: Applied Economics*, 9(2): 60–92.

- McGinn, Kathleen L., and Katherine L. Milkman.** 2013. "Looking Up and Looking Out: Career Mobility Effects of Demographic Similarity Among Professionals." *Organization Science*, 24(4): 1041–60.
- Milgrom, Paul, and Sharon Oster.** 1987. "Job discrimination, market forces, and the invisibility hypothesis." *The Quarterly Journal of Economics*, 102(3): 453–476.
- Peter, Laurence J., and Raymond Hull.** 1969. *The Peter Principle*. New York: William Morrow & Co.
- Petersen, Trond, and Ishak Saporta.** 2004. "The opportunity structure for discrimination." *American Journal of Sociology*, 109(4): 852–901.
- Player, Abigail, Georgina Randsley de Moura, Ana C Leite, Dominic Abrams, and Fatima Tresh.** 2019. "Overlooked leadership potential: The preference for leadership potential in job candidates who are men vs. women." *Frontiers in psychology*, 10: 755.
- Prendergast, Canice, and Robert Topel.** 1993. "Discretion and bias in performance evaluation." *European Economic Review*, 37(2-3): 355–365.
- Proudfoot, Devon, Aaron C Kay, and Christy Z Koval.** 2015. "A gender bias in the attribution of creativity: Archival and experimental evidence for the perceived association between masculinity and creative thinking." *Psychological science*, 26(11): 1751–1761.
- Roussille, Nina.** 2020. "The central role of the ask gap in gender pay inequality." *URL: https://ninaroussille.github.io/files/Roussille_askgap.pdf*.
- Sarsons, Heather.** 2017a. "Interpreting signals in the labor market: evidence from medical referrals." *Working paper*.
- Sarsons, Heather.** 2017b. "Recognition for group work: Gender differences in academia." *American Economic Review*, 107(5): 141–45.
- Sarsons, Heather, Klarita Gërkhani, Ernesto Reuben, and Arthur Schram.** 2021. "Gender differences in recognition for group work." *Journal of Political Economy*, 129(1): 101–147.
- Silzer, Rob, and Allan H Church.** 2009. "The pearls and perils of identifying potential." *Industrial and Organizational Psychology*, 2(4): 377–412.
- Tô, Linh T.** 2018. "The signaling role of parental leave." *Harvard University*.
- Yarnall, Jane, and Dan Lucy.** 2015. "Is the Nine Box Grid All About Being in the Top Right?" *Roffrey Park research report*.

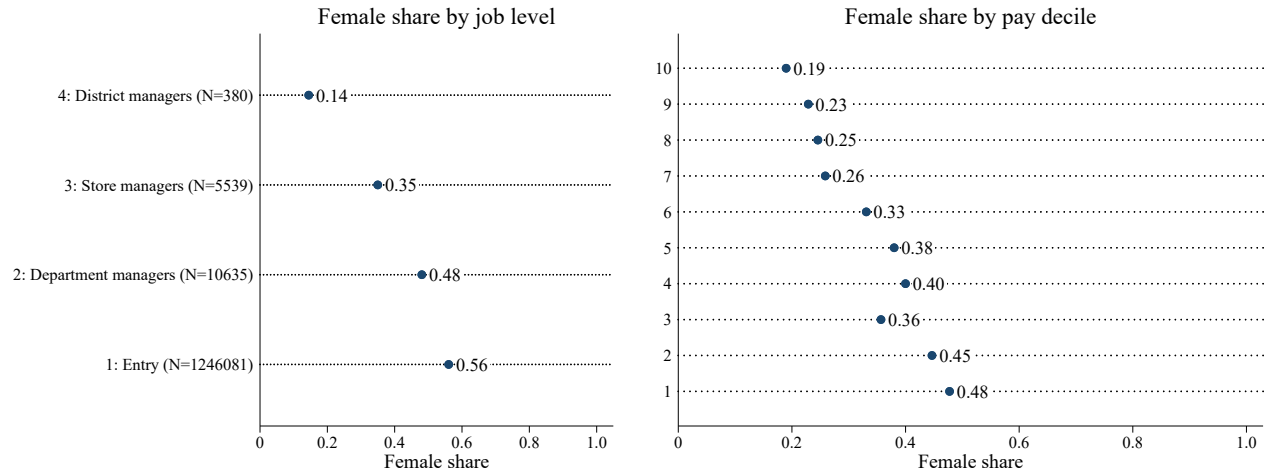
Yun, Myeong-Su. 2005. "A simple solution to the identification problem in detailed wage decompositions." *Economic inquiry*, 43(4): 766–772.

FIGURE 1: NINE BOX RATINGS AND LABELS

		Performance		
		1 (Low)	2 (Medium)	3 (High)
Potential	3 (High)	New hire	Delivering, strong potential	High performing, top talent
	2 (Medium)	Potential mismatch	Delivering, promotable	High performing, promotable
	1 (Low)	Underperforming	Delivering	High performing, critical resource

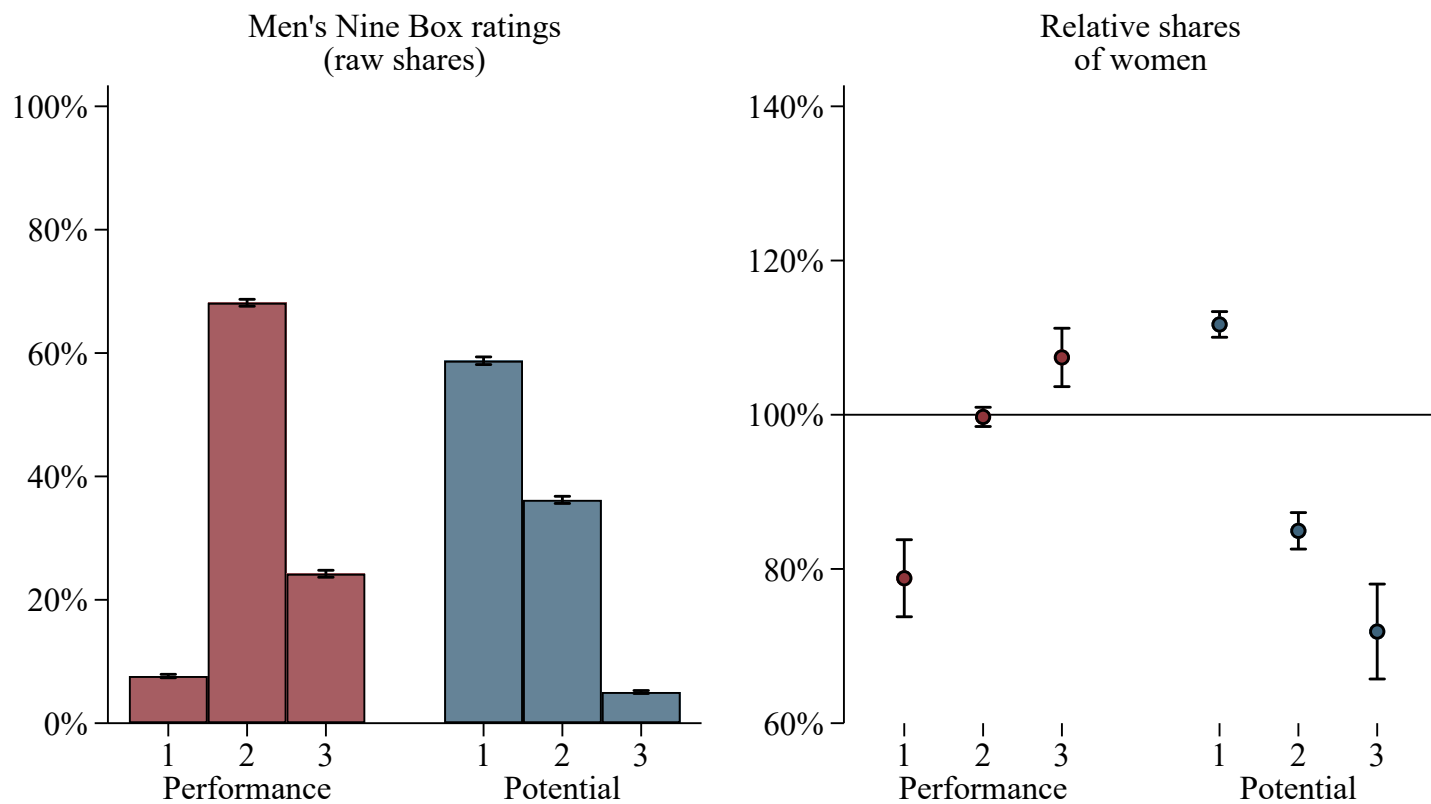
NOTES: This table reports the labels used by the data provider to describe each box within the Nine Box system.

FIGURE 2: FEMALE SHARES IN THE ORGANIZATIONAL HIERARCHY



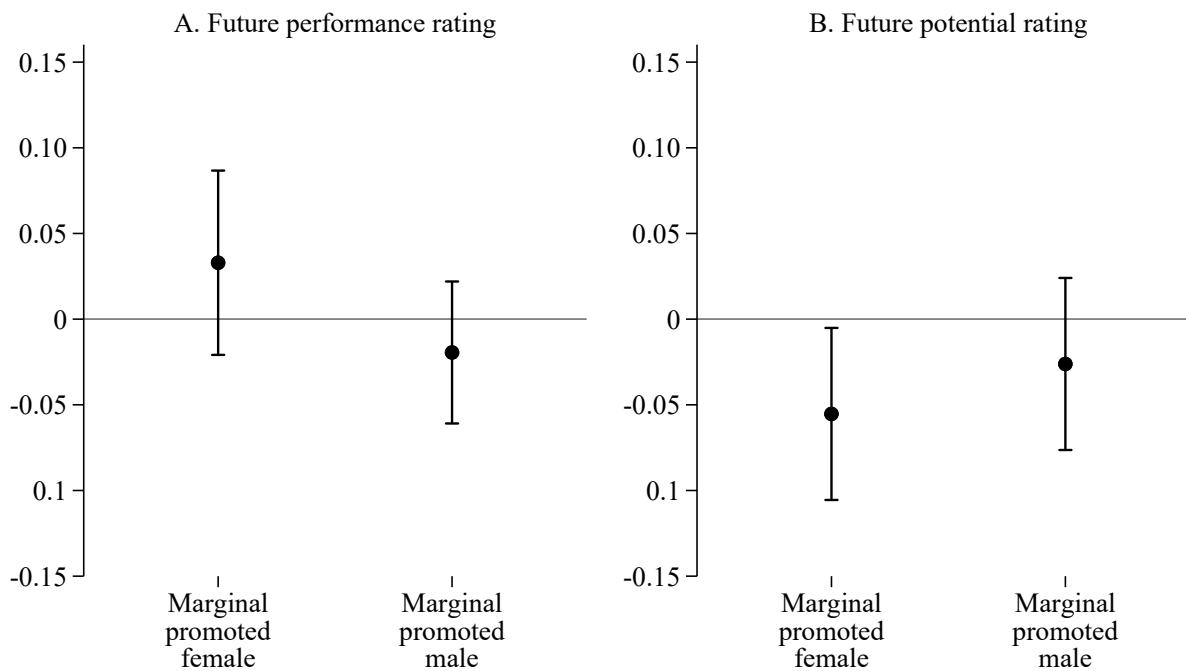
NOTES: The left panel reports the female share among retail operations workers. Department managers include all managers junior to the location's head manager, including associate managers overseeing departments and salaried assistant general managers. Counts include the number of unique workers who held a job at that level. The right panel reports the female share among all workers who receive Nine Box ratings. This population includes all regular, salaried workers, including corporate workers and field workers at the level of department managers and above, and excludes entry-level retail workers. The deciles are sorted by regular annual salaries.

FIGURE 3: GENDER GAP IN NINE BOX SCORES



NOTES: The left panel represents the distribution of Nine Box performance ratings and potential ratings assigned to male employees. The right panel represents the raw difference in shares for female employees relative to male employees. Vertical brackets represent 95% confidence intervals.

FIGURE 4: FUTURE RATINGS FOR MARGINAL PROMOTIONS



NOTES: This figure reports estimates for the IV coefficient on promotion status from Equation (3), as described in the text. Panel A focuses on a promoted worker's 12-month-ahead performance rating, whereas Panel B focuses on that worker's 12-month-ahead potential rating. Vertical brackets represent 95% confidence intervals.

TABLE 1: SUMMARY STATISTICS AND CORRELATIONS

Table 1: Descriptive statistics

Panel A: Data coverage							
Locations	> 4,000	Worker-months	900209				
Workers	29809	Promotions	8964				
Months (2011-2015)	58						
Panel B: Summary statistics		mean	sd	p25	p50	p75	
(a)	Female	.412	.492	0	0	1	
(b)	Promotion (annualized percent)	11.9	119.148	0	0	0	
(c)	Salary (annual dollars)	70691	101974	45000	59188	85000	
(d)	Potential rating	2.18	.536	2	2	2	
(e)	Performance rating	1.429	.578	1	1	2	
(f)	Age	44.4	10.834	35.8	45	53	
(g)	Tenure (months)	171.2	139.534	53	138	253	
(h)	White	.736	.441	0	1	1	
(i)	Black	.09	.287	0	0	0	
(j)	Hispanic	.103	.304	0	0	0	
(k)	Asian	.058	.234	0	0	0	
(l)	Other race	.012	.111	0	0	0	
Panel C: Correlations		(a)	(b)	(c)	(d)	(e)	(f)
(a)	Female	1					
(b)	Promotion (annualized percent)	-.007	1				
(c)*	Salary (annual dollars)	-.132	-.019	1			
(d)	Potential rating	.032	.025	.206	1		
(e)	Performance rating	-.071	.048	.176	.088	1	
(f)*	Age	.037	-.052	.193	.015	-.271	1
(g)*	Tenure (months)	.072	-.039	-.021	.097	-.215	.465

NOTES: Panel A reports data coverage. Tables and figures are estimated using the baseline sample consisting of observations at the worker-month level unless otherwise noted. Panel B presents summary statistics of variables used in our analysis. Panel C presents correlations between these variables. Asterisks denote that salary, age, and tenure are computed as log variables in Panel C and subsequent analyses.

TABLE 2: GENDER GAP IN PROMOTIONS

Promoted	(1)	(2)	(3)	(4)
Female	-1.644*** (0.267)	-1.837*** (0.266)	-1.027*** (0.255)	-1.079*** (0.280)
Performance rating				
2=Med		6.498*** (0.325)	6.061*** (0.331)	5.417*** (0.378)
3=High		11.35*** (0.424)	11.74*** (0.426)	10.99*** (0.482)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes
Location FEs				Yes
Observations	900209	900209	900209	900209

NOTES: This table reports a linear probability model for promotions. The dependent variable takes a value of 1200 if the worker is promoted in the following month, and zero otherwise, so that coefficients represent annualized percents. The omitted category for the performance rating is 1=Low. Demographic controls include log age, log tenure, and race/ethnicity fixed effects for White (omitted category), Black, Asian, Hispanic, and Other. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 3: GENDER PAY GAP AND THE ROLE OF PROMOTIONS

Log salary	(1)	(2)	(3)	(4)
Female	-0.118*** (0.00566)	-0.0364*** (0.00391)	-0.110*** (0.00508)	-0.0416*** (0.00367)
Performance rating				
2=Med			0.134*** (0.00577)	0.0653*** (0.00422)
3=High			0.291*** (0.00719)	0.153*** (0.00508)
Potential rating				
2=Med			0.164*** (0.00402)	0.0829*** (0.00287)
3=High			0.286*** (0.00957)	0.136*** (0.00633)
Fiscal year FEs	Yes	Yes	Yes	Yes
Job level \times year FEs		Yes		Yes
Observations	899023	899023	899023	899023

NOTES: This table reports regressions of log salary on the female indicator, performance rating indicators (the omitted category is 1=Low), potential rating indicators (the omitted category is 1=Low), fiscal year fixed effects, and/or job level interacted with fiscal year fixed effects. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 4: GENDER DIFFERENCES IN NINE BOX RATINGS

Panel A	Performance rating		Potential rating	
	(1)	(2)	(3)	(4)
Female	.0343*** (.0053)	.0151*** (.005)	-.083*** (.0057)	-.0527*** (.0052)
Mean of DV	2.1799 (.0006)	2.1799 (.0006)	1.4288 (.0006)	1.4288 (.0006)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		Yes
Observations	900209	900209	900209	900209
Panel B	Top performance rating		Top potential rating	
	(5)	(6)	(7)	(8)
Female	.0181*** (.0044)	.0061 (.0043)	-.0143*** (.0017)	-.0137*** (.0019)
Mean of DV	.2496 (.0005)	.2496 (.0005)	.0446 (.0002)	.0446 (.0002)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		Yes
Observations	900209	900209	900209	900209

NOTES: This table reports regressions of Nine Box performance and potential ratings on the female indicator and other control variables for fiscal year fixed effects, worker demographics, and location fixed effects as described in Table 2. Panel A uses the raw rating (1, 2, or 3) as the dependent variable whereas Panel B uses an indicator for whether the worker received the top performance or potential rating. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 5: POTENTIAL AND PROMOTIONS

Promoted	(1)	(2)	(3)	(4)
Female	-1.837*** (0.266)	-0.963*** (0.256)	-0.555** (0.256)	-0.726*** (0.279)
Potential rating				
2=Med		10.52*** (0.292)	7.343*** (0.282)	6.838*** (0.299)
3=High		19.50*** (0.864)	14.66*** (0.631)	13.57*** (0.650)
Performance rating				
2=Med	6.498*** (0.325)	6.856*** (0.329)	6.358*** (0.503)	5.921*** (0.536)
3=High	11.35*** (0.424)	10.09*** (0.417)	10.71*** (0.546)	10.38*** (0.588)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes
Location FEs				Yes
Observations	900209	900209	900209	900209

NOTES: This table replicates Table 2, with the addition of control variables for potential rating indicators (the omitted category is 1=Low). By comparing the coefficient on the female indicator in this table with the corresponding coefficient in Table 2, we estimate the fraction of the gender gap in promotions that can be explained gender differences in potential ratings. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 6: BIAS IN POTENTIAL RATINGS AND PROMOTIONS

Panel A	Full sample		Promoted sample	
Next performance rating	(1)	(2)	(3)	(4)
Female	.0328*** (.0046)	.0197*** (.0048)	.0285* (.0154)	.0279* (.0154)
Potential rating				
2=Med	.0913*** (.0048)	.1021*** (.0052)	.0678*** (.0162)	.0665* (.0163)
3=High	.1677*** (.0116)	.1974*** (.0118)	.1266*** (.0278)	.1275*** (.0282)
Performance rating				
2=Med	.3637*** (.0111)	.2613*** (.0116)	.2697*** (.0522)	.2609*** (.0513)
3=High	.7671*** (.0121)	.5801*** (.0126)	.5139*** (.0534)	.4975*** (.0524)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		
Observations	586338	586338	5222	5222
Panel B	Full sample		Promoted sample	
Next potential rating	(5)	(6)	(7)	(8)
Female	-.0482*** (.0047)	-.0346*** (.005)	-.0685*** (.018)	-.0536*** (.0175)
Potential rating				
2=Med	.4241*** (.0055)	.2871*** (.0059)	.2461*** (.0186)	.2074*** (.0184)
3=High	.7297*** (.0167)	.5388*** (.0164)	.4593*** (.0359)	.3816*** (.0353)
Performance rating				
2=Med	.2135*** (.0091)	.1668*** (.0096)	.1082* (.0592)	.0775 (.062)
3=High	.3147*** (.0099)	.2935*** (.0106)	.1795*** (.0602)	.175*** (.063)
Fiscal year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		
Observations	586338	586338	5222	5222

NOTES: Panel A reports a regression of Nine Box performance ratings 12 months in the future on control variables as described in Table 2. Columns 3 and 4 restrict the sample to worker-months corresponding to promotion events and do not control for location fixed effects due to the smaller sample size. Panel B is identical to Panel A except that the dependent variable is the Nine Box potential rating 12 months in the future. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 7: TURNOVER AND GENDER

Attrition	Full sample		High performers	
	(1)	(2)	(3)	(4)
Female	-0.000494* (0.000292)	-0.000379 (0.000298)	0.0000488 (0.000506)	0.000245 (0.000518)
Passed over		0.00688*** (0.00160)		0.00677** (0.00268)
Female × Passed over		-0.00427*** (0.00159)		-0.00651** (0.00261)
Potential rating				
2=Med	0.00233*** (0.000307)	0.00241*** (0.000308)	0.00633*** (0.000539)	0.00646*** (0.000541)
3=High	0.00529*** (0.000691)	0.00545*** (0.000692)	0.00899*** (0.00102)	0.00921*** (0.00102)
Performance rating				
2=Med	-0.0176*** (0.000689)	-0.0176*** (0.000690)		
3=High	-0.0225*** (0.000714)	-0.0224*** (0.000715)		
Fiscal year FEs	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes
Observations	886899	886899	221876	221876
DV mean	.022	.022	.017	.017

NOTES: This table reports regressions of whether a worker leaves the firm entirely in the next month on gender and other measures. The variable “Passed over” is equal to one if another worker sharing the same manager is promoted in the next month, but the focal worker is not. Regressions that include this variable also include a control for whether there is a promotion in that month for this team. Columns 1-2 report results for the full sample of workers (excluding the last year-month observation for any given location to avoid right truncation in our panel data). Columns 3-4 repeat this exercise, but for the subset of workers who receive a current period Nine Box performance rating of 3, the top rating. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 8: RISK OF LOSS AND POTENTIAL RATINGS

	(1) Attrition	(2) Risk of loss rating	(3) Next potential	(4) Promoted	(5) Log salary
Risk of loss rating					
2=Med	0.00414*** (0.000391)		0.0834*** (0.00644)	0.000489 (0.000299)	0.104*** (0.00507)
3=High	0.0115*** (0.000865)		0.0918*** (0.0136)	0.00324*** (0.000631)	0.134*** (0.00995)
Female		-0.0547*** (0.00685)	-0.0432*** (0.00561)	-0.000774*** (0.000262)	-0.117*** (0.00538)
Potential rating					
2=Med	-0.000855** (0.000375)	0.165*** (0.00665)	0.399*** (0.00653)	0.00919*** (0.000308)	0.105*** (0.00467)
3=High	-0.000921 (0.000855)	0.356*** (0.0175)	0.694*** (0.0194)	0.0164*** (0.000972)	0.191*** (0.0118)
Performance rating					
2=Med	-0.0193*** (0.000864)	-0.0900*** (0.0127)	0.228*** (0.0106)	0.00529*** (0.000349)	0.149*** (0.00672)
3=High	-0.0257*** (0.000886)	-0.00663 (0.0140)	0.323*** (0.0116)	0.00876*** (0.000441)	0.302*** (0.00811)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes	Yes
Observations	533780	533780	415683	533780	532996
DV mean	.019	1.421	1.399	1.421	1.421

NOTES: Column 1 reports regressions of actual attrition in the next month on “risk of loss” ratings assigned by the firm and other control variables. Risk of loss is categorized by the firm as 1-low, 2-medium, or 3-high. Column 2 regresses the risk of loss rating on the female indicator and other control variables. Columns 3-5 examine the relationship between risk of loss and gender with the 12-month-ahead potential rating, whether a worker is promoted in the following month, and log salary, respectively. See the appendix for a parallel analysis controlling for demographics and location fixed effects. The sample includes all worker months prior to the last month that given location is in our sample, to allow for observations of future behavior. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 9: LEAVE OF ABSENCES AND THE GENDER POTENTIAL GAP

	(1)	(2)	(3)	(4)
	Leave of absence	Potential rating	Potential rating	Potential rating
Female	5.431*** (0.430)	-0.0856*** (0.00562)	-0.0840*** (0.00562)	-0.0817*** (0.00562)
Past leaves			-0.0231*** (0.00405)	-0.0185*** (0.00396)
Future leaves				-0.0269*** (0.00298)
Fiscal year FEs	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes
Observations	886899	886899	886899	886899

NOTES: Column 1 present a regression of whether a worker takes a leave of absence in the next month on the female indicator and other control variables as described in Table 2. Columns 2-4 relate gender and leaves to Nine Box potential ratings. Past leaves measures the number of months of leave a worker has taken in their past history with the firm and future leaves measures the number of months of leave taken in the future within our data sample. Results with additional demographic and location controls are in the appendix. The sample includes all worker months prior to the last month that given location is in our sample, to allow for observations of future behavior. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 10: VARIATION BY MANAGER GENDER

	(1)	(2)	(3)	(4)	(5)
	Performance rating	Potential rating	Log salary	Promoted	Promoted
Female	0.0157*** (0.00564)	-0.0582*** (0.00582)	-0.168*** (0.00487)	-1.166*** (0.331)	-0.864*** (0.326)
Manager female	-0.0200*** (0.00695)	-0.0286*** (0.00732)	-0.117*** (0.00558)	-0.187 (0.481)	0.105 (0.477)
Female × Manager female	-0.000778 (0.00945)	0.0193** (0.00970)	0.0292*** (0.00731)	0.463 (0.622)	0.333 (0.615)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes	Yes
Worker ratings FEs					Yes
Observations	885353	885353	884200	885353	885353

NOTES: This table examines how the gender gaps in ratings, compensation, and promotions vary with the gender of the worker's immediate manager. Manager female is an indicator for whether the manager is female. Worker rating fixed effects are indicators for the worker's Nine Box performance and potential ratings. All other variables are as defined in previous tables. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE 11: VARIATION BY MANAGER RATINGS

	(1)	(2)	(3)	(4)	(5)
	Performance rating	Potential rating	Log salary	Promoted	Promoted
Female	0.0449** (0.0187)	0.00281 (0.0189)	-0.135*** (0.0123)	1.110 (1.264)	0.861 (1.254)
Manager potential (1-3)	0.0237*** (0.00374)	0.0374*** (0.00400)	0.0584*** (0.00261)	0.689** (0.272)	0.333 (0.271)
Manager performance (1-3)	0.0543*** (0.00449)	0.0364*** (0.00465)	0.0303*** (0.00297)	1.075*** (0.316)	0.571* (0.313)
Female × Manager potential (1-3)	-0.0167*** (0.00566)	-0.0122** (0.00601)	-0.00346 (0.00387)	-0.0327 (0.402)	0.123 (0.399)
Female × Manager performance (1-3)	-0.00131 (0.00668)	-0.0156** (0.00670)	-0.00675 (0.00429)	-0.911** (0.454)	-0.794* (0.451)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes	Yes
Worker ratings FEs					Yes
Observations	829429	829429	828417	829429	829429

NOTES: This table examines how the gender gaps in ratings, compensation, and promotions vary with the performance and potential ratings of the worker's immediate manager. Manager potential and manager performance are variables equal to 1, 2, or 3, representing the manager's Nine Box potential and performance ratings, respectively. All other variables are as defined in previous tables. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

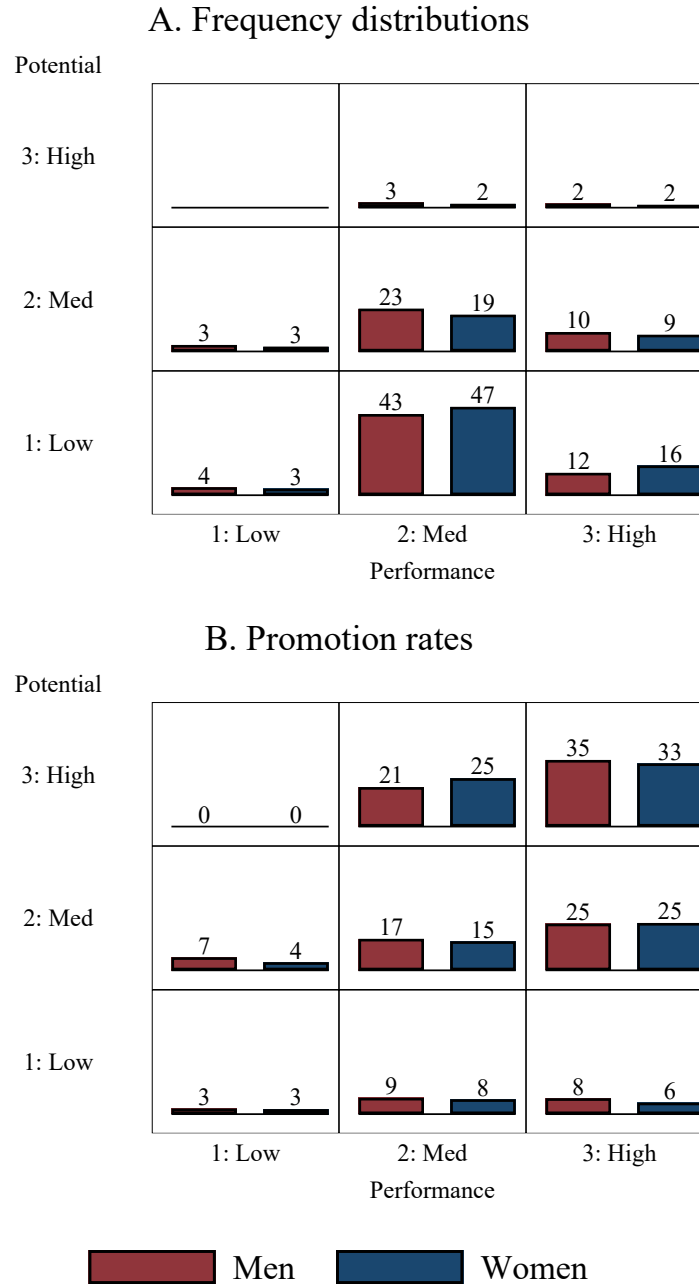
TABLE 12: PROMOTION AND PERFORMANCE UNDER COUNTERFACTUAL POLICIES

	Promotion rates		Expected next performance rating among promoted		
	(1)	(2)	(3)	(4)	(5)
	among men	among women	full sample	true unpromoted sample	true promoted sample
Baseline: current promotion policy	12.6232 (.1815)	10.9882 (.1426)	2.2933 (.0045)	2.2936 (.0045)	2.2743 (.0091)
Counterfactual 1: ignore potential scores and gender	12.2036 (.1492)	11.5865 (.1426)	2.2772 (.0041)	2.2774 (.0041)	2.2633 (.0092)
Counterfactual 2: add one to the potential scores of all women	12.6232 (.1776)	18.0554 (.3682)	2.2797 (.0041)	2.2798 (.0041)	2.2711 (.0091)
Counterfactual 3: add one to the potential scores of high performing women	12.6232 (.1732)	12.7829 (.2233)	2.3111 (.0046)	2.3115 (.0046)	2.2822 (.0093)

NOTES: This table reports expected promotion rates and expected future performance ratings under the firm's current promotion policy and counterfactual promotion policies. Details are provided in Section 6.2. Columns 1 and 2 provide the counterfactual expectations of promotion rates for men and women, respectively. Column 3 provides expectations of the 12-month-ahead performance ratings among the promoted, weighted by the current estimated promotion probabilities for all workers. Column 4 does the same, but for workers who were not promoted in the true sample. Column 5 does the same, but for workers who were promoted in the true sample. The baseline policy uses predicted values of promotion rates based on gender, performance ratings, potential ratings, year, age, tenure, and race/ethnicity. Counterfactual 1 uses the baseline policy, but omits gender and potential ratings when estimating promotion rates. Counterfactual 2 adds one to the potential ratings of women who receive potential ratings of 1 or 2 when estimating predicted promotion rates. Counterfactual 3 adds one to the potential ratings of women who receive potential ratings of 1 or 2, but only for women who receive performance ratings of 3. Bootstrapped standard errors, clustered by worker, are in parentheses.

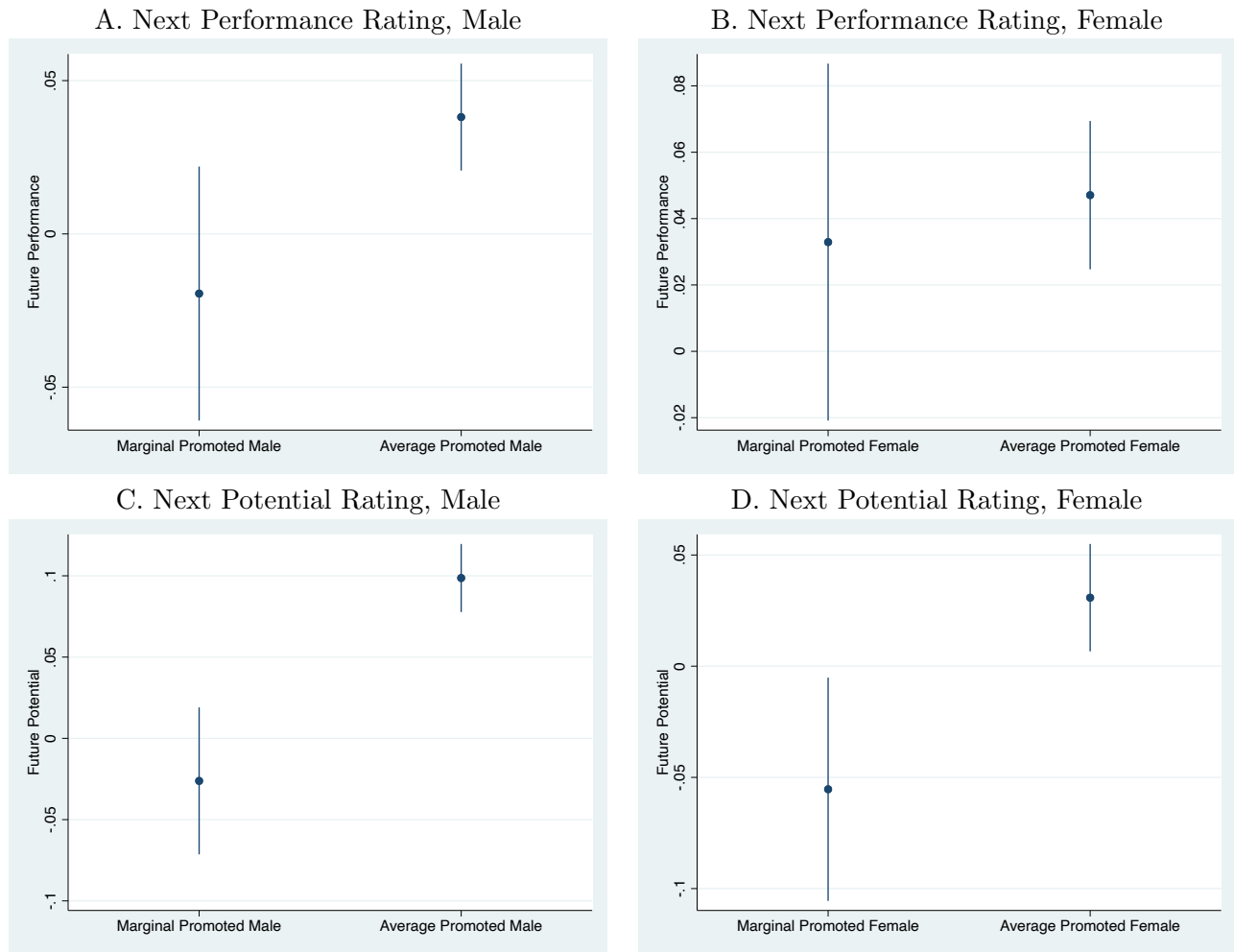
8 Appendix

FIGURE A1: DISTRIBUTION OF NINE BOX RATINGS AND PROMOTIONS



NOTES: The top panel provides the share of men and women receiving each Nine Box rating. The bottom panel provides the annual promotion rate conditional on receiving each Nine Box rating for men and women. We exclude observations rated a low performance and high potential (the top left box) from our sample, because that rating is reserved by our firm for new hires.

APPENDIX FIGURE A2: FUTURE PERFORMANCE AND POTENTIAL RATINGS FOR AVERAGE AND MARGINALLY PROMOTED WORKERS



NOTES: This figure reports estimates the OLS and IV coefficients on promotion status from Equation (3). Panels A and B focuses on a promoted worker’s future performance rating, whereas Panels B and C focuses on that worker’s future potential rating. In each case, outcomes for the average promoted worker exceed outcomes for the marginally promoted worker.

APPENDIX TABLE A1: DECOMPOSING THE EFFECT OF RATINGS ON THE PROMOTION GAP

	Coefficient	Standard error
Overall gap		
Men	12.623***	(.175)
Women	10.988***	(.196)
Difference	1.635***	(.263)
Endowments	.735***	(.077)
Coefficients	1.005***	(.252)
Interaction	-.105**	(.045)
Endowments		
Performance rating	-.159***	(.022)
Potential rating	.9***	(.07)
Fiscal year effects	-.006	(.016)
Coefficients		
Performance rating	-.313	(.195)
Potential rating	1.023*	(.554)
Fiscal year effects	.027	(.046)
Constant	.269	(.669)
Interaction		
Performance rating	-.023	(.016)
Potential rating	-.077*	(.044)
Fiscal year effects	-.005	(.004)

NOTES: This table reports results from a Oaxaca-Blinder-Kitigawa decomposition. Performance, potential, and fiscal year variables are treated as dummies with normalized effects. Standard errors are clustered by worker.

APPENDIX TABLE A2: VARIATION BY LOCAL CHARACTERISTICS

Potential rating	(1)	(2)	(3)
Female	0.135 (0.0937)	0.124** (0.0627)	-0.296*** (0.0521)
County's management gap	-1.310*** (0.106)		
Female × County's management gap	-0.379** (0.159)		
County's pay gap		-0.417*** (0.0303)	
Female × County's pay gap		-0.151*** (0.0450)	
County's female educational attainment			0.529*** (0.0566)
Female × County's female educational attainment			0.327*** (0.0812)
Fiscal year FEs	Yes	Yes	Yes
Observations	780753	780753	780753

NOTES: This table shows how the gender gap in potential ratings varies with county-level characteristics. *County management gap* is the fraction of men among workers with management standard occupational classification (SOC) codes. *County pay gap* is men's median earnings divided by women's median earnings. *County female educational attainment* is the fraction of women over the age of 18 with at least some college education. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

APPENDIX TABLE A3: FUTURE PERFORMANCE RATINGS, IV ESTIMATES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Future Performance	OLS Female	OLS Male	IV Female	IV Male	OLS Female	OLS Male	IV Female	IV Male
Promoted	0.0471*** (0.0114)	0.0381*** (0.00889)	0.0329 (0.0274)	-0.0195 (0.0211)	0.0465*** (0.0114)	0.0392*** (0.00888)	0.0293 (0.0276)	-0.0171 (0.0212)
Potential rating								
2=Med	0.000532*** (0.000206)	0.000620*** (0.000185)	0.000691* (0.000354)	0.00117*** (0.000284)	0.000633** (0.000253)	0.000478** (0.000228)	0.000785** (0.000346)	0.000863*** (0.000294)
3=High	0.000809 (0.000813)	0.00262*** (0.000594)	0.00109 (0.00109)	0.00371*** (0.000795)	0.000827 (0.000869)	0.00271*** (0.000632)	0.00106 (0.00112)	0.00365*** (0.000796)
Performance rating								
2=Med	0.000169 (0.000256)	0.000546* (0.000281)	0.000389 (0.000312)	0.000803*** (0.000289)	-0.0000524 (0.000333)	0.000433 (0.000328)	0.000145 (0.000377)	0.000568* (0.000328)
3=High	0.00218*** (0.000313)	0.00228*** (0.000333)	0.00208*** (0.000418)	0.00275*** (0.000393)	0.00200*** (0.000410)	0.00197*** (0.000388)	0.00180*** (0.000511)	0.00230*** (0.000437)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics and Location					Yes	Yes	Yes	Yes
Observations	228680	315104	228680	315104	228680	315104	228680	315104

NOTES: This table reports coefficients estimated from OLS and IV specifications related to Equation (3), with the outcome of interest being a worker's next period performance rating. As described in the main text, the OLS estimates are interpreted as the residualized post-promotion performance rating of the average male or female worker who is promoted in the next month. The IV estimates correspond to the post-promotion performance rating of the marginal male or female worker who is promoted. Marginal workers are defined as IV compliers, those who are promoted only because they received a lucky draw of the instrument. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

APPENDIX TABLE A4: FUTURE POTENTIAL RATINGS, IV ESTIMATES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Future Potential	OLS Female	OLS Male	IV Female	IV Male	OLS Female	OLS Male	IV Female	IV Male
[-0.5em] Promoted	0.0308** (0.0123)	0.0986*** (0.0107)	-0.0553** (0.0256)	-0.0262 (0.0231)	0.0312** (0.0124)	0.100*** (0.0107)	-0.0590** (0.0264)	-0.0296 (0.0238)
Potential rating								
2=Med	0.00117*** (0.000220)	0.00134*** (0.000211)	0.00197*** (0.000362)	0.00254*** (0.000327)	0.000507* (0.000279)	0.000789*** (0.000270)	0.000987*** (0.000374)	0.00171*** (0.000349)
3=High	0.00308*** (0.000972)	0.00531*** (0.000774)	0.00366*** (0.00125)	0.00810*** (0.00105)	0.00210** (0.00103)	0.00503*** (0.000820)	0.00220* (0.00129)	0.00756*** (0.00107)
Performance rating								
2=Med	0.000241 (0.000216)	-0.0000409 (0.000280)	0.000850*** (0.000272)	0.000537* (0.000313)	-0.000129 (0.000293)	-0.000213 (0.000346)	0.000307 (0.000344)	0.000130 (0.000372)
3=High	0.000858*** (0.000287)	0.000966*** (0.000344)	0.00152*** (0.000393)	0.00206*** (0.000446)	0.000945** (0.000385)	0.000898** (0.000433)	0.00152*** (0.000502)	0.00188*** (0.000527)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics and Location					Yes	Yes	Yes	Yes
Observations	228680	315104	228680	315104	228680	315104	228680	315104

NOTES: This table reports coefficients estimated from OLS and IV specifications related to Equation (3), with the outcome of interest being a worker's next period potential rating. As described in the main text, the OLS estimates are interpreted as the residualized post-promotion potential rating of the average male or female worker who is promoted in the next month. The IV estimates correspond to the post-promotion potential rating of the marginal male or female worker who is promoted. Marginal workers are defined as IV compliers, those who are promoted only because they received a lucky draw of the instrument. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

APPENDIX TABLE A5: ROBUSTNESS: MANAGER FIXED EFFECTS

Panel A	Full sample		Promoted sample	
Next performance rating	(1)	(2)	(3)	(4)
Female	.0294*** (.005)	.0306*** (.0054)	.021* (.0248)	.0201* (.025)
Potential rating				
2=Med	.1025*** (.0051)	.1051*** (.0054)	.0658** (.0271)	.0641* (.0273)
3=High	.1778*** (.0117)	.1887*** (.012)	.1138*** (.0434)	.1099** (.0447)
Performance rating				
2=Med	.2909*** (.0115)	.2308*** (.012)	.2695*** (.0931)	.2477*** (.0909)
3=High	.5892*** (.0125)	.4694*** (.0131)	.4325*** (.0945)	.4091*** (.0926)
Manager FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		
Observations	586014	586014	5196	5196
Panel B	Full sample		Promoted sample	
Next potential rating	(5)	(6)	(7)	(8)
Female	-.0508*** (.0051)	-.0273*** (.0055)	-.0884*** (.03)	-.0517* (.0293)
Potential rating				
2=Med	.3232*** (.006)	.2198*** (.0061)	.1517*** (.0318)	.1072* (.031)
3=High	.5813*** (.0163)	.4181*** (.016)	.3224*** (.0565)	.2294*** (.0554)
Performance rating				
2=Med	.1958*** (.0095)	.1606*** (.0098)	.0966* (.0979)	.0495 (.0997)
3=High	.3069*** (.0105)	.2835*** (.0109)	.1614*** (.1004)	.1317*** (.1018)
Manager FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Demographic controls		Yes		Yes
Location FEs		Yes		
Observations	586014	586014	5196	5196

NOTES: This table replicates Table 6 with the addition of fixed effects for the worker's immediate manager in each time period.

APPENDIX TABLE A6: INCREASING GENDER GAPS BY PAY DECILE

	(1)	(2)	(3)	(4)	(5)
	Performance rating	Potential rating	Log salary	Promoted	Promoted
Female	0.0864*** (0.0192)	0.0261 (0.0181)	-0.0399*** (0.00622)	8.713*** (1.690)	7.968*** (1.676)
Pay decile	0.0311*** (0.00203)	0.0245*** (0.00213)	0.103*** (0.000706)	-0.782*** (0.155)	-1.169*** (0.160)
Female \times Pay decile	-0.00486 (0.00337)	-0.0120*** (0.00340)	-0.000464 (0.00115)	-2.002*** (0.298)	-1.882*** (0.295)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes	Yes
Worker ratings FEs					Yes
Observations	160232	160232	160029	160232	160232

This table examines how the gender gaps in ratings, compensation, and promotions vary with pay decile. Pay decile is measured using the ranking of salary within each fiscal year. Column 5 includes controls for worker ratings fixed effects, consisting of performance rating indicators and potential rating indicators. All other variables are as defined in Table 2. Standard errors are clustered by worker. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

TABLE A7: TURNOVER AND GENDER, FULL CONTROLS

Attrition	Full sample		High performers	
	(1)	(2)	(3)	(4)
Female	-0.00142*** (0.000292)	-0.00131*** (0.000296)	-0.000755 (0.000543)	-0.000542 (0.000555)
Passed over		0.00866*** (0.00161)		0.00842*** (0.00273)
Female × Passed over		-0.00452*** (0.00161)		-0.00685** (0.00266)
Potential rating				
2=Med	-0.00205*** (0.000326)	-0.00199*** (0.000326)	0.000901 (0.000630)	0.00101 (0.000630)
3=High	-0.000508 (0.000702)	-0.000367 (0.000702)	0.00163 (0.00111)	0.00182 (0.00111)
Performance rating				
2=Med	-0.0189*** (0.000697)	-0.0189*** (0.000697)		
3=High	-0.0230*** (0.000732)	-0.0229*** (0.000732)		
Fiscal year FEs	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Observations	886899	886899	221876	221876
DV mean	.022	.022	.017	.017

NOTES: This table replicates Table 7 with the addition of control variables for demographics and location fixed effects.

TABLE A8: RISK OF LOSS AND POTENTIAL RATINGS, FULL CONTROLS

	(1) Attrition	(2) Risk of loss rating	(3) Next potential	(4) Promoted	(5) Log salary
Risk of loss rating					
2=Med	0.00285*** (0.000408)		0.0451*** (0.00659)	-0.000427 (0.000301)	0.105*** (0.00479)
3=High	0.00955*** (0.000879)		0.0321** (0.0134)	0.00198*** (0.000629)	0.132*** (0.00941)
Female		-0.0568*** (0.00685)	-0.0320*** (0.00586)	-0.000517** (0.000260)	-0.116*** (0.00512)
Potential rating					
2=Med	-0.00329*** (0.000406)	0.0998*** (0.00675)	0.277*** (0.00697)	0.00707*** (0.000317)	0.151*** (0.00451)
3=High	-0.00452*** (0.000888)	0.260*** (0.0172)	0.532*** (0.0191)	0.0132*** (0.000972)	0.267*** (0.0111)
Performance rating					
2=Med	-0.0193*** (0.000897)	-0.107*** (0.0127)	0.172*** (0.0111)	0.00482*** (0.000350)	0.161*** (0.00659)
3=High	-0.0242*** (0.000933)	-0.00640 (0.0139)	0.295*** (0.0124)	0.00905*** (0.000446)	0.300*** (0.00788)
Fiscal year FEs	Yes	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Observations	533780	533780	415683	533780	532996
DV mean	1.421	1.421	1.399	1.421	1.421

NOTES: This table replicates Table 8 with the addition of control variables for demographics and location fixed effects.

TABLE A9: LEAVES OF ABSENCE AND THE GENDER POTENTIAL GAP, FULL CONTROLS

	(1)	(2)	(3)	(4)
	Leave of absence	Potential rating	Potential rating	Potential rating
Female	5.300*** (0.438)	-0.0520*** (0.00507)	-0.0629*** (0.00510)	-0.0488*** (0.00507)
Past leaves			-0.0236*** (0.00378)	-0.0198*** (0.00365)
Future leaves				-0.0205*** (0.00258)
Fiscal year FEs	Yes	Yes	Yes	Yes
Performance FEs	Yes	Yes	Yes	Yes
Location FEs	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Observations	886899	886899	886899	886899

NOTES: This table replicates Table 9 with the addition of control variables for demographics and location fixed effects.

County-level labor market gender inequality measures

We construct labor market gender inequality measures for US counties based on the methodology in Human Development Reports (2021). The county level variables were collected from the 2019 US Census Bureau five year estimates from the American Community Survey (2019). In Human Development Reports, gender-based inequality is measured using fifteen variables in three dimensions, including many measures focused on health, fertility, and mortality. We focus on three variables tied to labor market outcomes with a focus on upper level management: *County management gap* is the fraction of men among workers with management standard occupational classification (SOC) codes. *County pay gap* is men's median earnings divided by women's median earnings. *County female educational attainment* is the fraction of women over the age of 18 with at least some college education.