Essays in Labor Economics

Dissertation Advisor:
Amanda Pallais

Author:
Lisa Abraham

Essays in Labor Economics

Abstract

I study the impact of gender in the labor market. My dissertation examines gender gaps in economic outcomes and provides insights on the underlying mechanisms.

Chapter 1, co-authored with Alison Stein, examines whether women are discouraged from applying to job postings because they believe they must meet all the listed qualifications. To test this hypothesis, we ran a randomized controlled trial on a sample of 60,000 potential applicants to over 600 of Uber's corporate U.S. job postings. We find that job seekers are responsive to removing optional qualifications and softening language about the intensity of the required qualifications. Our treatment increased the total number of applications. It also closed the gender skill gap: while female applicants in the control group are more likely to have graduate degrees than men applying for the same job, men and women in the treatment group are equally likely to have graduate degrees.

In Chapter 2, I use proprietary data from an employee performance management software company to examine gender differences in over 20,000 performance evaluations for a sample of approximately 200 companies. I find gender differences in the way that females and males publicly present themselves in formal evaluations: women rate themselves significantly lower than males after accounting for manager beliefs. This gender gap is larger for workers with lower tenure, who have had less time to build their reputation.

Chapter 3, co-authored with Matthew Gibson, provides evidence on the gender wage gap using a proprietary dataset of online self-reports from Glassdoor.com. These data allow us to control for detailed occupation and employer fixed effects, typically not permissible in publicly available data. Estimates of the gender wage gap range from 6 to 8 percent, smaller than in prior studies. Our analysis using Glassdoor reviews data finds that females are less likely than comparable men to list pay as a negative factor when providing anonymous feedback about their company. In addition to being paid less than their male counterparts, women seem to be less vocal about salary concerns.

Contents

List of Tables

List of Figures

Acknowledgments

I am deeply indebted to my advisors Claudia Goldin, Nathan Hendren, Larry Katz, and Mandy Pallais, for all their support, care, and guidance through every stage of the doctorate program. My advisors are exceptional because they believed in me and provided crucial encouragement while also setting high expectations for me, thereby motivating me to be a better researcher. I am lucky to have advisors who I admire and respect as much for their remarkable personal qualities as for their research and teaching.

In addition to my committee, I am very grateful to several people who have significantly impacted my time at Harvard. I thank Marty Feldstein for being a role model and mentor to me; Brenda Piquet for her care and support; and my friends at Harvard who have been there for me through the highs and lows of this six-year journey. I am grateful to my partners and data providers at each of the three companies I worked with for facilitating our unique collaboration. I also gratefully acknowledge generous financial support from the Weatherhead Initiative on Gender Inequality.

I would not be graduating without all the many people who have helped me on the long road to obtaining this PhD: my teachers at Franklin, Stanton, Reserve, and Wellesley who encouraged and supported me; my mentors at JPM and Treasury who have stayed in touch with me; and my incredible friends and family who have been a constant source of love and support. Thank you to my sister for always setting the bar high. And most of all, thank you to my mom and dad, the original Dr. Abrahams, for everything they have done for me and for teaching me the most important lessons in life. This dissertation is dedicated to my grandparents for all the gifts they have given to me.

To U.K. and Aymanam Ammachy & Appacha

Introduction

This dissertation is comprised of three chapters focused on gender differences in the labor market. All three chapters use novel data sources to examine gender gaps that exist at several critical stages in a worker's career: at the point of application before they have the job, and at the point of being reviewed and compensated once they obtain the job.

While existing literature has focused on gender differences at earlier points of human capital formation (e.g., in school) or at later points in a woman's career (e.g., deciding to take time off after having children), much less work examines differential sensitivity to language in job applications, which could materially impact longer-run economic outcomes. In the first chapter of my dissertation, we ran a randomized controlled trial with Uber to examine whether women are more sensitive than men in terms of meeting all of the listed qualifications in a job posting. Our analysis provides evidence that lower skill women are particularly sensitive to language about qualifications relative to their male counterparts, and so disproportionately hold themselves back from applying when the posting includes strong language about credentials and optional requirements. Particularly if hiring processes are noisy, our results suggest that females with a bachelor's degree or lower might be removing themselves from consideration for jobs that they could otherwise be hired for. Differential sensitivity to language materially affects economic outcomes.

Gender differences in behavior continue even after the individual has obtained the job and is participating in the performance review process. Using proprietary data from an employee performance management software company, the second chapter of my dissertation examines performance evaluations from a broad sample of companies and industries. I find that females review themselves more negatively relative to males in comparison to their manager's review. This gender gap is significantly larger for workers with lower tenure, who have had less time to build their reputation. If these results are due to females rating themselves lower than their true ability (as opposed to managers rating female workers higher than their true ability), this suggests that there is a gender gap in self-promotion. More research is necessary to confirm the source of these gender gaps (worker versus manager) and to understand the underlying mechanisms (e.g., backlash from managers, under-confidence, etc.).

Once on the job, women are also paid less: we provide evidence that a gender wage gap, on the order of 6 to 8 percent, still persists in a proprietary dataset of online self-reports from Glassdoor.com. These data critically allow us to control for detailed occupation and employer fixed effects, typically not available in publicly available data. Using a subsample of the salary reports where workers provided an anonymous company review, we find that women are significantly less likely than comparably paid men to list pay as a negative factor. In the same way that women might be less likely to self-promote, these results indicate that women are also less vocal about salary concerns, in line with existing literature on gender differences in the propensity to negotiate.

Taken together, these chapters provide suggestive evidence that women have a greater tendency than men to reset beliefs about themselves at each stage of their careers: females screen themselves out of getting the job, and then continue to rate their job performance lower than males (relative to their manager's review) once they have the job. They are also paid less and complain less about compensation once in the position. What are the implications of these behavioral differences at other points in an individual's life? Understanding these factors can help narrow gaps in economic outcomes for men and women.

Chapter 1

Words Matter: Experimental Evidence from Job Applications[1]

1.1 Introduction

Women are underrepresented in certain jobs, particularly in the technology sector and in STEM fields. The dropoff begins in school, as more women than men in the U.S. obtain a bachelor's degree (58 percent to 42 percent), but a lower share of these women receive a bachelor's degree in a STEM-related field (36 percent women to 64 percent men). Women in the U.S. hold less than 26 percent of computer and mathematical-related positions and less than 16 percent of architecture and engineering positions. At major tech companies such as Apple, Facebook, and Google, women represent less than 26 percent of technical positions.[2] This underrepresentation has motivated efforts to find ways of increasing the number of qualified women in these roles, thereby reducing gender disparities in employment without compromising firm productivity.

[1] Co-authored with Alison Stein (Uber)

[2] Education data are from the U.S. Department of Education, National Center for Education Statistics' 2015-2016 academic year (National Center for Education Statistics (2019)). See Ginther and Kahn (2017) for a detailed breakdown of differences within STEM. Occupational data are annual averages from the 2018 U.S. Bureau of Labor Statistics (U.S. Department of Labor Bureau of Labor Statistics (2019)). Data on Apple, Facebook, and Google reflect 2018 global workforce statistics (Apple (2019); Facebook (2019); Google (2019)).

This paper evaluates one potential mechanism responsible for female underrepresentation: many women in tech – drawing on their own personal experiences and anecdotal evidence – argue that they hold themselves to higher standards than their male counterparts when deciding whether or not to apply for a job.[3] Thus, job postings that ask for "exceptional" expertise and a slew of bonus qualifications may disproportionately discourage women from applying if women believe they must meet all the listed qualifications. Men, the hypothesis goes, are less affected by the intensity of the language, and so are more willing to apply. If this is the case, then women may be differentially screening themselves out of jobs for which they might be qualified solely due to differences in how each gender perceives language about the expected qualifications.[4]

We test the hypothesis that women and men are differentially sensitive to language about qualifications by running a randomized controlled trial (RCT). Partnering with the ride-sharing company Uber, we alter the language in job postings for Uber's U.S. corporate positions. From August 2018 through December 2018, all viewers who visited Uber's careers website (approximately 200,000 viewers) were randomized into either a Treated group or a Control group. Of these viewers, approximately 60,000 clicked on one of the 616 job postings included in our experiment. If randomized into the Control group, the viewer would see the original (control) version for all of the 616 job postings. If randomized into the Treated group, the viewer would see the "treated" version for all of the 616 job postings, where this treated version removed optional qualifications (e.g., "PhD preferred"), removed adjectives

[3]This paper does not consider many other reasons for female underrepresentation (e.g., differences in preferences over occupational tasks or workplace culture, penalties for time-off during and after childbirth, and lack of flexibility in working hours). See Blau and Kahn (2017) for a review of the literature on gender differences in the labor market.

[4]Sheryl Sandberg's *Lean In* (2013) provides anecdotal evidence of this self-screening mechanism (Sandberg (2013)). She cites an internal Hewlett-Packard report which found that men apply for jobs when they meet 60 percent of the requirements, whereas women only apply if they meet 100 percent of the listed requirements. See also Tara Sophia Mohr, "Why Women Don't Apply for Jobs Unless They're 100% Qualified", Harvard Business Review (Mohr (2014)), and simultaneous work by Coffman *et al.* (2019) who conduct an experiment on an online labor platform to test whether the addition of clearly stated qualifications affect female and male applications for an hour-long essay-writing job. Choices about which firms to apply to could also explain gender wage differences; evidence from Card *et al.* (2015) using Portuguese data suggests that female sorting into lower-paying firms relative to males explains 15 percent of the log gender wage gap.

describing qualifications (e.g., deleted "excellent" before "coding skills"), and softened the language describing qualifications (e.g., "SQL fluency" was changed to "experience with SQL").

Our treatment significantly increased the total number of applications by 7 percent: 3,057 applications in the Control group versus 3,273 applications in the Treated group. Applications from women increased by 5 percent and applications from men increased by 7 percent.[5] Because the treatment induced both men and women to apply in roughly equal magnitudes, the intervention did not significantly change the fraction of women who applied.

Changing the language, however, closed the gender skill gap. Women applying in the control group are 6 percentage points significantly more likely to have advanced educational degrees (i.e., higher than a bachelor's degree) relative to men applying for the same job. The same pattern holds for a composite measure of skill, constructed using all application attributes (e.g., educational attainment, eliteness of educational institution, educational degree field, technical skills, work experience, social skills, and foreign language skills) and information about whether or not the applicant received an interview. Female applicants seem to be more skilled than male applicants, particularly with respect to educational attainment.

Our treatment increases applications from women with a bachelor's degree or lower (10 overall percent increase), and decreases applications from women with advanced educational degrees (2 percent overall decrease). By contrast, the treatment causes men to apply independent of their level of education. As a result of the treatment's different effects by gender, treated women are on average equally likely as treated men to have advanced degrees for the same job. Treated females are also equally likely as treated males to be skilled using our measure of composite skill. Our treatment provides further evidence that choices about applying are correlated with skill for women but not men.

[5] There were 145 applications associated with applicants who chose to "opt-out" of providing their gender on the application form.

These results are consistent with a model in which job seekers' application decisions are influenced by their confidence in being qualified for the job, their perception of the relative competition, and their sensitivity to job match quality. Confidence about being qualified and/or sensitivity to job match quality seem to prevent women with a bachelor's degree or lower from applying relative to their male counterparts. In line with our model predictions, our treatment increases applications from lower skill females more so than from lower skill males. For high-skill job seekers, our model predicts ambiguous effects; empirically, we find that women with an advanced degree seem to be more sensitive to job match quality than their male counterparts.

Our RCT relates to three broad strands of literature. First, our paper adds to the literature examining female underrepresentation in certain high-paying, prestigious, competitive, and stereotypically male sectors (e.g., Mas and Pallais (2017), Goldin (2014), Flory *et al.* (2014), Bertrand *et al.* (2010)). Secondly, our paper adds to the literature on matching in the labor market at the point of application (e.g., Castilla and Rho (2015), Gee (2018)). Our paper is very related to simultaneous work by Coffman *et al.* (2019), which also examines gender differences in application behavior (discussed in more depth in Section 1.5). Finally, our paper contributes to the literature on gender differences in confidence (e.g., Coffman (2014), Avilova and Goldin (2018)). Specifically within the tech sector, Murciano-Goroff (2018) finds that equally experienced female programmers are less likely than males to self-report knowledge of programming languages on their resume. Also closely related to our study, Del Carpio and Guadalupe (2018) find that a female's beliefs about her ability to perform well in a tech job affects her application decision.

Our results confirm the importance of language in the self-screening process for job applications: words matter in different ways for women and men of different educational backgrounds, which materially affect these job seekers' economic outcomes. Our paper also opens up a research agenda for more rigorous work on which words influence the decisions of women and men of differing skill levels. There are many relevant contexts for this: students reading the course description for a technically challenging class and then deciding whether or not to enroll; workers reading a performance review from their manager and then deciding whether or not to ask for a raise; college students reading a graduate school's admission requirements and then deciding whether or not to submit an application. More research on which specific words matter to economic outcomes is crucial to improving matching in the labor market.

The rest of this chapter proceeds as follows. Section 1.2 details the setting and experimental design. Section 1.3 describes the treatment. Section 1.4 models how altering listed requirements affects job seekers' application decisions. Section 1.5 discusses the main results. Section 1.6 concludes.

1.2 Sample & Design of Experiment

1.2.1 Sample

Company Background

Uber is a large tech company focused on ride-sharing services. The company employed approximately 13,000 individuals in their U.S. corporate positions as of June 2019. Corporate positions at Uber span all types of roles (e.g., Software Engineer, Data Scientist, Operations Manager) and as of December 2019, were correspondingly organized into specific groups: Advanced Technologies; Business & Sales; Communications; Design; Engineering; Finance & Accounting; Global Community Operations; Legal; Marketing; Operations & Launch; People Operations; Product; Public Policy; Safety & Security; and University.[6] The hiring process for external applicants – those who apply via the Uber careers website – begins with a review of the candidate's resume, a recruiter phone interview, and a hiring manager phone interview before receiving an on-site interview and then an offer decision.[7]

[6]Beginning in April 2019, Uber reorganized the teams listed on their website as part of their website redesign; the reorganization included new groups such as Business Development, Data Science & Analytics, People & Places, and Safety & Insurance.

[7]See Appendix Document A.1 for more detail on Uber's hiring process.

Job Postings Included in Experiment

Our experiment ran from August 2018 through December 2018 on Uber's careers website. We treated 616 of the 2,000+ job postings opened over this time period. We could not treat the full set of opened jobs since not all job postings had the capacity to be treated (i.e., not all job postings included inessential modifiers to qualifications or optional qualifications). Additionally, the rate at which new jobs were released per week exceeded the rate at which we could manually treat them, as implementing the treatment required individually altering language on a job posting-by-job posting basis. At any given point in time, 300 of the 1,000 U.S.-based corporate positions were included in the experiment. Job postings were only removed from the experiment when they expired, were edited by a recruiter, or when a job posting with a particular technical issue (hereafter referred to as a "long-digit-id" posting) was opened.[8]

The job postings often utilized similar language for similar positions. Postings came from 13 of Uber's specific teams: Advanced Technologies; Business & Sales; Design; Engineering; Finance & Accounting; Global Community Operations; Legal; Marketing; Operations & Launch; People Operations; Product; Public Policy; and Safety & Security.[9] Table 1.1 details relevant characteristics of the job postings in our sample.

Viewers Included in Experiment

Our experiment includes individuals who directly viewed a job posting on Uber's careers website (i.e., they opened a browser and directly navigated to the U.S. version of the careers website, "www.uber.com/careers/"). We exclude from our experiment those who applied via a third-party website (e.g., LinkedIn, Glassdoor, Indeed) due to technical implementation

[8]See Appendix Document A.2 for more discussion of job posting editing and technical issues related to long-digit-id job postings. Applications from eight of the original 624 job postings in our sample were not included due to the job posting being edited by a recruiter or hiring manager before the A/B test was initialized. Of the 616 job postings, 328 were initialized on the first day of the experiment, and the remaining 288 were added over the experiment duration.

[9]We did not include jobs from the Communications subgroup (due to lack of volume in job postings opened) or the University subgroup (due to the slightly different nature of internship-type positions).

constraints. We also exclude sourced and referral applications due to the fundamentally different application process for these candidates. Individuals who have ad-blocking browser extensions were also automatically excluded from the experiment. Due to the setup of our experiment, we do not have any data on the characteristics of third-party, sourced, or referral applications.

Table 1.1: *Share of Job Postings Included in Experiment, by Job Characteristics*

	Number of Jobs	Number of Apps
In Experiment	616	6,330
In Experimental Data	532	6,330
	Share of Jobs	Share of Apps
Technical Job	0.247	0.215
Entry-Level Job	0.380	0.594
Pipeline Job	0.252	0.408
Advanced Technologies Group	0.065	0.061
Business & Sales	0.070	0.135
Design	0.015	0.003
Engineering	0.187	0.153
Finance & Accounting	0.144	0.139
Global Community Operations	0.133	0.190
Legal	0.058	0.026
Marketing	0.028	0.021
Operations & Launch	0.117	0.135
People Operations	0.062	0.054
Product	0.084	0.064
Public Policy	0.002	0.0003
Safety & Security	0.036	0.018

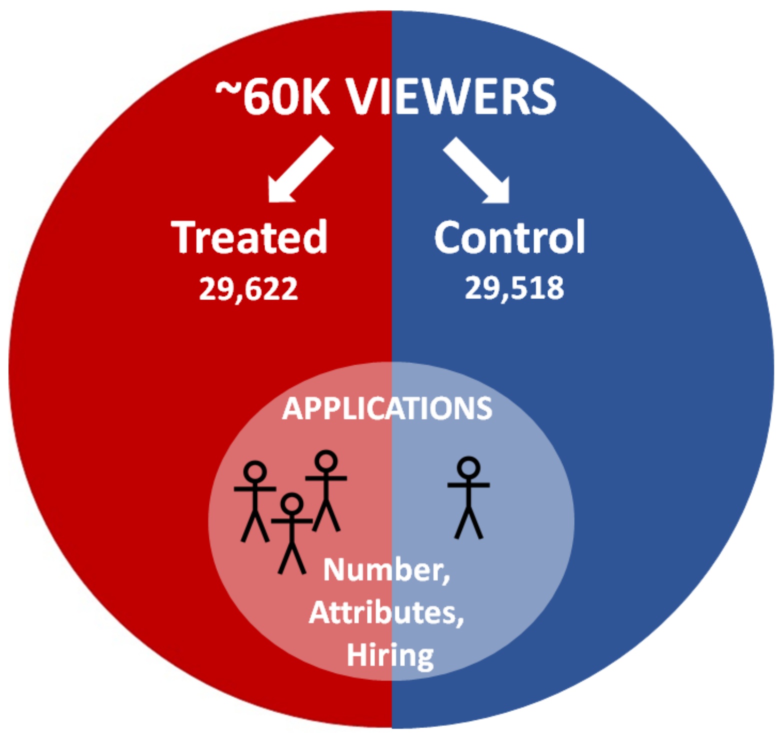
Notes: This table shows job posting characteristics as a share of all 616 job postings (Column 1) and 6,330 applications (Column 2). Of the 616 job postings, 532 had at least one application (i.e., 84 jobs had no applications to either the Treated or Control groups). Job posting characteristics include technical jobs (defined as having "Engineer","Data", or "Tech" in the job title), entry-level jobs (defined using Uber's internal classification system), pipeline jobs (defined as being in the experiment for longer than 3 months), and Uber-defined job group.

1.2.2 Design of Experiment

Randomization of Viewers

Viewers who visited Uber's careers website from August 13, 2018 through December 14, 2018 were randomized into the Treated or Control group. There were 203,232 viewers randomized at the device-browser level over this four-month period. Approximately 30 percent of these viewers (59,140) saw a job posting in our experiment, which is in line with the approximately 30 percent of U.S. job postings included in our experiment at any given point in time (Figure 1.1).

Figure 1.1: *Randomization Diagram*



Notes: This figure depicts the experiment execution. There were 203,232 viewers (as identified by browser-device) that were randomized into the Treated or Control group over the experiment duration. Of these viewers, 59,140 saw a job posting in the experiment (comprised of 29,622 in the Treated group and 29,518 in the Control group). Viewers saw the Treated or Control version for all job postings included in the experiment. For viewers who chose to apply, we were able to examine the gender, resume text, and progression through the hiring process.

Viewers randomized into the Control group saw the original (unedited) version for each of the 616 job postings included in our experiment (i.e., clicking on the job posting title from the list of jobs on the careers website would take them to the unedited version). Conversely, viewers in the Treated group saw the treated (edited) version for each of the 616 job postings included in our experiment. Through the use of cookies, the same treatment was served to a viewer who returned to the careers website at a later date.[10]

Regardless of randomization assignment, viewers saw the same list of jobs and search options on the main careers page (Appendix Figure A.1). Importantly, nothing about job search changed: searching by keyword (e.g. "PhD preferred") did not result in differential search results for Treated and Control viewers even if certain words were deleted in the Treated version of the job postings. The treatment was only present at the point of reading a job posting included in our experiment.

Fifty-fifty randomization of viewers into the Control and Treated groups ensures that any differences between the groups are solely due to the effects of the treatment. Put differently, if the treatment had no effect, we would expect an equal number of applications from the Treated and Control groups. We would also expect an equal distribution of skills both in general and by gender across Treated and Control applications.[11]

In Appendix Document A.2, we provide a discussion of potential slippage (e.g., discrepancies in treatment assignment and what was viewed in the job posting). To the extent that there is slippage, it is one-directional (only viewers assigned to the Treated group experienced these discrepancies if they instead saw the Control version of the job posting). Thus any slippage would at most cause us to find no differences between the two groups, and so is not generating our results. We document all technical details related to our experiment in Appendix Document A.3.

---

[10]This is conditional on the viewer using the same browser and device and not clearing local storage. See Appendix Document A.2 for the details of technical execution. See Appendix Table A.1 for device and browser detail.

[11]We are not able to performance balance tests traditionally performed in randomized controlled trials because we are not able to collect any information on viewers that did not apply.

Application Process for Viewers

Individuals had four methods of applying after clicking the "Apply Now" button on a job posting: (1) by logging into their LinkedIn profile (whereby data from the individual's LinkedIn profile would be transmitted into their Uber application); (2) by signing in using their existing account on the Uber careers website; (3) by uploading their resume; or (4) by manually filling out the application form (Appendix Figures A.2 and A.3).[12] All applicants were required to disclose their gender: "Male", "Female", or "Decline to Self Identify." Applicants, as identified by email address, were able to apply to more than one job posting in our experiment. We track the progression of applications at all stages of the hiring process. Recruiters and hiring managers were blind to the experiment throughout its duration.[13]

1.3 Treatment

Our treatment removed unnecessary language describing qualifications for the job postings in our sample. In implementing the treatment, our goal was to eliminate the inessential modifiers while preserving the core functions of the job. We executed the treatment manually, creating a treated version for each of the 616 job postings included in our experiment (i.e., manually removing optional criteria and softening language on a job posting-by-job posting basis). We describe this process below.

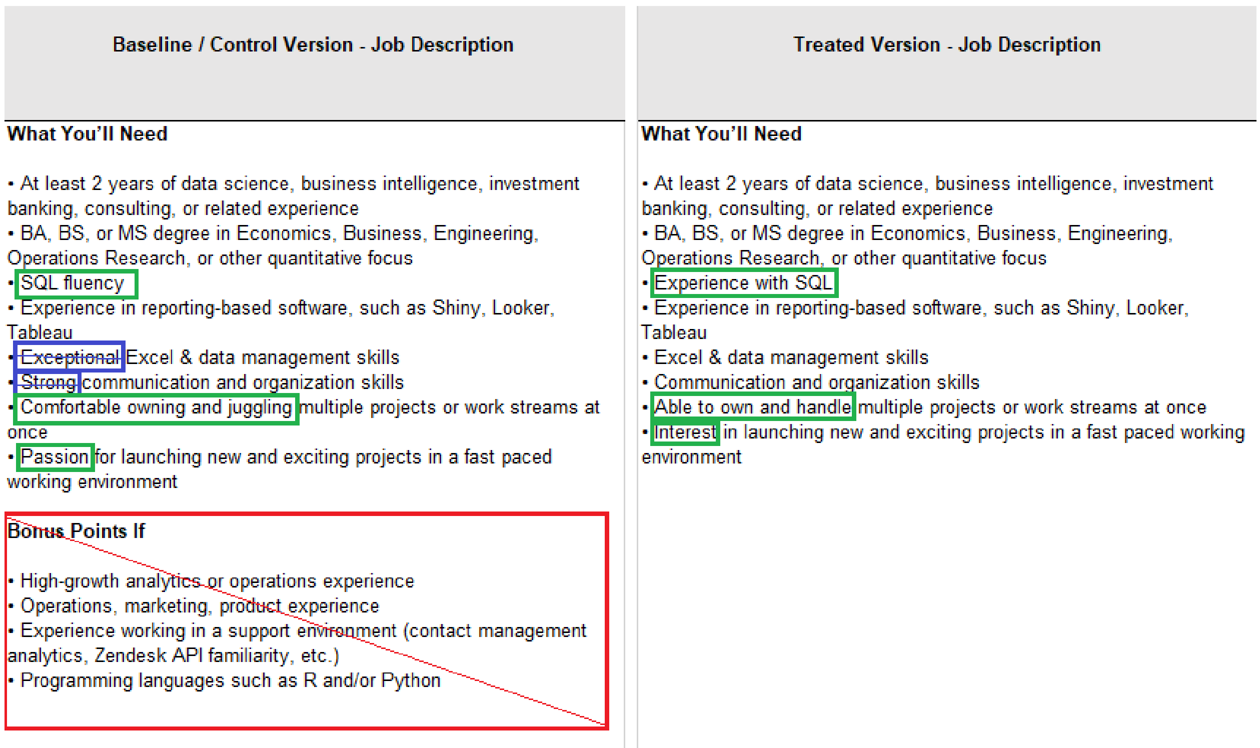
All Uber job postings in our experiment contained four sections: "About Uber," "About the Role," "What You'll Do," and "What You'll Need" (Appendix Figure A.4). We only altered language in the "What You'll Need" section, which lists the required and desired qualifications. This format, particularly the separation of required and desired qualifications, is common in job postings at other major tech firms (e.g., Google, Facebook, and Amazon). Some job postings contained language about listed qualifications outside of the "What You'll

[12]The LinkedIn application method is not the same as applying through a third-party website, since the candidate is still submitting the application from the Uber careers website.

[13]Recruiters and hiring managers could not view our treatment to the job postings from the internal system they use. There are approximately 475 recruiters at Uber, 107 of whom were responsible for a job posting in our experiment.

Need" section, which we could not change. However, the presence of intensely worded or optional qualifications outside of the "What You'll Need" section in treated job postings would only tend to dampen the effect of the treatment.

Figure 1.2: *Control vs. Treated Qualifications*



Baseline / Control Version - Job Description	Treated Version - Job Description
What You'll Need • At least 2 years of data science, business intelligence, investment banking, consulting, or related experience • BA, BS, or MS degree in Economics, Business, Engineering, Operations Research, or other quantitative focus • SQL fluency • Experience in reporting-based software, such as Shiny, Looker, Tableau • ~~Exceptional~~ Excel & data management skills • ~~Strong~~ communication and organization skills • Comfortable owning and juggling multiple projects or work streams at once • Passion for launching new and exciting projects in a fast paced working environment **Bonus Points If** • High-growth analytics or operations experience • Operations, marketing, product experience • Experience working in a support environment (contact management analytics, Zendesk API familiarity, etc.) • Programming languages such as R and/or Python	**What You'll Need** • At least 2 years of data science, business intelligence, investment banking, consulting, or related experience • BA, BS, or MS degree in Economics, Business, Engineering, Operations Research, or other quantitative focus • Experience with SQL • Experience in reporting-based software, such as Shiny, Looker, Tableau • Excel & data management skills • Communication and organization skills • Able to own and handle multiple projects or work streams at once • Interest in launching new and exciting projects in a fast paced working environment

Notes: This figure depicts the Control qualifications text (left) and the Treated qualifications text (right) for a job posting in our experiment. As depicted in the figure, we manually created the Treated qualifications by deleting optional qualifications (in red), deleting adjectives describing qualifications (in blue), and softening subjective language about qualifications (in green). We chose not to edit other qualifications that could potentially change the position itself (e.g., years of experience).

Our treatment consisted of three changes: deleting optional credentials (e.g., "PhD preferred"), deleting adjectives describing skills (e.g., deleting "deep" before "analytical expertise"), or softening credentials to soften the intensity of the qualification (e.g., changing "SQL fluency" to "experience with SQL") (see Figure 1.2 for an example). Table 1.2 summarizes the share of job postings that were affected by the different dimensions of our treatment. We made as many of the three changes as possible to each of our treated postings.

The Treated and Control version of the qualifications section for all 616 job postings were documented concurrently as the experiment was running in our AEA registry.

Table 1.2: *Summary of Treatment*

	Share of Job Postings (%)
Language Edited	97
Bonus Point Deleted	99
Education Bonus Point Deleted	31
Higher than Bachelor Bonus Point Deleted	22
Technical Skill Bonus Point Deleted	44
Certification Skill Bonus Point Deleted	15
Average Length Change (Control-Treated) in characters	268
Average Number of Bonus Points Deleted	3

Notes: This table shows the fraction of job postings in our experiment (out of the total 616 job postings) that were altered by different aspects of our treatment, as well as the average length change (in characters) in the Treated and Control qualifications text and average number of bonus points deleted.

Importantly, in executing the treatment, we wanted to delete the inessential modifiers while preserving the necessary functions of the position. In keeping with this aim, we did not alter years of experience or other hard skills that were not explicitly denoted as optional or preferred in the original job posting. We also chose to keep adjectives describing the company and culture, as we only wanted to change perceptions about the job qualifications.

1.4 Model: How Does Altering Language About Qualifications Affect Application Decisions?

We model how altering the language about qualifications affects job seekers' application decisions. The model frames three channels under which our treatment impacts our outcomes of interest: the number, gender composition, and skill distribution of applications. These three channels allow us to derive predictions for each of our outcomes under a given set of assumptions about applicant behavior. Our empirical results support the predictions of our model, allowing us to confirm our assumptions about how low- and high-skill job seekers of each gender respond to language changes about qualifications.

1.4.1 Model Setup

Consider a single job and a single employer. Job seekers vary in two dimensions: gender g (which can be female, f, or male, m) and skill s, where $s \sim N(\mu_s; \sigma_s^2)$. We assume job seekers and the employer can perfectly observe s.

Let R represent the requirements listed in the job posting. Let $V_g(s, R)$ denote the perceived value of the job, which represents the utility the job seeker derives from the position. Let $\gamma_g(s, R)$ denote the perceived probability (ranging from 0 to 1) of being hired for the job. We can further decompose $\gamma_g(s, R)$ into two subparts: $p_g(s, R)$, which represents the job seeker's perceived confidence that (s)he is qualified for the job, and $n_g(s, R)$, which represents the job seeker's perceptions about the relative competition from the rest of the applicant pool. Let c represent the fixed time cost of applying. Without loss of generality, assume $c \sim N(\mu_c; \sigma_c^2)$.

1.4.2 Effect of the Treatment on the Application Decision

Throughout the model, we assume that the job seeker believes the employer will choose to hire one candidate from the pool of qualified candidates who have applied for the position.[14] The job seeker must decide to apply or not to apply. A job seeker of gender g with skill level s will choose to apply if the expected utility from the job (the perceived probability of being hired times the value of the job) is greater than the fixed time cost of applying. This is mathematically represented as follows[15]:

[14] The results are analogous if we assume the employer can hire a finite number of candidates from the pool of qualified applicants. "Qualified" refers to having the minimum qualifications, and includes those who have more than the minimum.

[15] Changing the form of the equation to $(p_g(s, R) \cdot n_g(s, R))V_g(s, R) - c > 0$ does not meaningfully alter the model's predictions (see Appendix Document A.4).

$$\gamma_g(s,R)V_g(s,R) - c = (p_g(s,R) + n_g(s,R))V_g(s,R) - c > 0 \tag{1.1}$$

Our experiment decreased R by removing optional requirements and toning down the language of the listed requirements. To understand how our experiment affects the probability of applying, we take the partial derivative of c^* with respect to R, where c^* denotes the maximum c for which $(p_g(s,R) + n_g(s,R))V_g(s,R) - c = 0$, and so defines the threshold type for a given skill level s. The partial derivatives below determine how the application behavior of the threshold type c^* varies with different parameters.

$$\frac{\partial c^*}{\partial R}\Big|_s = \Big(\frac{\partial p_g(s,R)}{\partial R} + \frac{\partial n_g(s,R)}{\partial R}\Big)V_g(s,R) + \frac{\partial V_g(s,R)}{\partial R}\Big(p_g(s,R) + n_g(s,R)\Big) \tag{1.2}$$

The above equation highlights the three channels through which reducing R affects the probability of applying:

(1) Confidence Channel: $\frac{\partial p_g(s,R)}{\partial R} < 0$ for all s
Reducing R increases a job seeker's perceived confidence about being qualified for the job by making the job seeker more optimistic about his or her skill level relative to what the job requires.[16] This increases the job seeker's perceived probability of being hired and the probability of applying;

(2) Competition Channel: $\frac{\partial n_g(s,R)}{\partial R}$ ambiguous for all s
Reducing R changes a job seeker's perception of the competitiveness of the applicant pool. If the job seeker believes that a greater number of qualified individuals will apply, this would decrease the job seeker's perceived probability of being hired; even if the job seeker believes she is highly qualified for the role, a draw from a larger pool of qualified candidates would still lower her odds of being hired. Conversely, if the job seeker believes reducing R changes the skill distribution of the applicant pool such that a fewer number of qualified individuals apply, this would increase the job seeker's perceived probability of

[16] We ignore changes in the job seeker's perceived likelihood of being a good "fit" for the position (i.e., the job seeker believes she is more (or less) tailored for the role), because we assume the job seeker believes the employer will choose among the qualified candidates. If the employer instead hires based on fit for the position with penalty for being overqualified, there would be another term in the job seeker's application decision, $\frac{\partial f_g(s,R)}{\partial R}$; including this term does not materially impact the model's predictions (see Appendix Document A.4).

being hired, as the employer would have a smaller pool of qualified individuals to choose from thereby increasing her odds of being chosen. The net effect on the perceived probability of being hired and the probability of applying from this channel is ambiguous;

(3) Job Match Quality Channel: $\frac{\partial V_g(s,R)}{\partial R} < 0$ for low skill s and > 0 for high skill s
Reducing R increases a lower skill job seeker's perceived value of the job by increasing job match quality, and decreases a higher skill job seeker's perceived value of the job by reducing job match quality. The probability of applying increases for low skill job seekers and decreases for high skill job seekers as a result of this channel.

1.4.3 Two-by-Two-Case Predictions: Change to Number of Applications, Gender Composition & Skill Distribution

We can use assumptions to derive predictions about the number, gender composition, and skill distribution as a result of changing R. For simplicity, consider the 2x2 case: a high-skill job seeker of gender g and a low skill job seeker of gender g. The qualitative results are the same in the continuous case of skill s (see Appendix Document A.4). Figure 1.3 summarizes how each component of $\frac{\partial c^*}{\partial R}$ changes with a decrease in R under a given set of assumptions about applicant behavior.

Column 1 of Figure 1.3 reflects the assumption that $\frac{\partial p_g(s,R)}{\partial R} < 0$ regardless of gender or skill level, since lowering the listed requirements R would tend to increase a job seeker's confidence about being qualified, holding all else constant. We further assume that $\left|\frac{\partial p_f(s,R)}{\partial R}\right| > \left|\frac{\partial p_m(s,R)}{\partial R}\right|$, since females are assumed to be more sensitive to language about qualifications, and so are assumed to exhibit greater responsiveness along this margin.

Column 2 reflects the assumption that $\frac{\partial n_g(s,R)}{\partial R}$ is ambiguous for all job seekers regardless of gender or skill level. Lowering the listed requirements R could either *decrease* $n_g(s,R)$, since the job seeker may perceive that there are more qualified applicants vying for the same position, or *increase* $n_g(s,R)$, since the job seeker may now perceive that there are fewer qualified applicants that apply under a lower R.

Column 3 reflects the assumption that $\frac{\partial V_g(s,R)}{\partial R}$ is < 0 for low skill s and > 0 for high

Figure 1.3: *Model Predictions*

	$\frac{\partial p_g(s,R)}{\partial R}$	$\frac{\partial n_g(s,R)}{\partial R}$	$\frac{\partial V_g(s,R)}{\partial R}$	$\frac{\partial c^*}{\partial R}$	$\frac{\partial c^*}{\partial R}$ Assuming No Competition Channel
$s_{g,l}$	< 0	Ambiguous	< 0	Ambiguous	< 0
$s_{g,h}$	< 0	Ambiguous	> 0	Ambiguous	Ambiguous

Notes: This figure depicts the change in the probability of applying as a result of changing R for high ($s_{g,h}$) and low ($s_{g,l}$) skill job seekers of gender g. The first three columns depict the directional effect of changing R from each of the channels: the Confidence channel ($\frac{\partial p_g(s,R)}{\partial R}$), the Competition channel ($\frac{\partial n_g(s,R)}{\partial R}$), and the Job Match Quality channel ($\frac{\partial V_g(s,R)}{\partial R}$)). The net effect - $\frac{\partial c^*}{\partial R}$ - is presented in the last two columns, assuming non-negligible and then negligible effects of the Competition channel.

skill s. Reducing requirements for low skill s would tend to increase the perception that they are a better fit. Conversely, lowering R would tend to lower job match quality for high skill s.

Given these assumptions, Column 4 summarizes these effects to show the net effect of changes in R on the probability of applying for the marginal applicant of gender g and skill level s. If we assume a non-negligible *Competition* channel, then all of the predictions for our main outcomes are ambiguous.

If we assume a negligible effect of $\frac{\partial n_g(s,R)}{\partial R}$ on the probability of applying, we attain more concrete predictions of differences between the two skill types. Assuming a negligible effect of the *Competition* channel seems realistic since job seekers typically have no access to information about the applicant pool for a given job posting (both in terms of number of other applicants or their quality), which makes updating their beliefs harder and thus might lead them to ignore it entirely.[17] Under this scenario, lowering R increases the number of low skill s for both males and females since the *Confidence* channel and the *Job Match Quality* channel move in the same direction. Conversely, the number of high skill s that are induced to apply is ambiguous since the *Confidence* channel and the *Job Match Quality* channel move in opposite directions. For both high and low skill s, it is an empirical question if the magnitude of these changes alters the fraction of female applications for their skill type. The relative changes in high and low s job seekers also illustrate how the total skill distribution changes by gender.

We discuss how our predictions compare to the results of our experiment in Section 1.5. In particular, we can examine whether low skill job seekers of both genders are induced to apply as a result of the treatment, and whether female low skill job seekers are more responsive to the treatment than their male counterparts, in line with the predictions outlined here. We can also examine how the treatment affects high skill female and male job seekers; the magnitude of these effects will enable us to determine whether the *Confidence* channel or the *Job Match Quality* channel dominates for female and male high skill job seekers.

1.5 Empirical Strategy & Results

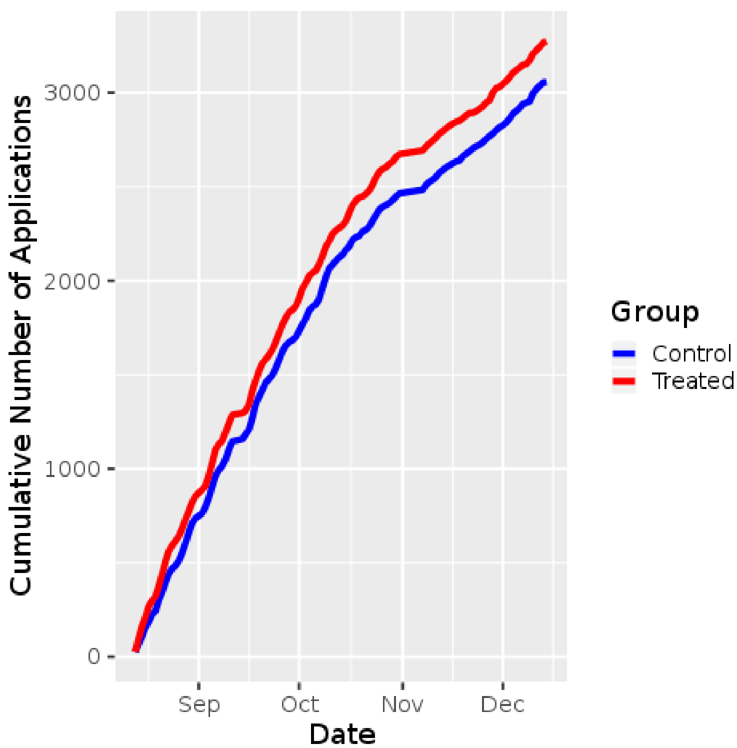
If the treatment had no effect, we would expect an equal number of applications from Treated and Control groups and an equal profile of skills across Treated and Control applications. Instead, we find a disproportionate number of Treated applications, suggesting the intervention encouraged individuals to apply. Further, we find that there is a different skill distribution by gender in the Treated group relative to the Control group.

In our experiment, approximately 8 percent of viewers who saw a job posting in our experiment ultimately applied (4,928 unique applicants out of 59,140 unique viewers). All of our analyses cluster at the individual applicant level, in line with assignment of treatment at

[17] The competition channel would also be negligible if we assumed that the employer is not restricted to hiring one applicant (or a finite number of applicants).

the individual level, and are robust to account for heteroskedasticity (Abadie *et al.* (2017)).[18]

Figure 1.4: *Number of Applications Over Time By Treatment Status*



Notes: This figure depicts the cumulative applications over time for the Control and Treated groups over the duration of our experiment (August 13, 2018 through December 14, 2018).

1.5.1 Descriptive Effect of the Treatment

Our treatment increased the total number of applications by 7.1 percent. The cumulative number of Treated applications is consistently higher than the cumulative number of Control applications (Figure 1.4). Summary statistics in Table 1.3 show relatively small differences between Control and Treated applications in terms of application method or gender composition.

[18]There are 46 applicants who had a Treated tag for one job they applied to in our experiment and a Control tag for another job they applied to in our experiment, which is possible if these individuals cleared their local

Table 1.3: *Summary Statistics by Treatment Status*

	Control	Treated	p-value
Female	0.375	0.368	0.565
Male	0.606	0.607	0.917
Missing Gender	0.020	0.026	0.128
LinkedIn	0.101	0.095	0.371
Resume	0.265	0.294	0.012
Manual	0.023	0.022	0.874
SignIn	0.610	0.589	0.088
Total Number of Applications	3,057	3,273	-
Average Number of Applications by Job	4.96	5.31	-

Notes: This table shows the fraction of Control and Treated applications by self-reported gender (Female, Male, Missing Gender) and application method. There are four methods for applying: "LinkedIn", "SignIn", "Apply with Resume", and "Apply Manually." Applying with LinkedIn imports the candidate's LinkedIn data into the application form, applying manually asks the candidate to fill out the application form directly, and applying with resume imports the data from the candidate's resume into the application form. Applying with the SignIn option pulls the latest existing information in the Uber system (entered from a previous application) and so could reflect any of the three application methods. The third-column shows the p-value from a two-sample t test between the Control and Treated mean values.

1.5.2 Effect of the Treatment on Total Number of Applications

To obtain the effect of changing language and removing optional qualifications on the number of total applications, we construct the ratio of the fraction of treated applications to the fraction of control applications and calculate the associated standard error of this ratio using the delta method (Table 1.4).[19] Column 1 of Table 1.4 shows that the number of applications significantly increases by 7.1 percent as a result of our treatment.[20] Column 2 of Table 1.4 repeats this analysis but restricts the sample to the first time that a unique

storage or used a different browser or device. We conservatively cluster at the individual applicant level.

[19]The reported p-values are associated with the null hypothesis that the ratio is 1.

[20]The point estimate from this method is equivalent to regressing the number of applications in a job-treatment cell on treatment status, controlling for job posting fixed-effects.

Table 1.4: *Effect of Treatment on Total Number of Applications*

	Ratio of Fraction Treated/Fraction Control		
	All Apps	1st Time Apps	2nd+ Apps
All	1.071*	1.085***	1.020
	(0.036)	(0.031)	(0.091)
p-value	0.051	0.006	0.825
Observations	6,330	4,928	1,402
Male	1.073	1.108***	0.964
	(0.047)	(0.041)	(0.107)
p-value	0.119	0.008	0.741
Observations	3,837	2,953	884
Female	1.051	1.039	1.095
	(0.058)	(0.048)	(0.171)
p-value	0.382	0.414	0.579
Observations	2,348	1,864	484
Missing Gender	1.377	1.313	1.615
	(0.289)	(0.253)	(0.713)
p-value	0.194	0.219	0.394
Observations	145	111	34
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of the treatment on the total number of applications for the whole sample and by gender. We obtain the effect of the treatment by calculating the ratio of the fraction of Treated applications to the fraction of Control applications (e.g., 1.071 in the first row and first column indicates a 7.1% increase in the number of applications between control and treatment for the whole sample). Robust standard errors are calculated using the delta method and are clustered by applicant. The reported p-values are associated with the null hypothesis that the ratio is 1. The first column represents all of the applications in our sample. The second column restricts the sample to the first time a unique individual (as identified by email address) applied to a job posting in our experiment. The third column restricts the sample to the second (or more) time(s) that a unique individual applied to a job posting in our experiment.

individual applied for a job in our experiment, and the third column restricts the sample to the second (or more) time(s) that a unique individual applied for a job in our experiment. The significant increase in the second column highlights that the treatment increases applications by drawing in new applicants.[21]

A higher number of applications in the Treated versus Control group is consistent with the *Confidence* channel increasing the likelihood of applying for all job seekers if they feel more confident they are qualified, the *Job Match Quality* channel increasing the likelihood of applying for lower skill job seekers, and the *Competition* channel increasing the likelihood of applying if job seekers feel that fewer qualified individuals will apply.

1.5.3 Effect of the Treatment on Gender Composition of Applications

The treatment has similar effects on the number of applications by gender: total applications from women increase by 5.1 percent, and total applications from males increase by 7.3 percent (Column 1 of second and third panel of Table 1.4). The treatment draws in new job seekers for both genders (Column 2).

Given that the treatment increases applications from both female and male job seekers, there is no effect of the treatment on the fraction of female applications. We show this formally through the following regression:[22]

$$\textit{Female Application}_i = \beta \textit{Treated}_i + \textit{Job FE}_j + \epsilon_i$$

where $\textit{Female Application}_i$ is a dummy variable equaling 1 if the individual associated with the application reported being female, $\textit{Treated}_i$ is a dummy variable equaling 1 if the application is associated with a Treated viewer, and $\textit{Job FE}_j$ represents job posting fixed-effects. β represents the percentage point change in the probability that a Treated application

[21]The treatment is insignificant in all Uber subgroups except for the Business & Sales subgroup, where it significantly increases the total number of applications by 26 percent; while a positive increase can be rationalized by our model, this result could also be due to chance given the number of subgroups we examine.

[22]The point estimates of this regression are equivalent to regressing the fraction of female applications in each job posting-treatment cell on an indicator for treated status, weighting by the number of applications in the given cell.

is female.

Table 1.5: *Effect of Treatment on Fraction Female Applications*

	Female
Treated	−0.004
	(0.016)
Job FE	X
Control Mean	0.382
Observations	6,185
R^2	0.172
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: This table shows the effect of the treatment on the fraction of female applications. The dependent variable is a dummy variable equaling 1 if the individual reported being a female, and 0 if male; the sample is restricted to those who report their gender. The independent variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. Robust standard errors are clustered by applicant. Given the approximately 6,000 applications in our sample, we can rule out a change in the fraction female of female applications greater than approximately 3.5 percentage points.

As shown in Table 1.5, the treatment did not significantly alter the gender composition of applications in our experiment. Given the number of total applications in our sample, we can rule out a change in the fraction of female applications greater than 3.5 percentage points.[23] There were no significant changes in the fraction female for the entry-level, non-entry level, or technical jobs. There were also no significant changes in the fraction female for any Uber-defined job groups, except for the Legal subgroup which had a significant 18 percentage point increase in the fraction of female applications; however, the sample size for the Legal group is quite small (160 applications), and so this could also occur by chance given the number of subgroups we examine (Appendix Table A.2).

[23]See Appendix Document A.5 for the associated power calculations.

1.5.4 Effect of the Treatment on Educational Attainment

While we do not see significant differences in the gender composition of applications, our analysis of the associated resumes reveals significant differences by gender in terms of skill as a result of the treatment. While women applying in the Control group are significantly more likely to have higher than a bachelor's degree relative to men applying for the same job, women in the treatment group are equally likely as men to have advanced degrees.

Table 1.6 displays the effect of the treatment on the educational attainment of applications by gender. We classify every application according to highest educational attainment: having lower than a bachelor's degree, having only a bachelor's degree, or having higher than a bachelor's degree (defined as having a Masters, JD, MBA, or PhD).[24] We then run the following regression for each of the three educational attainment categories:

$$\textit{Educ. Attainment}_i = \beta_1 \textit{Treated}_i + \beta_2 \textit{Female}_i + \beta_3 \textit{Treated} \cdot \textit{Female}_i + \textit{Job FE}_j + \epsilon_i$$

where $\textit{Educ. Attainment}_i$ is a dummy variable equaling 1 if the individual associated with the application reported having the given level of educational attainment, $\textit{Treated}_i$ is a dummy variable equaling 1 if the application is associated with a Treated viewer, $\textit{Female}_i$ is a dummy variable equaling 1 if the individual associated with the application reported being female, $\textit{Treated} \cdot \textit{Female}_i$ is a dummy variable for the interaction of the treated and female terms, and $\textit{Job FE}_j$ represents job posting fixed-effects. β_1 reflects the percentage point change in the likelihood that a Treated male has the given educational attainment relative to a Control male. β_2 reflects the percentage point change in the likelihood that a Control female has the given educational attainment relative to a Control male. β_3 reflects the incremental effect of being a Treated female on the likelihood of having the given educational attainment.

As shown in the third column, females in the Control group are 5.8 percentage points significantly more likely to have higher than a bachelor's degree relative to their Control male counterparts after controlling for job posting fixed-effects. This effect is attributable

[24]We conduct a robustness check that classifies every application's educational attainment using the education substring of the resume text, and our results are similar in magnitude and significance (Appendix Table A.3).

Table 1.6: *Effect of Treatment on Educational Attainment*

	Educational Attainment		
	< Bachelor	Bachelor Only	> Bachelor
Treated	−0.014	−0.006	0.019
	(0.014)	(0.020)	(0.020)
Female	−0.006	-0.053^{**}	0.058^{**}
	(0.016)	(0.024)	(0.023)
Treated X Female	0.023	0.035	-0.058^{*}
	(0.022)	(0.031)	(0.030)
Job FE	X	X	X
Control Male Mean	0.146	0.391	0.463
Observations	6,185	6,185	6,185
R^2	0.294	0.186	0.296
Note:			*p<0.1; **p<0.05; ***p<0.01

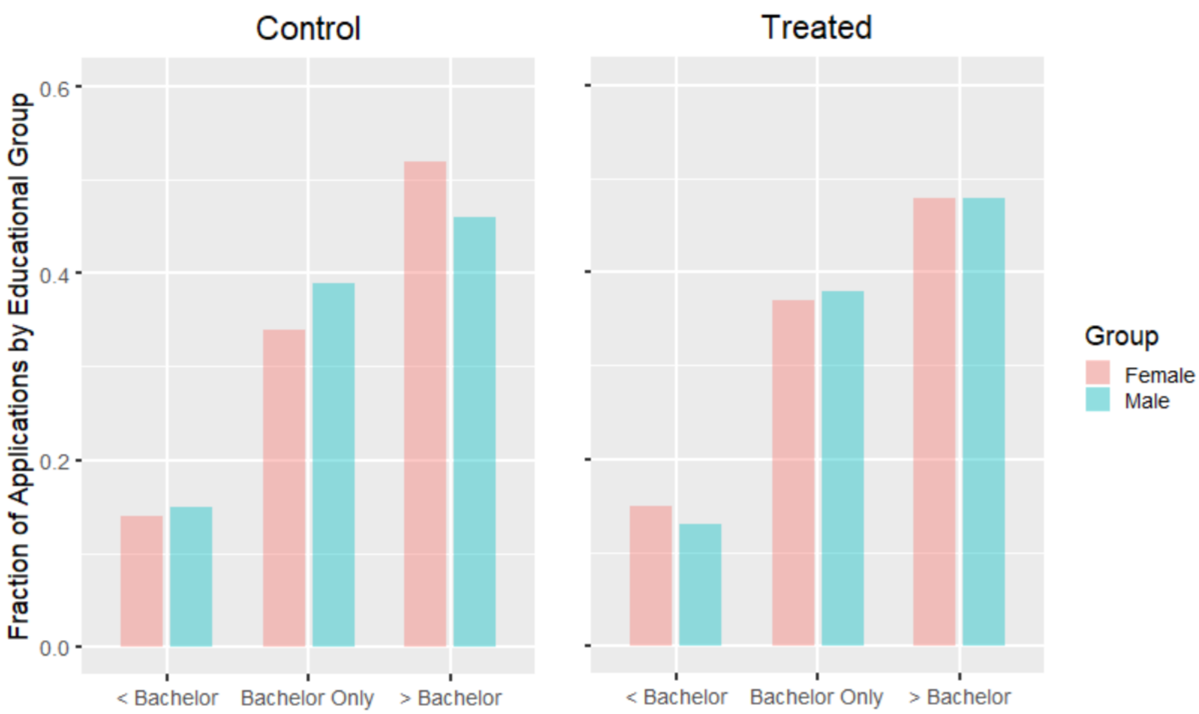
Notes: This table shows the effect of treatment and female status on educational attainment. The dependent variable in the first column is a dummy variable equaling 1 if the application is associated with less than a bachelor's degree. The dependent variable in the second column is a dummy variable equaling 1 if the application is associated with only a bachelor's degree. The dependent variable in the third column is a dummy variable equaling 1 if the application is associated with having higher than a bachelor's degree – defined as having a Master's, MBA, JD, or PhD. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

to those who list a Masters degree or MBA (Appendix Table A.4). This pattern is also stronger for technical, stereotypically male subgroups; Control females are respectively 15.7 percentage points, 25.3 percentage points, and 16.9 percentage points more likely to have higher than a bachelor's degree within technical jobs (as defined by the job posting title), jobs in the Advanced Technologies Group, and jobs in the Engineering group. This is in line with the hypothesis that females might be particularly sensitive to meeting listed qualifications in stereotypically male domains.

The treatment erases the educational attainment differences between women and men

(Figure 1.5). The coefficient on *Treated* · *Female* is negative 5.8 percentage points, which offsets the premium exhibited by females in the Control group. This effect is largely due to the treatment inducing more women with a bachelor's degree or lower to apply, as well as a slight outflow of women with higher than a bachelor's degree (Appendix Figure A.5).[25]

Figure 1.5: *Educational Distribution by Gender and Treatment Status*



Notes: This figure depicts the change to the educational distribution for males and females. The x-axis shows each of the three educational levels (Lower than Bachelor, Bachelor Only, and Higher than Bachelor). The y-axis shows the share with the given educational level after controlling for job fixed-effects. Control females are more likely to have an advanced degree relative to their male counterparts applying for the same job, but Treated females are equally likely as their male peers to have an advanced degree. This is because the treatment increases applications from females with a bachelor's degree or lower, and causes fewer females with higher than a bachelor's degree to apply, whereas the treatment increases applications from males independent of their educational attainment.

[25]We do not find significant differences in this effect of the treatment on entry versus non-entry level jobs. We do find that the effect is pronounced in the largest Uber job group, Global Community Operations (GCO), which could be due to the large sample size of the GCO subgroup (comprising approximately 20 percent of all applications).

The inflow of female applications from those with a bachelor degree or lower as a result of the treatment is in line with the *Confidence* and *Job Match Quality* channels drawing in lower skill female job seekers. Softening the qualifications makes these females more confident about being qualified for the job and/or increases the suitability of the job, thereby inducing them to apply. Conversely, the slight outflow of females with higher than a bachelor's degree as a result of the treatment accords with the *Job Match Quality* channel dominating the *Confidence* channel among higher skill female job seekers; sensitivity to match quality outweighs the increased confidence these females have about being qualified for the job. Importantly, we do not see similar patterns for men, as the treatment induced them to apply irrespective of educational attainment. Our results are consistent with high skill men having a weaker *Job Match quality* channel, perhaps due to having a lower sensitivity to matching all of the given qualifications.

Our treatment effects on high skill females versus males is analogous to simultaneous findings by Coffman *et al.* (2019) (CCK). Serving as the employer for an hour-long essay writing job on Upwork, CCK invite individuals to apply for either an "intermediate" ($70/hour) or "expert" ($150/hour) essay writing job. Of these invited individuals, those that were randomized into either of the two treatment conditions saw a single sentence inviting or encouraging job seekers with above a 3.75 score on Upwork's Management or Analytical skills test to apply for the "expert" version of the job. Since all the individuals in CCK's sample already took these skill tests, their treatment adds a precise, quantifiable qualification, and thus there is less room for self-confidence about having the qualification to influence application behavior. Despite differences in the setting and treatment, CCK find that highly qualified females (defined as having above a 3.75 test score) are more likely than highly qualified males to apply when the skills test score qualification is added. This evidence is in line with our findings that job match quality is a more important factor for high skill females than high skill males.

1.5.5 Effect of the Treatment on an Additional Application Attributes and a Composite Measure of Skill

Effect of Treatment on Educational Background/Performance, Technical Skills, Work Experience and Social Skills

We also examine the effect of the treatment on other application attributes: "eliteness" of educational institution, listed GPA, self-reported technical skills, social skills (communication, teamwork, managerial skills, and a measure of praiseworthy language used in one's resume), foreign language skills, and work experience (defined in more detail in Appendix Document A.6). We classify these attributes into *Objective* and *Subjective* categories to highlight the inherent challenges of identifying true skill from an applicant's resume. Objective measures include educational attainment, degree institution, GPA, and listing a GitHub account, as these measures seem more difficult to artificially report or inflate. Subjective measures include self-reported technical skills and work experience (due heterogeneity in experience and difficulties associated with assessing actual work experience from textual data). Social skills reflect words associated with the given attribute in the resume text; while likely not a perfect measure, these attributes do capture the extent to which an individual uses the associated words, and thus is perhaps more aware of the attribute's value.

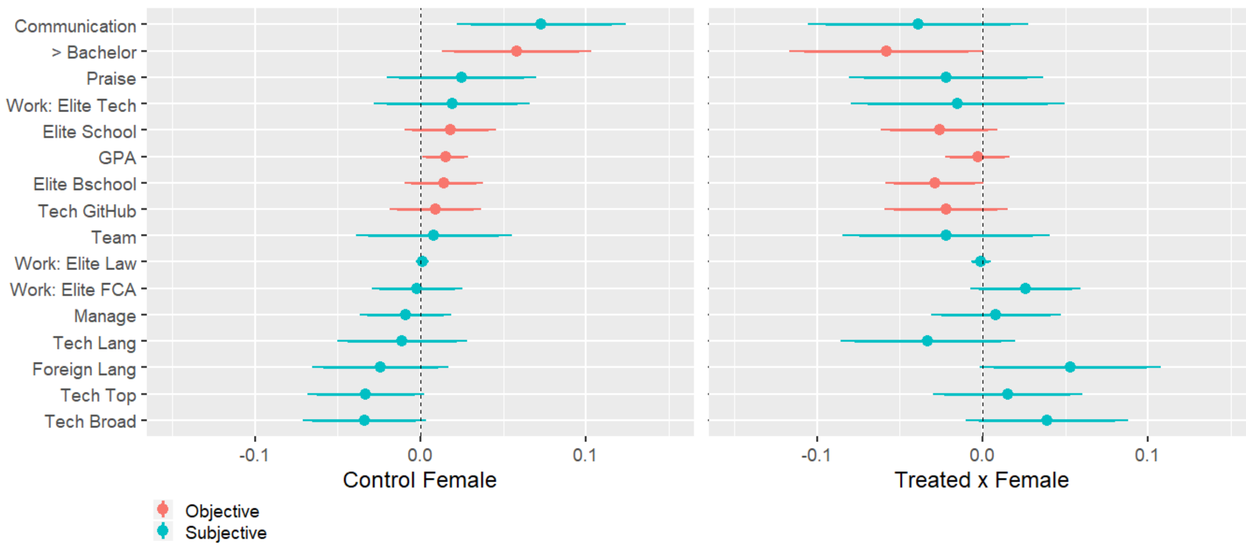
For each attribute, we perform the following regression:

$$Attribute_i = \beta_1 Treated_i + \beta_2 Female_i + \beta_3 Treated \cdot Female_i + Job\ FE_j + \epsilon_i$$

where $Attribute_i$ is a dummy variable for having the attribute (for all attributes except GPA, which is a continuous measure ranging from 0 to 1, e.g., 3.5/4), $Treated_i$ is a dummy variable equaling 1 if the application is associated with a Treated viewer, $Female_i$ is a dummy variable equaling 1 if the individual associated with the application reported being female, $Treated \cdot Female_i$ is a dummy variable for the interaction of the treated and female terms, and $Job\ FE_j$ represents job posting fixed-effects.[26]

[26]For GPA, we trim values greater than 1 and less than 0.3, and we calculate the average if two or more GPAs are listed in the same application.

Figure 1.6: *Distributional Changes for Application Attributes*



Notes: This figure depicts the changes to other application attributes for females as a result of our treatment. These estimates reflect individual regressions of the given attribute on a dummy variable for being Treated, being Female, and being Treated and Female, controlling for job fixed-effects and robust standard errors clustered by applicant. The left panel displays the coefficient on the Female variable, which represents the Control Female premium relative to Control Males (along with 90 and 95% confidence intervals). The right panel displays the coefficient on the Treated · Female variable, which represents the incremental effect of the treatment for females (along with 90 and 95% confidence intervals). > Bachelor refers to listing an advanced degree. Elite School refers to listing an elite educational institution, Bschool refers to listing an elite business school, and GPA refers to the fractional form of the listed GPA, ranging between 0 and 1. Tech Top refers to listing one of the most commonly listed technical skills in our sample's job postings, Tech Broad refers to listing one of a broad set of over 40 technical skills sourced from our sample's job postings, Tech Lang refers to listing technically inclined language sourced from our sample's job postings, and Tech GitHub refers to listing "github." Work: Elite Tech refers to listing an elite tech company, Work: Elite FCA refers to listing an elite finance, consulting, or accounting company, and Work: Elite Law refers to listing an elite law firm. Social skills (Communication, Team, and Manage) refer to listing words describing these attributes in the resume text, and Praise refers to listing proactive, self-promoting, and praising language in the resume text. Foreign Language refers to listing one of 33 foreign languages in the resume text. See Appendix Document A.6 and Appendix Tables A.5 through A.8 for more detail.

Figure 1.6 represents the results of these regressions (see also Appendix Tables A.5 through A.8 for each attribute's regression). The left panel plots the coefficient of the *Female* variable (along with 90 and 95 percent confidence intervals), and so displays the premium (deficit) that Control females have relative to Control males for the given attribute.

Besides having a higher than bachelor's degree, Control females list more words related to communication skills. All the *Objective* measures of skill also indicate a slight premium for Control females relative to Control males, though the point estimates are not estimated with precision. At the other end of the spectrum, Control females seem to self-report fewer technical skills than Control males, in line with recent work showing that females are more likely than men to under-report technical skills (Murciano-Goroff (2018)).

The right panel displays β_3, the coefficient on the *Treated* $\cdot$ *Female* variable, reflecting the incremental change to the skill premium (deficit) as a result of being a Treated female. The treatment seems to significantly "undo" the skill premium exhibited by Control females for having a higher than bachelor's degree and listing an elite business school, both by increasing applications from females in the lower part of the distribution and causing fewer higher skilled females to apply. The other *Objective* skills seem to exhibit the same pattern, though the point estimates are not estimated with precision. Conversely, our *Subjective* measures exhibit less of a consistent pattern. The effect of the treatment on male attributes is shown in Appendix Figure A.6; for males, the tendency to apply does not appear significantly correlated with any specific attributes; furthermore, to the extent that these estimates could be made more precise, the treatment seems to increase applications from more skilled males in terms of the *Objective* attributes, in direct contrast to the treatment effects on females.

A final aspect of skill concerns the propensity for applications to have the bonus qualifications that we deleted in the Treated version of the job postings. We do not find significant differences between Treated and Control applications for this outcome. However, this seems in part due to the limited sample size for our analysis, as it is not possible to code and match on all the listed bonus points in all job postings (Appendix Table A.9). This is particularly true for bonus points which are more sophisticated and described in more detail; for example, it is not possible to determine whether an applicant has "experience developing complex software systems scaling to millions of users" (a bonus point listed in several of our experimental job postings).

Effect of Treatment on Composite Skill Index

How do observable attributes actually affect the probability of success in the hiring process? To assess this question, we construct a complementary metric of skill which takes into account employer weights on the observable attributes and subjectivity in recruiters' decisions about who to advance in the hiring process. This subjectivity could encompass additional information that is too difficult to quantify (e.g., detailed information about a candidate's career progression).

We construct this skill measure by using information about how applications fare at the first stage of the hiring process – the recruiter phone interview – which only approximately 9 percent of applications in our sample progress to. We regress a dummy variable for receiving a recruiter phone interview on all aforementioned application attributes (educational attainment, educational institution eliteness, educational degree field, technical skills, work experience, social skills, and foreign language skills) to obtain the predicted likelihood of receiving a recruiter phone interview.[27]

$$\textit{Phone Interview}_i = \beta X + \textit{Job FE}_j + \epsilon_i$$

where *Phone Interview*$_i$ is an dummy variable for receiving a recruiter phone interview, X is a vector of application attributes (each attribute is included linearly), and *Job FE*$_j$ represents job posting fixed-effects.

For Control applications, we predict the probability of receiving an interview for applicant i using all the Control applications minus applicant i, in accordance with the standard leave-one-out procedure (Abadie *et al.* (2018)). For Treated applications, we use the coefficients from a regression with only the Control applications to predict the probability of a phone interview.[28] Higher values for the skill index correspond to a higher likelihood

[27]The sample for this regression is restricted to those who report their gender. We do not include GPA as an application attribute since it is only present for a fraction of the sample.

[28]Note that we remove all applications from applicant i in our implementation of the leave-one-out procedure. We also apply the leave-one-out procedure to Treated applicants that had a Control application for another job (46 applicants), instead of using all the Control applications in the prediction. Job fixed-effects are used to obtain the coefficients on the application attributes but not to calculate the prediction.

of receiving a recruiter phone interview.

Table 1.7: *Effect of Treatment on Composite Skill Index*

	Skill Index
Treated	0.358^{**}
	(0.179)
Female	0.518^{**}
	(0.220)
Treated X Female	-0.514^{*}
	(0.290)
Control Male Mean	5.75
Job FE	X
Observations	6,185
R^2	0.225
Note:	*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on composite skill. The dependent variable is a continuous measure of the predicted likelihood of receiving a recruiter phone interview (the first stage of the hiring process), constructed by regressing a dummy variable for reaching the recruiter phone interview on the application attributes (e.g., educational attainment, eliteness of educational institution, educational degree field, technical skills, work experience, social skills, and foreign language skills). See Appendix Document A.6 and Appendix Tables A.5 through A.8 for more detail about the application attributes. We multiply this measure of composite skill by 100 for presentational purposes. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

The effect of treatment and female status on this measure of composite skill is depicted in Table 1.7 (note that we multiply the skill index by 100). The magnitude and sign of the coefficients parallel our broad findings on educational attainment and our *Objective* application attributes: Control females seem more skilled than their Control male counterparts conditional on the job, as seen in the positive coefficient on *Female* in Table 1.7, and the treatment counteracts this premium for females, as seen in the negative coefficient on $Treated \cdot Female$. Interestingly, Treated males appear to be more skilled relative to Control males, as seen in the positive coefficient on *Treated*; this could be the case if lower skill males are overconfident relative to higher skill males, in-line with theory from psychology – the Dunning-Kruger effect – which finds an inverse correlation between confidence and skill at lower parts of the skill distribution (Dunning and Kruger (1999)).

Broadly, these findings can be rationalized by our model: female applicants in the Control group are more skilled than their male counterparts, and the treatment causes relatively equal increases in the number of females and the number of males (due to the net effects of the *Confidence* channel, *Competition* channel, and *Job Match Quality* channel). If there are no additional high skill female job seekers to bring in or if the *Job Match Quality* channel for higher skill female job seekers dominates the *Confidence* channel, then the marginal female under the treatment will be of lower skill.

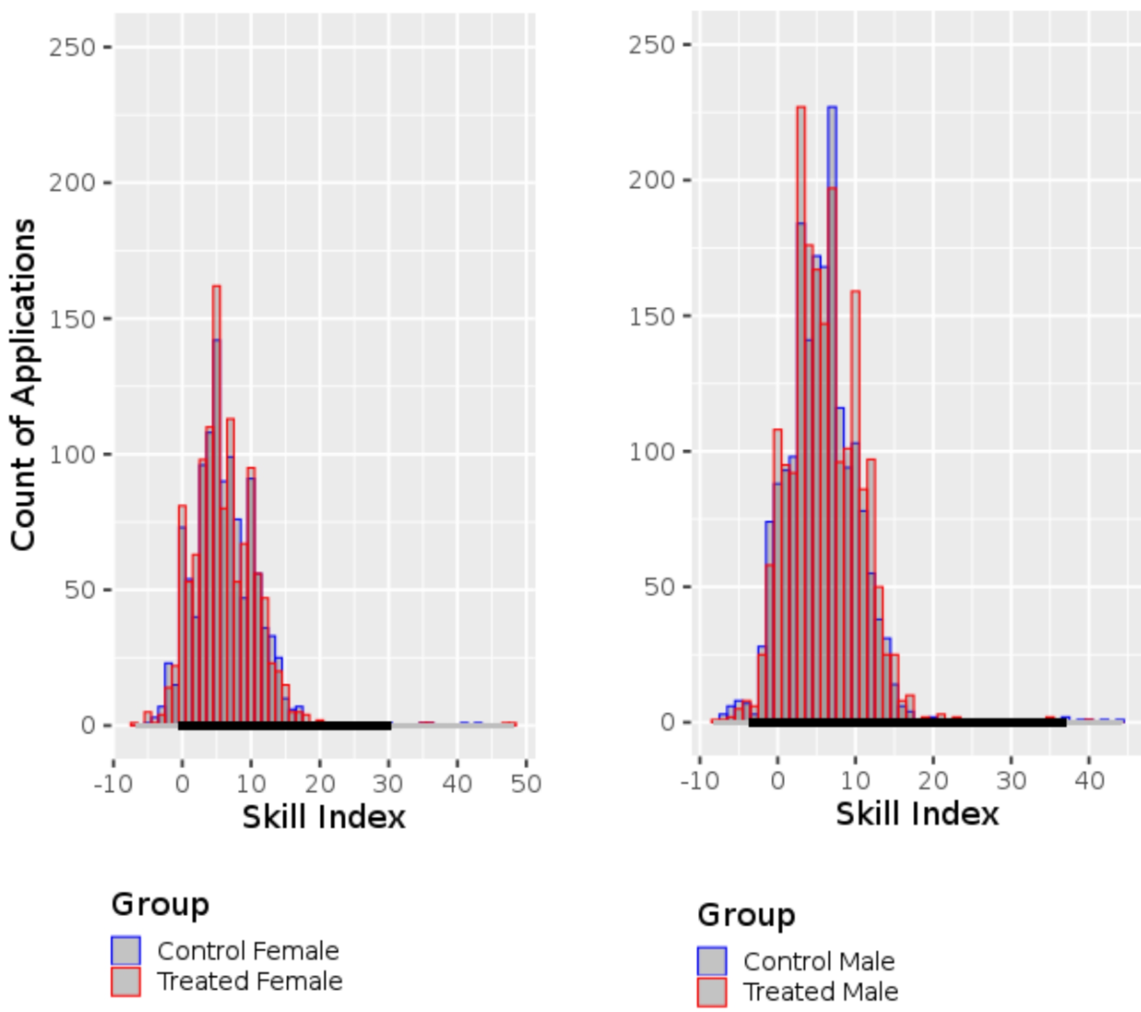
Do Treated Applications Accrue in the Interviewing Range?

An important related question asks whether the treatment brings in job seekers in the skill range in which Uber actually conducts recruiter phone interviews. We plot the raw count of applications for both Treated and Control groups by gender for all values of the skill index in Figure 1.7. The black line represents the skill index range in which Uber actually offered a recruiter phone interview for Control applications in our sample. It is clear that the treatment brings in applications in the interviewing range for both females and males. The wideness of the interviewing range indicates that there is noise and/or subjectivity in the recruiter decision-making process, which results in a range of applications receiving a first-round interview beyond what would be predicted by looking at a narrow set of observable application attributes.

1.5.6 Effect of the Treatment on Progression through Hiring Process

Table 1.8 shows the effect of the treatment and female status on the probability of reaching each of the four stages in Uber's hiring process: recruiter phone interview, team phone interview, team interview, and offer. We do not find consistent patterns by treatment or

Figure 1.7: *Skill Index: Total Application Counts for Females and Males*



Notes: This figure depicts the total number of applications for each value of our composite skill index for females (left panel) and males (right panel). The y-axis reflects the count of total applications. The x-axis reflects a composite measure of skill constructed by regressing a dummy variable for reaching the recruiter phone interview (the first stage of the hiring process) on the application attributes (e.g., educational attainment, eliteness of educational institution, educational degree field, technical skills, work experience, social skills, and foreign language skills). Blue bars represent Control applications and red bars represent Treated applications. The black line overlaying the x-axis reflects the full recruiter phone interviewing range for Control applications (i.e., the skill range in which Uber conducted recruiter phone interviews for Control applications in our sample).

gender.[29] To obtain these estimates, we run the following regression:

$$\textit{Reach Hiring Stage}_i = \beta_1 \textit{Treated}_i + \beta_2 \textit{Female}_i + \beta_3 \textit{Treated} \cdot \textit{Female}_i + \textit{Job FE}_j + \epsilon_i$$

where $\textit{Reach Hiring Stage}_i$ is an dummy variable for reaching the given stage, $\textit{Treated}_i$ is a dummy variable equaling 1 if the application is associated with a Treated viewer, $\textit{Female}_i$ is a dummy variable equaling 1 if the application is associated with a female application, $\textit{Treated} \cdot \textit{Female}_i$ is a dummy variable for the interaction of the treated and female terms, and $\textit{Job FE}_j$ represents job posting fixed-effects.

Even though none of the point estimates from these regressions is significant, the approximately 6,000 applications in our RCT enable us to bound the null effects at each stage with some degree of precision (though still a relatively large share of the Control male mean). For example, we can bound the increase in the probability that a Treated female (relative to a Control female) reaches the recruiter phone interview stage to less than 3.3 percentage points (approximately 40 percent of the Control female mean). We can also bound the increase in the probability that a Treated male (relative to a Control male) reaches the recruiter phone interview stage to less than 2.6 percentage points (approximately 30 percent of the Control male mean). See Appendix Document A.5 for more detail on the bounding procedure.

1.6 Conclusion

Given that females are underrepresented in certain occupations, it is important to understand whether differential sensitivity to language about qualifications could be a contributing factor. We find that job seekers are sensitive to the language in listed qualifications in a sample of 60,000 viewers to over 600 job postings for Uber's U.S. corporate positions. Our

[29] There are a small fraction of applications (107, or less than 2 percent) that were "moved" into a different job posting by recruiters or hiring managers if the recruiter or hiring manager believes the candidate is a better fit for a different role; we can track these applications for however long they remain in our experimental sample for the job posting they initially applied to. Approximately 65 applications, or 1 percent of our sample, self-withdrew at some point during the hiring process; we record their progression for however long they remain in the sample before withdrawing.

treatment significantly increased the number of applications in our experiment by 7 percent. Because both males and females are induced to apply, there was no significant change in the fraction of female applications. There are, however, significant gender differences in who decides to apply as a result of the treatment. Control females are significantly more likely than their male counterparts to have high educational qualifications. Our treatment increases applications from women with a bachelor's degree or lower and slightly reduces applications from women with a higher than bachelor's degree, whereas men are induced to apply independent of educational attainment; as a result, Treated females are equally likely as Treated males to have an advanced degree. These patterns for females appear consistent for other *Objective* application attributes and a measure of composite skill.

Table 1.8: *Effect of Treatment on Progression through Hiring Process*

	Recruiter Phone Interview	Team Phone Interview	Interview	Offer
Treated	−0.0003	−0.002	−0.001	0.0003
	(0.009)	(0.006)	(0.005)	(0.003)
Female	−0.005	−0.003	−0.003	0.003
	(0.011)	(0.007)	(0.006)	(0.004)
Treated X Female	0.006	−0.003	−0.0001	−0.001
	(0.015)	(0.009)	(0.008)	(0.005)
Job FE	X	X	X	X
Control Male Mean	0.090	0.033	0.025	0.007
Observations	6,185	6,185	6,185	6,185
R^2	0.175	0.148	0.175	0.097
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on the probability of reaching the subsequent hiring stages. The dependent variable in each column represents a dummy variable for receiving an interview at the given stage. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

If we extrapolate our findings to the broader population, evidence that females hold themselves back from applying for jobs relative to their educationally equivalent male counterparts has significant implications for female representation and advancement in the labor market. It would be useful to understand what other language in job postings besides the qualifications section women might be more sensitive to than men, such as descriptions of the job responsibilities or the culture of the team.

Another fruitful direction of future work would be to better understand which *specific* words are likely to influence perceptions about the work required in the job and suitability for the position, versus which specific words are likely to influence perceptions about the odds of getting the job. Understanding the way that language influences both channels could suggest interventions that would retain the high-skill female applicants while pulling in more women with a bachelor's degree or less.

Finally, our evidence that language influences the economic outcomes of decision-makers differently by gender and skill invites a plethora of related research questions in a range of settings. In school, for example, students routinely receive written feedback about their work and signals about their performance which affect their self-assessment; if women are more sensitive than men to the language used to convey this feedback, this could impact choices about which classes to take and which fields of study to pursue. Analogously, in a professional working environment, women may be more sensitive to the language used in a performance review, and may correspondingly make different decisions about putting themselves forward for a promotion or asking for a challenging work assignment. More research is needed to understand how differential responses to language in a broad range of settings impact economic outcomes.

Chapter 2

Do Women Underrate their Performance?: Evidence from HR Data

2.1 Introduction

When workers perform a task, they are evaluated on their performance. In many firms, these formally written evaluations are a central component of tracking a worker's productivity and progress. For managers, they are an opportunity to collect feedback from their subordinates and justify compensation, promotion, and firing decisions. For workers, performance appraisals provide managerial feedback, and in the context of a self-review, enable workers to strategically project an opinion about their performance to their boss. The information contained in these reviews – how workers are evaluated and how they evaluate themselves – can potentially illuminate reasons for existing gender, racial, and other demographic disparities in the labor market.

Despite the richness of performance evaluations, there is relatively little research using this type of data. This is due to the highly confidential and sensitive nature of performance evaluations, which make them unavailable for research. It is also difficult to obtain both objective and subjective reviews of performance for a given worker, making it hard to disentangle a worker's actual performance from their perceived performance.

I provide new insights into how women and men behave in the performance evaluation process using a unique dataset of approximately 20,000 performance evaluations from a large employee performance management software company.[1] Critically, these data contain workers' self-reviews as well as reviews from their manager, peers, and direct reports. Unlike prior studies which focus on a limited number of companies, my sample reflects evaluations from 170 companies across a range of industries, representing reviews for over 15,000 unique workers from nearly 5,000 unique managers.

I use these data to ask whether women provide different self-assessments of their performance versus men, relative to their own managers' review. This metric evaluates whether men (or women) overrate or underrate their performance in comparison to their manager's beliefs. I find that women are more pessimistic than males about their comparative performance by approximately a tenth of a standard deviation. I find the same pattern for specific skills (e.g., *communication & collaboration*, *management & leadership*, *technical skills*), and no significant differences between female versus male-typed skills, which is suggestive that the gap is present regardless of true competency. Interestingly, I find significantly larger gender gaps for workers with lower tenure, who have had less time to develop their reputation. This could be explained by managers defaulting to their priors about females and rating female workers with low tenure higher than their true performance. It could also be explained by female workers with less on-the-job experience underreporting their performance relative to their male counterparts.

To the extent that gender gaps in self-evaluation reflect the latter, then my findings have significant implications for other settings where women are faced with the decision to self-promote. These decisions occur at many points in the labor market, such as how a worker describes her skills in a job interview, or whether she feels comfortable claiming

[1] The main analyses reflect approximately 20,000 managerial evaluations, but robustness checks, which include self, peer, and direct report reviews, reflect approximately 100,000 evaluations.

credit for work done in a team setting. If managers react endogenously to males' and females' outward displays of self-promotion, then gender differences in worker behavior could lead managers to treat men and women in different ways, which could affect their career trajectories.

This paper contributes to several strands of literature. First, my paper relates to work that empirically connects performance evaluation data to labor market outcomes. Cappelli and Conyon (2018) use data from a large U.S. company to show that performance appraisals are positively related to employee outcomes such as compensation, promotions, and worker exit decisions. Frederiksen *et al.* (2017) obtain performance evaluation data from six large companies and also find a positive correlation between managerial evaluations and compensation and promotions. Benson *et al.* (2020) use microdata of sales workers to provide evidence that firms base promotion decisions more on current worker performance versus other predictors of managerial ability.

My focus on gender disparities in performance evaluations also relates to a rich literature on gender and labor market outcomes. In addition to occupation (Goldin (2014)) and employer-specific factors (e.g., Card *et al.* (2015)), evidence of gender differences in preferences for flexibility (e.g., Mas and Pallais (2017), Bolotnyy and Emanuel (2019)), confidence (e.g., Avilova and Goldin (2018), Coffman (2014)), competition (e.g., Niederle and Vesterlund (2007)), negotiation (e.g., Leibbrandt and List (2015)), acceptance of job tasks (e.g., Babcock *et al.* (2017b)), and job application behavior (e.g., Abraham and Stein (2020), Coffman *et al.* (2019)) highlight how females behave differently from males in ways that might meaningfully impact their careers. My paper adds to this literature by documenting gender differences at the moment of a worker's performance evaluation.

Finally, my paper relates to several economics papers and a rich social psychology literature on gender differences in self-promotion.[2] Exley and Kessler (2019) find persistent gender differences in self-promotion in an experiment with 900 MTurk workers. Interestingly,

[2]See Rudman and Phelan (2008), Bowles *et al.* (2007), and Rudman and Glick (2001) for examples in the psychology literature.

the authors state that this gender gap cannot be explained by confidence, incentives to self-promote, or information about average self-promotion, but hypothesize that females' perception of backlash (internalized from females' experiences outside of their study) might be driving the observed effects (even though gender is not revealed to employers in the context of their experiment). This paper provides evidence in the field consistent with Exley & Kessler's laboratory findings, and in a context where backlash concerns might be present because gender is known to employers. Relatedly, Babcock *et al.* (2017a) examine whether levying penalties for not agreeing to do a low-promotability task magnifies gender differences in task allocation; while they replicate females' receipt and acceptance of more low-promotability tasks than men, they do not find that penalties exacerbate these gender gaps (though they note that the penalties in their setting might be too small to matter).

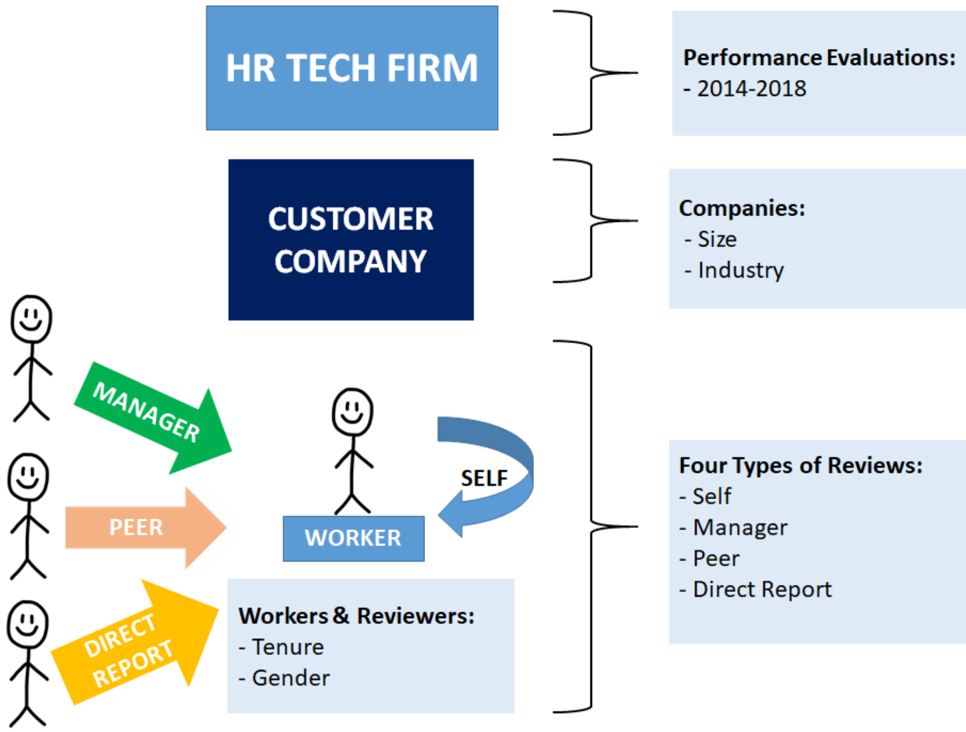
The rest of this chapter proceeds as follows. Section 2.2 details the sample. Section 2.3 describes the empirical strategy, main analysis, and potential mechanisms. Section 2.4 concludes.

2.2 Performance Evaluation Sample

My performance evaluation data come from an employee performance management software company. This company's primary product is a software platform used by its customer companies to manage the performance evaluation process. Through the use of this platform, the company stores performance management data for a broad array of companies (see Figure 2.1).

A given row of data represents a performance review for a worker from herself (i.e., a self-evaluation) or from a manager, peer, or direct report. The review can consist of textual and/or quantitative measures. The quantitative measures reflect numeric scores in response to specific questions, which are then standardized and averaged to yield a 0 through 1 score for each "competency" category (e.g., 0.6 for "communication"); competency scores are then averaged to form a single 0 through 1 reviewer score (see Figure 2.2). Each review can be tailored by the customer companies in terms of who reviews whom, the

Figure 2.1: *Employee Performance Management Software Company Data Structure*

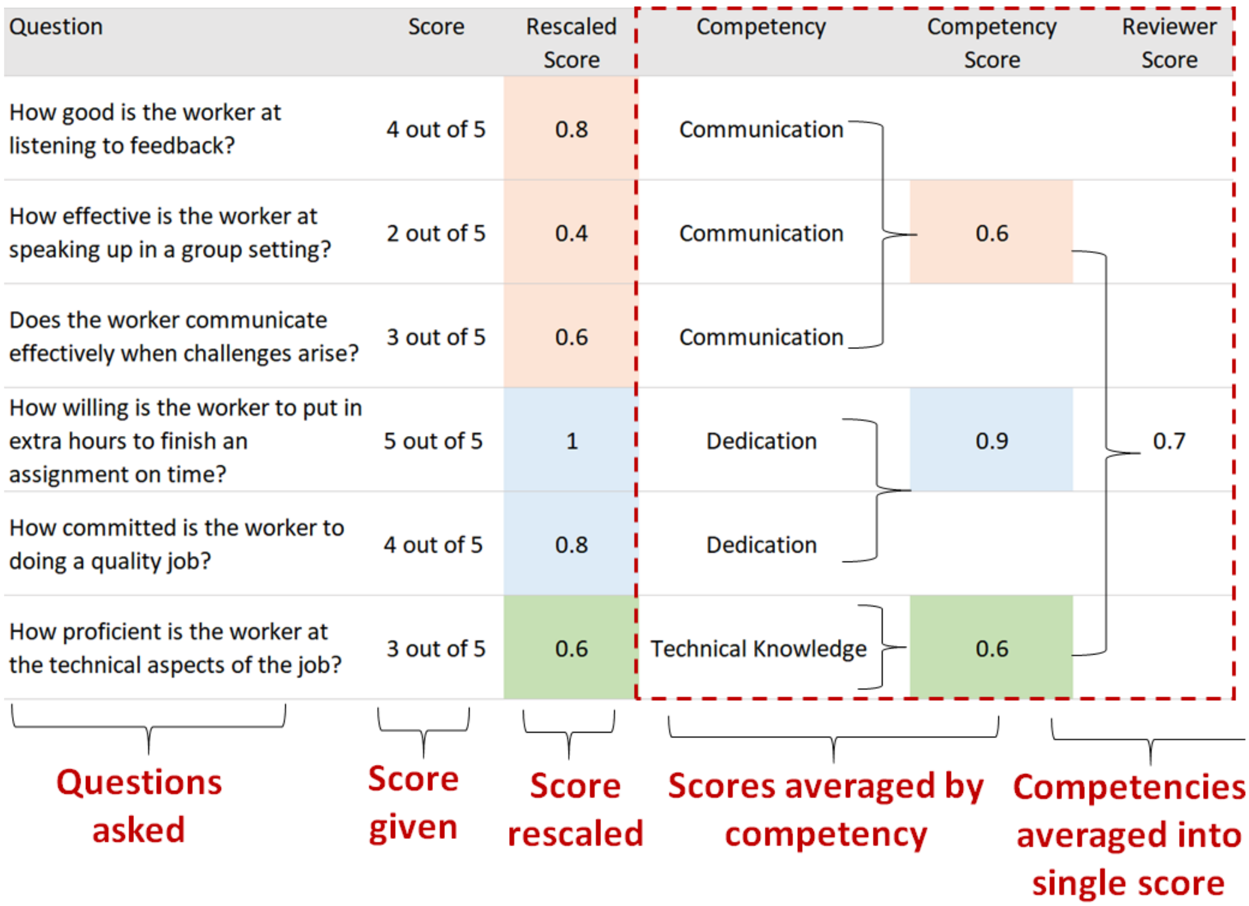


Notes: This figure displays the data in my sample. The employee performance management software company's platform stores customer company performance evaluations. For each customer company, there are four potential types of reviews: self, manager, peer, and direct report. I also have information on company size, industry, and worker/reviewer gender and tenure at the company.

questions and competency categories, and what is ultimately shared with the workers (see Appendix Figure B.1). While I do not have timestamp data for when a review is started and completed, discussions with the employee performance management software company suggest that self, peer, and direct report reviews are submitted before a manager's review. Ninety percent of the dataset reflects reviews where a manager could see all other reviews (self, direct report, and peer) for the worker before writing his or her own review.

I have access to a subset of variables for each performance review. In particular, I have information on the type of evaluation (e.g., self, manager, etc.), a numerical standardized score for the evaluation, numerical standardized scores for each competency category, labels

Figure 2.2: *Creation of Reviewer Score*



Notes: This figure displays the creation of the reviewer score. Scores for each question are rescaled, averaged within a competency category, and then averaged across competency categories to yield a single 0 through 1 value.

for each competency category, the date the evaluation was given, worker tenure and reviewer tenure at the company at the point of evaluation, worker termination date and reviewer termination date, and a company identifier.[3] Information on gender is determined from worker and reviewer names using standard gender-inferring techniques (see Appendix Document B.1 for more detail). I retain individuals in the dataset who have at least one self

[3] I am not able to distinguish between terminations that represent firings versus voluntary departures.

evaluation and a manager evaluation in the sample (these evaluations do not have to be in the same review cycle).[4]

Table 2.1: *Summary Statistics*

	N	St. Dev.
Companies	170	–
Worker (Self) Reviews	24,372	–
Unique Workers	15,709	–
Manager Reviews	23,446	–
Unique Managers	4,993	–
	Mean	St. Dev.
Share Female Workers	0.38	–
Share Female Managers	0.33	–
Worker Tenure (yrs)	2.50	3.29
Manager Tenure (yrs)	3.01	3.78
Competencies	2.52	2.28
Questions	5.95	5.64
Reviewer Score	0.70	0.16
Reviews per Company	281.28	671.91
Reviews Given by a Manager	4.70	5.04
Reviews Given by Self	1.55	1.06

Notes: This table displays the summary statistics for the final sample. The top panel displays the number of companies, workers, and managers. The bottom panel displays key statistics on a reviews-level basis: the female share, average tenure (in years), average number of competencies, average number of questions, average reviewer score, average number of reviews per company (inclusive of self and manager reviews), average number of reviews given by a manager, and average number of self reviews given by the worker.

Table 2.1 presents summary statistics. The final sample represents 170 companies over the 2014 through 2018 time period.[5] These companies span a range of industries, with slightly more concentration in the Computer Software and Advertising & Marketing sectors as a share of reviews (Appendix Table B.1).[6] Companies also exhibit a range of sizes, with

[4]See Appendix Document B.1 for additional detail on the data cleaning procedure.

[5]Most observations come from the 2016-2018 time period; see Appendix Figure B.2.

[6]Industry information is available for approximately 73 percent of the sample.

most falling in the 100 to 500 employee-range as a share of reviews (Appendix Figure B.3).[7] There are over 24,000 self-reviews (representing over 15,000 unique workers) and over 23,000 manager-reviews (given by approximately 5,000 unique managers). Thirty-eight percent of the self-reviews belong to female workers, and 33 percent of manager-reviews belong to female managers. Average worker tenure is 2.5 years and average manager tenure is approximately 3 years. The mean reviewer score is 0.7 (out of 1) (the full distribution is shown in Figure 2.3).[8]

2.3 Empirical Strategy & Results

I find that female workers are more likely than male workers to give themselves lower scores relative to their own manager's review. In the analysis that follows, all dependent variables are converted to z-scores for comparability (i.e., each dependent variable is standardized to have a mean of 0 and standard deviation of 1). Robust standard errors are conservatively clustered at the company level.

2.3.1 Effect of Worker Gender on Performance Evaluations

Gender Differences in Aggregate and at Worker-level

I first examine the effect of worker gender on self and manager scores in aggregate. I estimate this using the following equation:

$$ReviewerScore_{r,i} = \alpha + \beta Female_i + \gamma Tenure_i + \delta_c + \varepsilon_{ic} \tag{2.1}$$

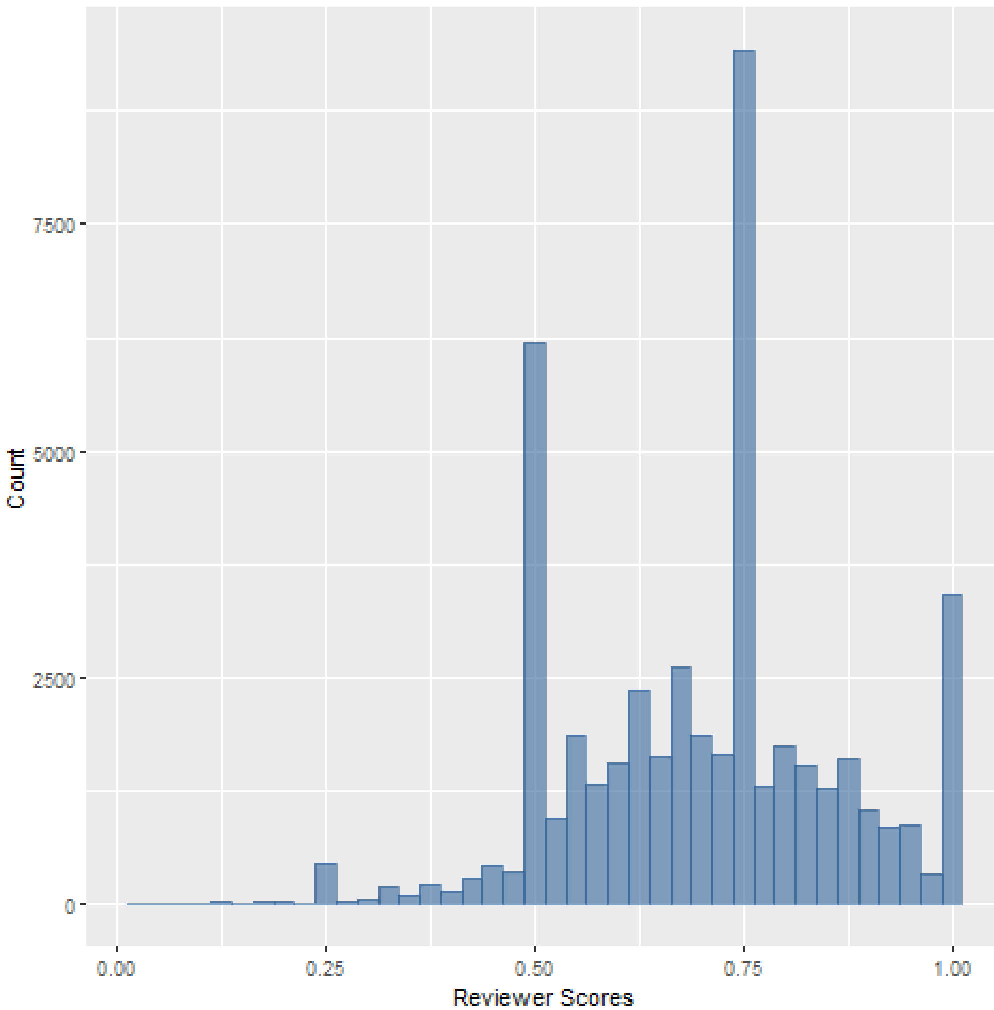
In this equation, $ReviewerScore_{r,i}$ represents the numerical score given by reviewer r to individual worker i, $Female_i$ represents a dummy variable for the worker's gender, and $Tenure_i$ represents a continuous variable for the worker's tenure (in years) at the point of

---

[7]Company size information represents a count of all the employees in the company, even if not all of them are part of the employee performance management software company dataset.

[8]Seventy percent of the sample reflects workers who are reviewed once per year.

Figure 2.3: *Reviewer Score Distribution*



Notes: This figure displays the raw distribution of the reviewer score in the full sample, and includes both self and manager scores.

evaluation. The coefficient of interest is β. I also include a vector of fixed effects for company, δ_c (or reviewer).

Table 2.2 presents the results of this analysis. Column 1 of Table 2.2 shows that female workers rate themselves by 0.033 standard deviations lower than male workers. Conversely, female workers are rated 0.030 standard deviations higher than their male counterparts by their managers (Column 2). The coefficient on *Female* is significantly different between these two regressions ($\chi^2_{(1,N=47,818)} = 7.8, p = .005$). The magnitude of the *Female* coefficient is slightly larger once I control for differences across managers (Column 3). These patterns suggest that there are differences in how female relative to male workers both rate themselves and are rated by their managers.

Table 2.2: *Effect of Worker Gender on Self and Manager Score*

	Self	Mng	Mng
Female Worker	−0.033	0.030	0.051*
	(0.028)	(0.021)	(0.029)
Company FE	X	X	
Reviewer FE			X
Observations	24,372	23,446	23,446
Adjusted R^2	0.120	0.187	0.330
Note:	*p<0.1; **p<0.05; ***p<0.01		

Notes: This table displays a regression of the Reviewer z-score listed in the column on a female dummy variable and a continuous measure of worker tenure. Robust standard errors are clustered at the company level.

To more precisely control for differences in worker productivity, I construct a "gap" measure for each worker by subtracting the manager's score of the worker from the worker's self review, i.e., a *Self-Manager* score. I construct this score in two ways: (1) by taking the worker's average self score across all review cycles minus the manager's score in a particular cycle; and (2) by taking the worker's self score minus the manager's score, where both scores are from the same review cycle ("within cycle"). I then estimate the following equation:

$$(Self - Manager)_{r,i} = \alpha + \beta Female_i + \gamma Tenure_i + \delta_r + \varepsilon_{ir} \tag{2.2}$$

As in equation (2.1), $Female_i$ represents a dummy variable for the worker's gender, $Tenure_i$ represents a continuous variable for the worker's tenure (in years) at the point of review, and δ_r represents manager (or company) fixed effects.[9]

Table 2.3: *Effect of Worker Gender on Self Minus Manager Score*

	Self-Manager Score			
	(1)	(2)	(3)	(4)
Female Worker	-0.071^{***}	-0.077^{***}	-0.105^{***}	-0.112^{***}
	(0.023)	(0.021)	(0.022)	(0.021)
Company FE	X	X		
Manager FE			X	X
Within Cycle		X		X
Observations	23,446	20,929	23,446	20,929
Adjusted R^2	0.075	0.070	0.198	0.176
Note:		*p<0.1; **p<0.05; ***p<0.01		

Notes: This table displays regressions of the Self-Manager z-score listed in the column on a female dummy variable and a continuous measure of worker tenure. The Self-Manager z-score in columns 1 and 3 reflects the average self score across all review cycles for a given worker minus the manager's score for the worker in a given review cycle. The Self-Manager z-score in columns 2 and 4 reflects the self score minus manager score for a given worker where both scores are in the same review cycle. Robust standard errors are clustered at the company level.

Table 2.3 presents the results of this analysis. The preferred specification in Column 4 uses the "within cycle" gap metric and takes into account manager fixed effects, thereby controlling for variation across reviewers using a potentially more precise metric of worker and manager differences because both scores come from the same point in time. The coefficient on *Female* from this regression shows that women rate themselves approximately a tenth of a standard deviation lower than their male counterparts after accounting for a proxy of externally assessed ability (the manager score). While I cannot quantitatively

[9] An alternative equation regresses the *Self* score on a dummy variable for the worker's gender, a continuous variable for the worker's tenure, and a continuous variable for the *Manager* score. Analogously, one could regress the *Manager* score on a dummy variable for the worker's gender, a continuous variable for the worker's tenure, and a continuous variable for the *Self* score. Both equations are equivalent to my main specification, except they additionally account for the *level* of worker performance (as proxied by either the *Self* score or *Manager* score). Including the level of worker performance as a control variable does not materially change the results.

attribute this gap to female workers rating themselves lower conditional on ability – rather than managers rating female workers higher conditional on ability – the observed patterns are in line with concurrent research showing that females are less likely to self-promote (Exley and Kessler (2019)).

Gender Differences by Skill & Industry

Does female underreporting of performance (conditional on externally assessed ability) occur for skills which females are generally thought to be better at than men, such as *communication*? I examine heterogeneous effects by skill by constructing the same *Self-Manager* score for five specific skills: *communication & collaboration*, *management & leadership*, *technical skills*, *work ethic*, and *proactive attitude*. These skills represent common words in the competency categories and those that are of particular interest to assess by gender.[10]

Table 2.4: *Skill Heterogeneity: Effect of Worker Gender on Self Minus Manager Score*

	Self-Manager Score				
	Collaborate	Manage	Tech	Work Ethic	Proactive
Female Worker	-0.111^{**}	-0.120^{*}	-0.194^{***}	-0.185^{**}	-0.143^{***}
	(0.046)	(0.067)	(0.068)	(0.073)	(0.053)
Manager FE	X	X	X	X	X
Within Cycle	X	X	X	X	X
Observations	4,413	1,856	759	2,377	1,914
Adjusted R^2	0.164	0.135	0.170	0.152	0.136
Note:					*p<0.1; **p<0.05; ***p<0.01

Notes: This table displays regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each skill group: communication & collaboration, management & leadership, technical skills, work ethic, and proactive attitude. Robust standard errors are clustered at the company level.

Table 2.4 displays the results of this analysis. Interestingly, I see the same patterns in

[10] Both the self score and manager score are associated with the given skill. See Appendix Document B.2 for more detail on the competency words included in each skill group.

Table 2.3 across all of the skills (though significance is slightly lower for the *management & leadership* subgroup). The *Female* coefficient exhibits a slightly higher magnitude for the *technical skills* subgroup; while not statistically significantly different from the other *Female* skill coefficients, this pattern is in line with evidence that females exhibit less confidence in stereotypically male domains (Coffman (2014)), as technical skills are often thought to be more male-specific.

I also examine heterogeneous effects by industry. Figure 2.4 plots the *Female* coefficient for each industry from a regression of the *Self-Manager* score on gender (see Appendix Table B.2).[11] The *Female* coefficient is negative for nearly all of the industries. There do not seem to be significant observable correlates with industries that have a more negative *Female* coefficient; under a model of managerial discrimination, we might have expected industries with a higher share of female workers to exhibit a smaller negative *Female* coefficient (because there is less scope for managers to discriminate along gender lines), but this pattern does not appear in the data.

Taken together, the lack of striking differences between skill groups or industries suggests that a common, external factor might be driving the observed gender differences, as opposed to differing beliefs about gender-specific skills or industry-specific factors.

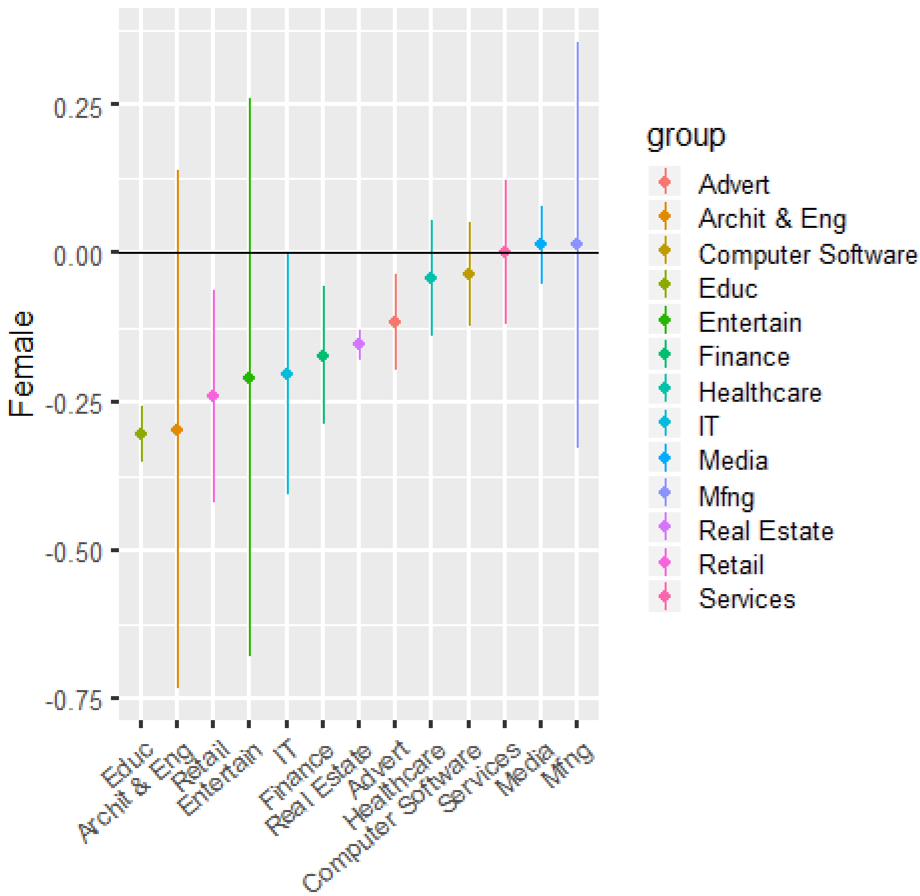
2.3.2 Mechanisms for Female Underreporting of Performance

Heterogeneity in Worker Tenure

Why would women underrate their performance relative to men? One hypothesis is that women, more so than men, anticipate backlash if they rate themselves higher than what they expect their manager will rate them. If women are worried about appearing overconfident or arrogant to their boss, then they might shade down their self-evaluation relative to men. If this hypothesis is true, one might expect gender differences in self-evaluation to be larger

[11] Industry data is available for approximately 73 percent of the sample. I omit two industries from this analysis (*Consumer Goods* and *Telecom / Communication Services*) because of too few observations in the industry group.

Figure 2.4: *Industry Heterogeneity: Effect of Worker Gender on Self Minus Manager Score*



Notes: This figure displays the point estimates and 95 percent confidence interval for the Female coefficient from regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each industry group: Advertising & Marketing, Architecture & Engineering, Computer Software, Education, Entertainment, Finance, Healthcare, IT Services, Manufacturing, Media, Real Estate, Retail, and Services. I omit two industries from this analysis (Consumer Goods and Telecom / Communication Services) because of too few observations in the industry group. All regressions include manager fixed effects, and robust standard errors are clustered at the company level.

for those with lower tenure, since these workers have less information about what their manager thinks of them and less information about how their manager might react if they appear arrogant. Alternatively, we might see gender differences in the *Self-Manager* score by tenure if managers learn about worker ability over time, and so rate a female worker at the beginning of her tenure higher than a comparable male worker based on managerial priors about female workers being more competent.

I examine gender differences in the *Self-Manager* score for workers of differing tenure in Table 2.5 and Figure 2.5 and find patterns consistent with these potential mechanisms. The magnitude of the *Female* coefficient is largest in magnitude for workers with tenure lower than 6 months. The $<$ *6 months Female* coefficient is significantly different from the *Female* coefficient for workers with between 2 and 3 years of experience ($\chi^2_{(1,N=6{,}466)} = 6.3, p = .012$) and 3 or more years of experience ($\chi^2_{(1,N=8{,}179)} = 7.3, p = .007$) (the two far right bars in Figure 2.5).[12]

Table 2.5: *Worker Tenure: Effect of Worker Gender on Self Minus Manager Score*

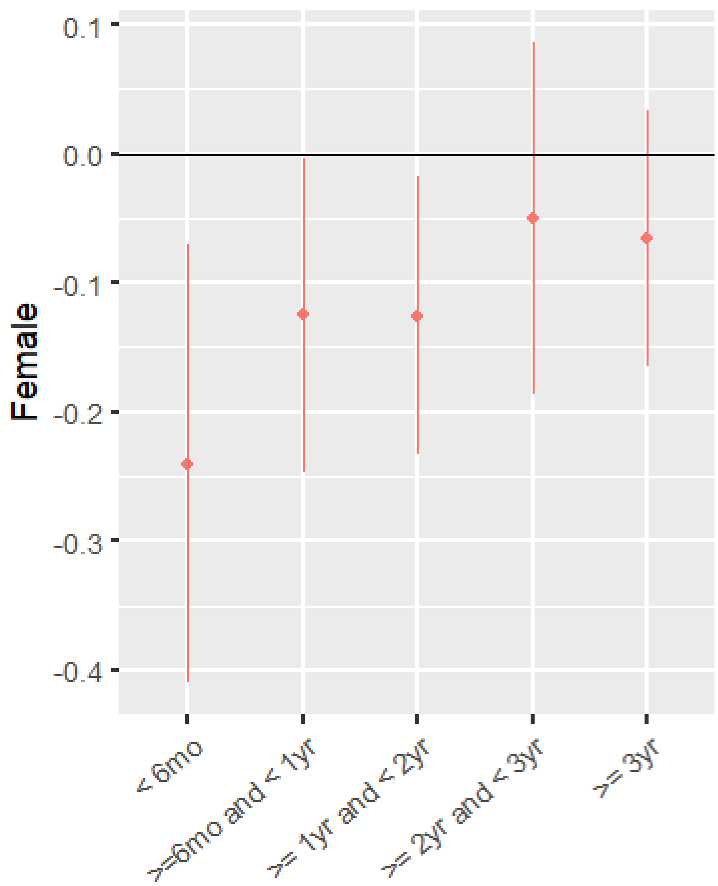
	Self-Manager Score				
	<6 mo	6 mo to <1 yr	1 to <2 yr	2 to <3 yr	3+ yr
Female Worker	-0.240^{***}	-0.125^{**}	-0.125^{**}	-0.051	-0.066
	(0.087)	(0.062)	(0.055)	(0.069)	(0.051)
Manager FE	X	X	X	X	X
Within Cycle	X	X	X	X	X
Observations	3,135	3,998	5,421	3,331	5,044
Adjusted R^2	0.197	0.166	0.203	0.200	0.244
Note:					*p<0.1; **p<0.05; ***p<0.01

Notes: This table displays regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each tenure group: <6 months, 6 months to <1 year, 1 to <2 years, 2 to <3 years, and 3 or more years. Robust standard errors are clustered at the company level.

We can also examine whether learning over time is reflected in the variance of the scores. The standard deviation of the raw *Self-Manager* score declines from 0.176 for workers with $<$ *6 months* tenure to 0.154 for workers with 3 or more years of tenure, suggestive that some degree of learning over time about worker ability could be occurring on the part of

[12]This pattern is not present when I include company fixed effects instead of reviewer fixed effects, indicating that variation across reviewers within a company is an important aspect of heterogeneity in the *Self-Manager* score.

Figure 2.5: *Worker Tenure: Effect of Worker Gender on Self Minus Manager Score*



Notes: This figure displays the point estimates and 95 percent confidence interval for the Female coefficient from regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each tenure group (e.g., 6 months or less, etc). All regressions include manager fixed effects, and robust standard errors are clustered at the company level.

managers and/or workers. Examining the two scores separately, the standard deviation of the *Self* score is largest for workers with less than 6 months on the job (0.171), declines to 0.164 for workers with between 6 months and 1 year of experience, and then drops further to 0.161 for workers with 1 to 2 years of experience before leveling off. Conversely, the standard deviation of the *Manager* score remains the same for workers in the first three tenure groups (approximately 0.163), and only starts to decline for workers with between

2 and 3 years of experience. These patterns are consistent with workers and managers updating their beliefs about worker ability, though the updating appears more immediate for workers relative to managers.

Reviewer Relationship

Do females shade down their reported performance relative to men in the context of reviews from other colleagues, such as peers and direct reports? If the same gender gaps in self-evaluation exist for other worker-reviewer relationships, then this might suggest that female workers are generally underconfident about their abilities relative to males.[13]

I examine this by constructing analogous *Self-Peer* and *Self-DirectReport* measures, and performing the same analysis as in Table 2.3. The results in Table 2.6 show that the *Female* coefficient is both smaller and magnitude and insignificant in these samples when reviewer fixed effects are included.[14] This suggests that the worker-manager relationship, in particular, is driving part of the gender gap observed in the *Self-Manager* score.

Alternative Explanations: Other Reviewer Characteristics

Other reviewer characteristics do not appear to drive the main findings. Appendix Table B.3 separates the sample into managers with varying lengths of tenure, and shows no differences across the subsamples.

Appendix Table B.4 examines the *Self-Manager* score gap separately for female and male managers. While the *Female* coefficient is significantly larger under female managers than male managers ($\chi^2_{(1,N=20,929)} = 4.7, p = .029$), both are significantly negative. More

[13]In the cross-section, peers rate female workers 0.04 standard deviations significantly higher (p=.010) after controlling for reviewer fixed effects; direct reports do not rate females significantly higher, but this estimate may be partly constrained by sample considerations (since most direct reports only review one supervisor). For both subsamples, female workers do not rate themselves significantly lower than males after controlling for company fixed effects.

[14]I include reviewer fixed effects in the *Self-Peer* and *Self-DirectReport* regressions for comparability with our baseline specification. On average, 4 reviews are given by each peer. A smaller number of reviews, 1.4 on average, are given by each direct report, and so the *Female* coefficient is estimated off a smaller sample of direct reports who give two or more reviews to workers with different genders, making the estimate more imprecise.

Table 2.6: *Worker-Reviewer Relationship: Effect of Worker Gender on Self Minus Reviewer Score*

	Self-Mng	Self-Peer	Self-DirectReport
Female Worker	-0.112^{***}	-0.040	-0.082
	(0.021)	(0.033)	(0.154)
Reviewer FE	X	X	X
Within Cycle	X	X	X
Observations	20,929	43,506	12,336
Adjusted R^2	0.176	0.228	0.391
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table displays regressions of the Self-Reviewer z-score on a female dummy variable and a continuous measure of worker tenure for each reviewer group: managers, peers, and direct reports. Robust standard errors are clustered at the company level.

research is needed to understand why the gap is larger under female managers.

I also examine whether having more information about the worker from a peer or direct report evaluation in the same cycle affects the *Self-Manager* gender gap. Appendix Table B.5 shows no significant differences for reviews with and without a peer or direct report.[15] This suggests that the results are not being driven by managers altering their score of the worker after seeing additional data.

2.4 Conclusion

Performance evaluations are an important aspect of labor markets, as they impact decisions related to worker productivity, compensation, promotion, and long-term career choices. Using data from an employee performance management software company, I find that the gap between a workers' numerical self-assessment of their overall work performance and their managers' assessment of them is more negative for female workers than for male

[15] I restrict this analysis to the sample of reviews where managers are able to concurrently view peer or direct report evaluations.

workers, indicating that females rate themselves lower than their male counterparts after accounting for their manager's beliefs.

I find no significant gender differences in terms of particular skills (e.g., communication). This suggests the mechanism for observed effects might be more tied to external factors as opposed to gender differences in beliefs about specific competencies. I find that gender differences in the *Self-Manager* score gap is only present between workers and managers, as there are no significant differences in other worker-reviewer relationships (i.e., peers and direct reports) once I control for reviewer fixed effects. Moreover, the gender gap is significantly larger for workers with lower tenure.

The gap I identify could be due to managers rating female workers higher than male workers, or female workers rating themselves lower than male workers. If the mechanism is more attributable to the latter, this would be in-line with existing research about female underconfidence and aversion to self-promote relative to males. Moreover, if gender differences at the point of worker evaluations are correlated with gender differences in how women and men publicly present themselves, this can have meaningful implications for workers' labor market trajectories, such as the type and amount of work they are given and other opportunities to advance their careers. More work is needed to better understand drivers of the gender score gap I identify in this sample, the ways that women might engage in lower self-promotion outside of the performance review process, and how managers correspondingly respond to gender differences in workers' self-assessments and self-promotion.

Chapter 3

The Gender Earnings Gap: New Evidence from Online Self-Reports[1]

3.1 Introduction

Over fifty years after the Equal Pay Act of 1963, inequities in wages paid to men and women remain a highly relevant topic in the United States. Much research has focused on the reasons for the gender pay gap. Existing evidence points to a smaller difference in earnings once observable factors, such as age, education, and occupation, are accounted for. But a female-to-male earnings ratio of approximately 80 cents on the dollar still persists in measures of publicly available data.

Using a unique dataset, we further probe the extent of the gender wage gap. Our data come from Glassdoor.com, a company that crowd-sources salary data, and provides potentially richer occupation information using detailed occupation and employer-level identifiers relative to publicly available data. Moreover, Glassdoor does not top-code the higher end of the earnings distribution, unlike publicly available data, and so could provide a better picture of earnings at the upper end of the income distribution.

Our analyses provide evidence of a gender wage gap on the order of 6 to 8 percent

[1]Co-authored with Matthew Gibson (Williams College)

once detailed occupational and employer information are accounted for, smaller than in prior studies. Similar gaps exist for the upper end of the income distribution. Given the selected nature of the data, we construct a detailed mapping of Glassdoor occupations to the Occupational Employment Statistics (OES) and Current Population Survey (CPS). We find comparable female wage penalties between Glassdoor and the relevant CPS subsample (high school+, restricted to Glassdoor-present occupations), indicating that Glassdoor estimates are a useful data point in terms of understanding gender wage differentials.[2]

Glassdoor data also includes company reviews provided by users. We examine the reviews data linked to a subsample of individuals who also provided salary reports, and find that females are less likely than males to list pay as a negative factor when providing anonymous feedback about their company. This highlights that males might be more vocal about salary concerns than women, which parallels existing evidence on gender differences in willingness to negotiate and self-promote.

This work relates to a broad literature examining gender differences in pay. Despite substantial decreases in the gender wage gap over the last 50 years, the female-to-male earnings ratio remains at 0.81, as estimated by the Bureau of Labor Statistics using publicly available data.[3] Considerable research has been devoted to examining potential reasons for the residual gap. These include: employer-specific factors such as taste-based and statistical discrimination (Becker (1957), Aigner and Cain (1977)); worker-specific factors, such as differences in labor market experience (Cook *et al.* (2018)), motherhood (Bertrand *et al.* (2010), Batchelder *et al.* (2010), Kleven *et al.* (2019)), and gender-specific preferences for competition (Niederle and Vesterlund (2007)), negotiation (Leibbrandt and List (2015)), and job flexibility (Mas and Pallais (2017), Bolotnyy and Emanuel (2019)); and occupation/industry-specific

[2]Work by Karabarbounis and Pinto (2018) compare Glassdoor data to the Quarterly Census of Employment and Wages and the Panel Study of Income Dynamics; while industry employment shares differ, they find mean salary industry-level correlations of approximately 0.9.

[3]This ratio is calculated by dividing the average female median usual weekly earnings by the average male median usual weekly earnings for full-time wage and salary workers. Changes in earnings must be considered in the broader historical context of women's advancement into the labor force, the shaping of their identities, and their increased human capital investment, which Goldin (2006) has described as the "most momentous change in the labor force of the twentieth century."

factors, such as differential sorting by gender into particular jobs and industries (Blau and Kahn (2017)). Of particular relevance is what happens within occupations; Goldin (2014) shows that flexible work and discontinuous hours come at a high cost in certain sectors (e.g., financial, corporate, and legal), which can perpetuate gender pay discrepancies relative to professions that do not penalize as much for flexible work (e.g., pharmacy, discussed in Goldin and Katz (2016)).

Our paper focuses on the role of occupational detail in quantifying the gender wage gap. It is most closely related to work by Marinescu and Wolthoff (2020) who use data from CareerBuilder.com and find that detailed job titles capture 90 percent of the variance in wages and over 80 percent of variance in applicant education and experience across job openings; this aligns with our findings that more detailed occupational titles provide valuable information about a worker's job. Our research is also related to work examining the role of employers: using Portuguese data, Card *et al.* (2015) find that firm-specific factors are an important aspect of the gender wage gap, and Goldin *et al.* (2017) use the Longitudinal Employer-Household Dynamics data to decompose the widening of the gender wage gap among U.S. college graduates into differential movement across establishments (44 percent) versus earnings differences within establishments (56 percent).

The rest of this chapter proceeds as follows. Section 3.2 describes the Glassdoor sample and presents descriptive statistics relative to publicly available data. Section 3.3 describes our empirical strategy and presents our main results on the gender wage gap in Glassdoor data. Section 3.4 concludes.

3.2 Glassdoor Sample

3.2.1 Salary Report and Review Data in Glassdoor

Our data come from Glassdoor.com. Glassdoor data is crowd-sourced: users anonymously submit demographic information about themselves, their salaries, occupations, employers, and job interviews through a survey administered through their website.[4] Unlike publicly administered survey data which employs random sampling, the Glassdoor sample reflects self-reports from online job-seekers, and so is non-random and less likely to represent workers from the lower-end of the earnings distribution.[5] However, Glassdoor earnings data are also not top-coded, which potentially provides more information about the upper end of the income distribution.

Our data reflects salary reports over the 2008 to 2019 period.[6] Salary reports contain information on an individual's employer, employment status (full-time, part-time, intern, etc.), whether it is a current or former job, work location, job title, salary, and years of experience (see Figure 3.1).[7] Age, education, and gender are optional information that users can voluntarily provide through their account settings, though few users actually do so; instead, most of the information on these variables is obtained from users who login to Glassdoor through Facebook.[8]

We also have access to company review data from Glassdoor. Reviews data are provided through a similar survey on Glassdoor's website (see Figure 3.2).[9] Information solicited includes company name, current employment status, job title, overall company rating (a 1 through 5 rating, with 5 being the highest value), and textual data on pros, cons, and advice to management.

---

[4]The salary survey is provided here: https://help.glassdoor.com/article/Adding-a-salary/.

[5]Some salary report submissions are part of Glassdoor's "give-to-get" model, whereby access to the website is available for users who contribute information about a company or job position.

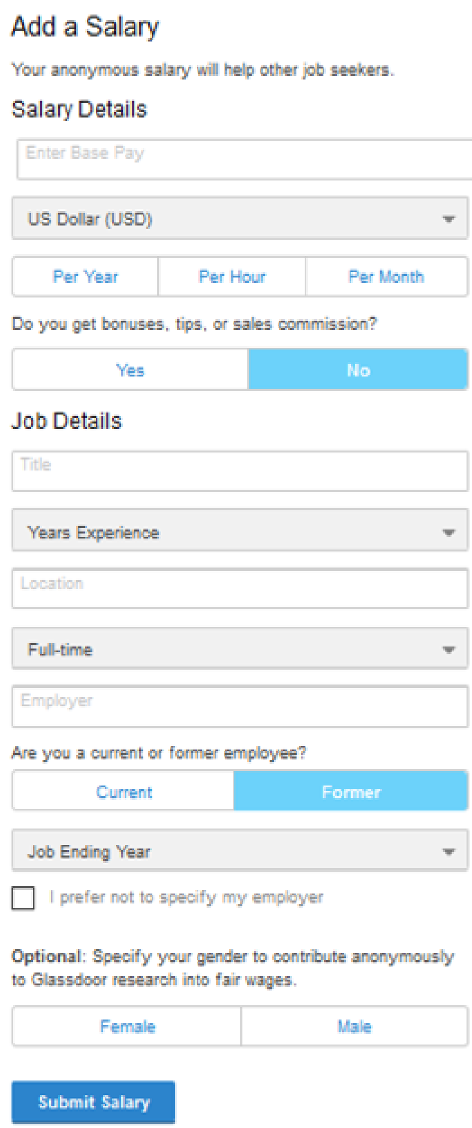
[6]See Appendix Table C.1 for detail on observations per year.

[7]Base pay is the only salary measure that we use in our analysis. Other forms of compensation are provided if the user voluntarily submitted information on bonuses, sales, and tips; we only utilize base salary information (not inclusive of bonus pay) for comparability with the publicly available data. Individuals can provide a salary report for a job held up to four years in the past, though the year that the job ended must be identified. If an employer is not already in the Glassdoor database, the individual must provide information about the employer, including company website, headquarters, employer type (public, private, etc.), and firm size. Users can submit multiple salary reports, though almost all of our sample (99 percent) comprises users who submitted one salary report while at a given employer and holding a given job title.

[8]Race and data on hours worked per day or week are not provided.

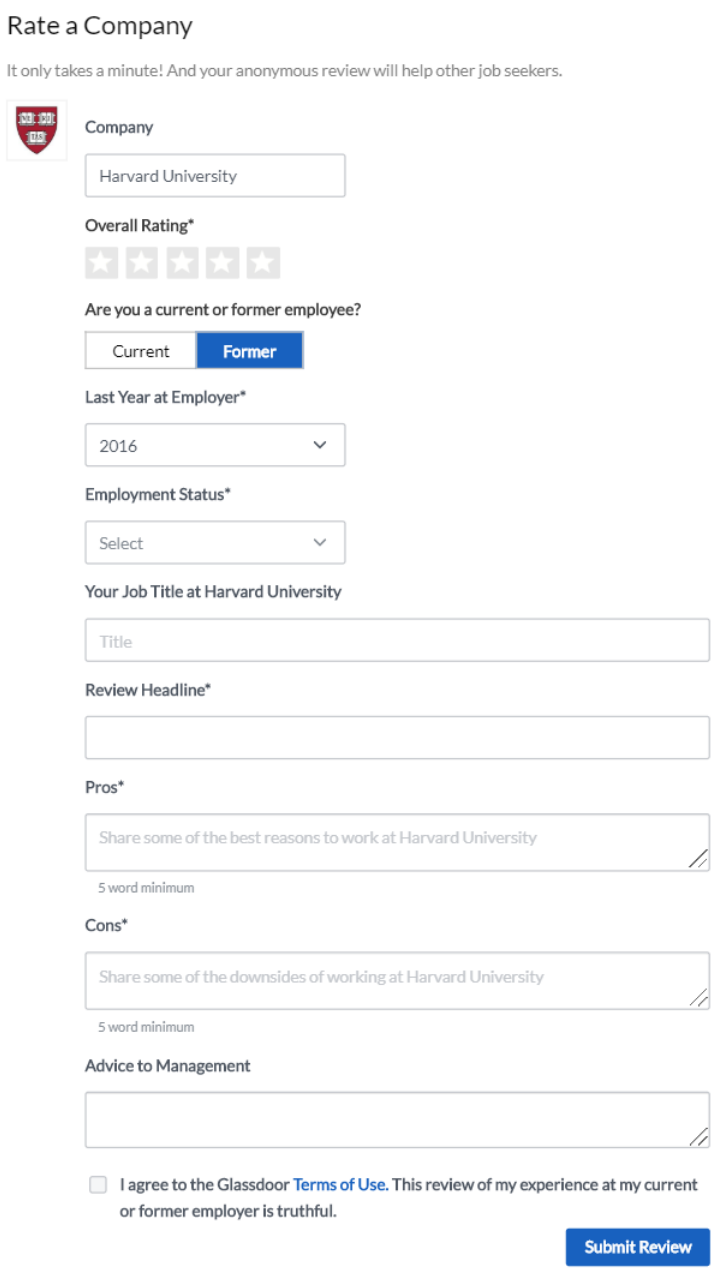
[9]The review survey is provided here: https://help.glassdoor.com/article/Writing-a-company-review.

Figure 3.1: *Glassdoor Salary Report Survey*



Notes: This figure represents the voluntary salary report that an individual can submit on Glassdoor's website. Information solicited includes pay, job details (i.e., title, years of experience, location, full-time/part-time status, company name), and optional gender information.

Figure 3.2: *Glassdoor Review Survey*



Rate a Company

It only takes a minute! And your anonymous review will help other job seekers.

Company

Harvard University

Overall Rating*

Are you a current or former employee?

Current Former

Last Year at Employer*

2016

Employment Status*

Select

Your Job Title at Harvard University

Title

Review Headline*

Pros*

Share some of the best reasons to work at Harvard University

5 word minimum

Cons*

Share some of the downsides of working at Harvard University

5 word minimum

Advice to Management

I agree to the Glassdoor Terms of Use. This review of my experience at my current or former employer is truthful.

Submit Review

Notes: This figure represents the voluntary company review that an individual can submit on Glassdoor's website. Information solicited includes company name, current employment status, job title, overall company rating, pros, cons, and advice to management.

3.2.2 How Does Glassdoor Compare to Publicly Available Data?

Given the fundamentally different data collection process, there are notable differences between Glassdoor and publicly available data, as depicted in Table 3.1.[10] In particular, the average hourly wage is higher in Glassdoor relative to the Current Population Survey (\$39.30 versus \$27.67), and individuals in the Glassdoor sample tend to be younger (35 versus 43 years old) and primarily reflect bachelor's degree holders. Figure 3.3 shows that the distribution of Glassdoor hourly earnings is more skewed to the right relative to the CPS.

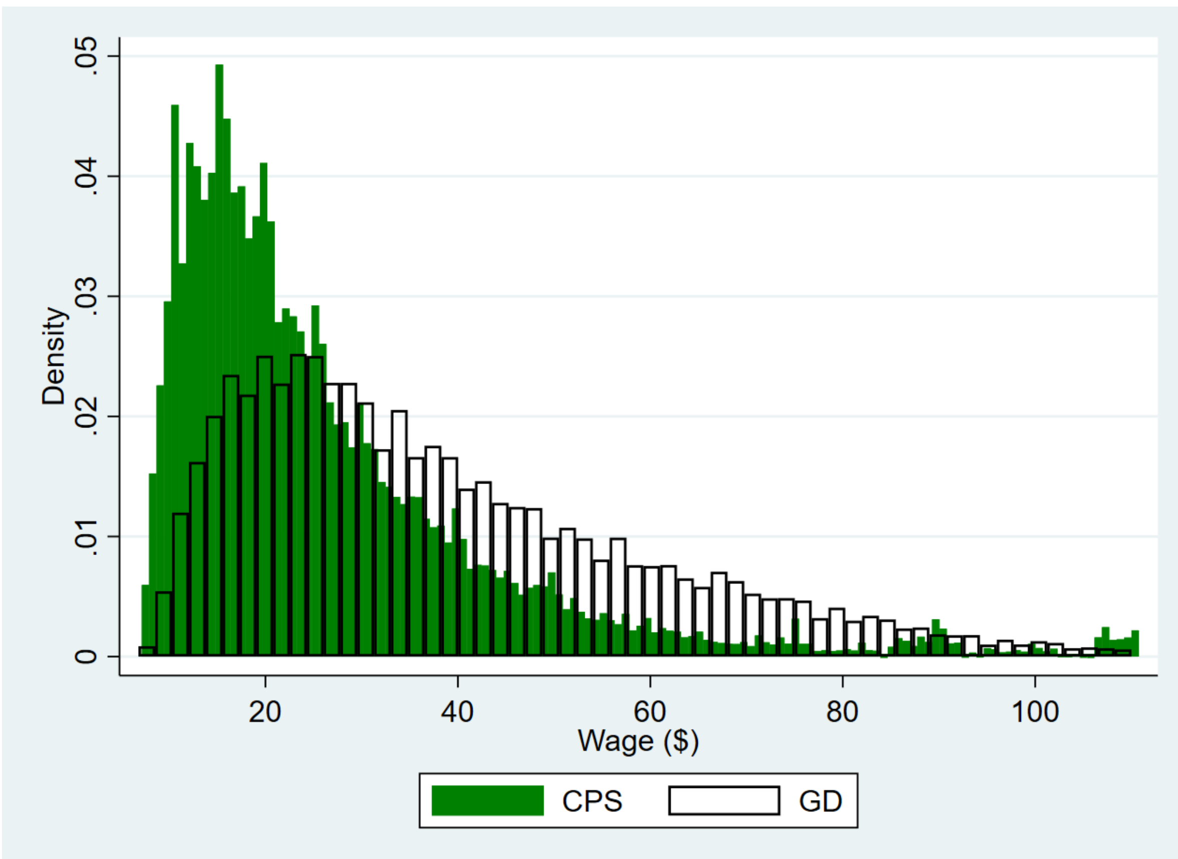
Table 3.1: *Glassdoor versus CPS: Descriptive Statistics*

	GD Mean	CPS Mean
Wage (\$2018 dollars)	39.30	27.67
Share Female	0.41	0.43
Age (years)	34.92	42.62
Experience (years)	6.63	–
< High School	–	0.07
High School	0.09	0.40
Associates	0.03	0.11
Bachelors	0.63	0.26
Masters or MBA	0.24	0.12
JD or MD	0.004	0.02
PhD	0.01	0.02
Observations	906,971	366,492
Employers	40,057	–
Glassdoor Industries	141	–
Glassdoor Occupations	1,684	–

Notes: This table displays summary statistics for the Glassdoor sample compared to the Current Population Survey (CPS) sample. Both samples reflect full-time, non-self-employed workers aged 25 to 64 in the United States. The Glassdoor sample reflects data collected between 2008 and 2019, while the CPS sample reflects data from the 2013-2018 time period. See Appendix Document C.1 and C.2 for more detail on sample construction.

[10]Both samples reflect full-time, non-self-employed workers aged 25 to 64 in the United States. See Appendix Document C.1 and C.2 for details on our cleaning procedures for Glassdoor and the CPS.

Figure 3.3: *Glassdoor versus CPS: Wage Distribution*



Notes: This figure represents the Glassdoor hourly wage distribution compared to the hourly wage distribution for workers in the Current Population Survey (see Appendix Document C.2 for details on the CPS sample). Earnings are in 2018 dollars. This graphical distribution is truncated at the 99th percentile of earnings in both samples.

While Glassdoor data represents a selected subsample, it is still useful to examine given the additional richness in occupational detail. Every salary report requires the individual to provide a job title associated with their occupation, which Glassdoor then categorizes into an aggregated Glassdoor occupation category using its proprietary machine-learning model.[11] Glassdoor occupations provide greater occupational detail along two dimensions: by providing job specificity (e.g., listing *sonographer* or *radiologic technologist* versus the comparable broader occupational category in the CPS, *diagnostic related technologists and technicians*), and by providing more detail on the hierarchical level of the job (e.g., *compliance manager* versus *compliance officer*). Figure 3.4 displays the top 10 Glassdoor occupation categories (gocs) in our sample. While these occupations are slightly more general in nature, several reflect hierarchical information.[12]

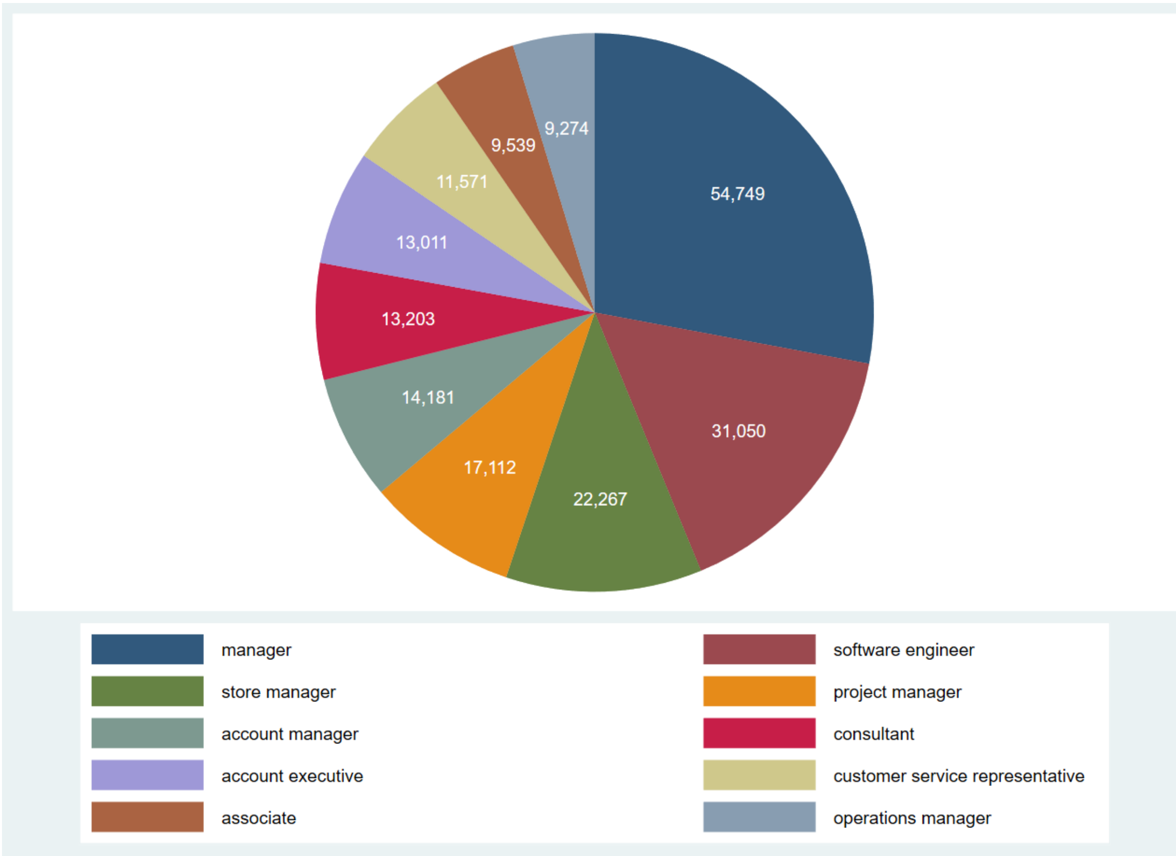
In order to better assess the occupational differences between Glassdoor and publicly available data, we manually constructed a mapping between the 1,684 Glassdoor occupation categories and the 455 broad occupational categories and 22 major occupational categories in the 2018 Occupational Employment Statistics (OES).[13] Table 3.2 compares the employment shares between the major OES categories and Glassdoor. Unsurprisingly, Glassdoor data is more heavily skewed toward Management, Business & Financial Operations, and Computer & Mathematical occupations relative to the publicly available data. We also compare wage levels between Glassdoor and the OES at the broad occupation level. Two hundred and twenty-three of the 455 OES broad occupation categories are represented in Glassdoor. Of these, 64, or approximately 30 percent are "aligned", defined as OES occupations where the

[11]The model uses common phrases in the job titles as well as clicking behavior by users on the "Jobs" portion of the website to create a "distance metric" between similar job titles. Critically, there is no mechanical correlation between salaries and clicking behavior. The model only provides categorization when it is 70% to 80% accurate.

[12]There are some Glassdoor occupation categories which are similar, though they are recognized as distinct using the additional data that feeds into Glassdoor's machine learning algorithm; future work could test the robustness of our results after merging comparable categories. See Appendix Figure C.1 for the top 10 Glassdoor industry categories in our sample.

[13]The OES data reflect detailed employment and wage estimates from a survey of nonfarm establishments. Where ambiguous, we make our best estimate of the associated OES category and use the O*NET-SOC auto-coder website for additional help in the classification process.

Figure 3.4: *Glassdoor's Top 10 GOCs*



Notes: This figure represents the top 10 Glassdoor occupation categories (gocs) and associated number of salary reports in our data. These occupation categories account for approximately 22 percent of the full sample.

corresponding Glassdoor wage is within 15 percent of the OES' hourly wage distribution at the 10th, 25th 50th, 75th and 90th percentiles.[14]

We marry our occupational mapping with the demographic information provided in the CPS to further assess Glassdoor's comparability with publicly available data. In particular, we extract the relevant occupations in the CPS (i.e., CPS occupations that correspond to OES occupations in our Glassdoor sample) and restrict the CPS to high school+ degree-holders. We then employ the DiNardo-Fortin-Lemieux (DFL) procedure to reweight

[14]Depending on how many Glassdoor occupations are associated with an OES occupation, multiple Glassdoor occupation wages may be grouped together for this analysis.

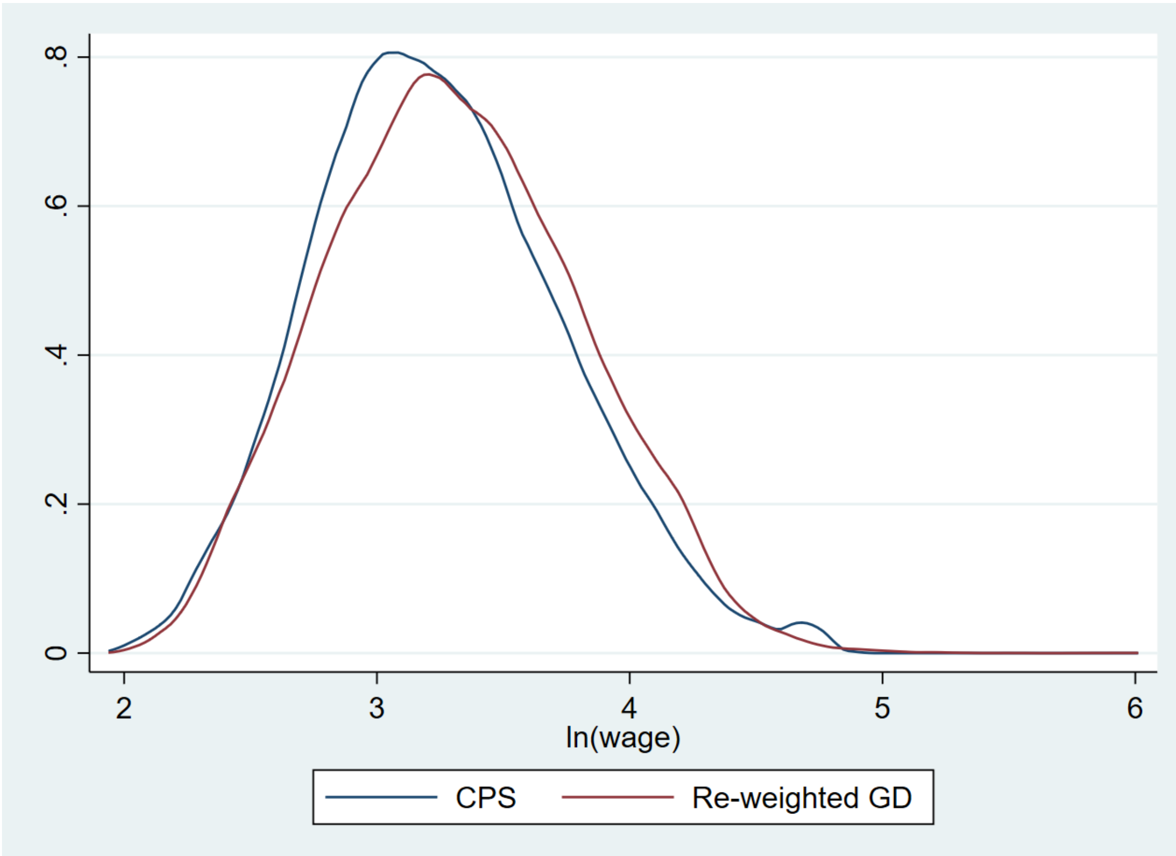
the Glassdoor sample to have the education, occupation, and state characteristics of the CPS subsample. Figure 3.5 displays the CPS subsample and the reweighted Glassdoor sample; while there are still differences between the two, the earnings distribution is more similar than the unrestricted, unweighted version (Figure 3.3).

Table 3.2: *Glassdoor Compared to OES Employment Shares*

Occupation	GD Employment	OES Employment
Management	31.6	5.3
Business and Financial Operations	23.5	5.3
Computer and Mathematical	9.8	3.0
Architecture and Engineering	4.9	1.8
Life, Physical, and Social Science	2.6	0.8
Community and Social Service	0.6	1.5
Legal	0.6	0.8
Education, Training, and Library	1.8	6.1
Arts, Design, Entertainment, Sports, and Media	3.7	1.3
Healthcare Practitioners and Technical	3.2	6.0
Healthcare Support	0.4	2.8
Protective Service	0.6	2.4
Food Preparation and Serving Related	0.8	9.2
Building and Grounds Cleaning and Maintenance	0.1	3.1
Personal Care and Service	0.2	3.8
Sales and Related	4.8	10.0
Office and Administrative Support	8.5	15.1
Farming, Fishing, and Forestry	< 0.01	0.3
Construction and Extraction	0.1	4.1
Installation, Maintenance, and Repair	0.4	3.9
Production	0.5	6.3
Transportation and Material Moving	1.2	7.1

Notes: This table displays the Glassdoor occupational distribution relative to the 2018 Occupational Employment Statistics (OES) employment shares. We manually categorize 1,684 Glassdoor occupations into the OES's 22 broad occupational groups to facilitate this comparison.

Figure 3.5: *Glassdoor Reweighted to CPS Distribution*



Notes: The above figure represents the output from the DiNardo-Fortin-Lemieux (DFL) reweighting procedure. We restrict the CPS sample to individuals with a high-school degree or higher who work in the same occupations represented in the Glassdoor sample. We then re-weight the Glassdoor sample to reflect the education, state, and occupational distribution in the restricted CPS sample, and graph the ln(hourly wage) for both distributions.

3.3 Glassdoor Gender Wage Gap

3.3.1 More Precise Occupational Titles Narrow the Gender Wage Gap

We estimate the female-male wage gap under increasingly rich fixed effects, beginning from the following equation:

$$\ln\left(wage_{isy}\right) = \alpha + \beta Female_i + \gamma_s + \delta_y + \varepsilon_{isy} \tag{3.1}$$

Above, i indexes individual, s state, and y year. The coefficient of interest is β. In this initial equation we include vectors of fixed effects for state γ_s and year δ_y. Subsequent estimating equations incorporate progressively more fixed effects, until we arrive at the following:

$$\ln\left(wage_{ioy}\right) = \alpha + \beta Female_i + \gamma_s + \delta_y + \theta_1 age_i + \theta_2 age_i^2 + \theta_3 experience_i + \theta_4 experience_i^2 + \eta_i + \rho_o + \sigma_k + \varepsilon_{isyok} \quad (3.2)$$

This specification adds demographic controls for age, experience, highest education fixed effects η_i, occupation fixed effects ρ_o, and employer (or industry) fixed effects σ_k. While the variation identifying our estimate of β is not a causal estimate, because we cannot rule out selection by gender into particular companies and occupations, detailed occupational and employer controls potentially provide more precision in estimating the female wage penalty.

Table 3.3 presents the results of our main specification, which regresses the natural log of the hourly wage on a female dummy variable and controls. The coefficient on the female dummy variable starts out at -0.207 (or a female-to-male earnings ratio of approximately 0.81) with no controls except for state and year fixed effects. When age and human capital variables (age, age-squared, experience, experience-squared, and fixed effects for highest educational attainment) are included, the coefficient decreases to -0.177 (a ratio of 0.84).[15] This is roughly in-line with estimates for the female coefficient using publicly available data.

The subsequent columns control for Glassdoor-specific occupation, industry, and employer detail. The female wage penalty is reduced to -0.076 (a ratio of 0.92) when we include fixed effects for Glassdoor's 1,684 specific occupations (Column 3). The wage penalty is further reduced to between -0.065 and -0.059, a ratio of 0.94, when we additionally include industry or employer fixed effects (Columns 4 and 5). This is a smaller female wage penalty than has traditionally been estimated in the literature (Goldin (2014), Blau and Kahn (2017)).

[15] The educational attainment categories are High School, Associates, Bachelors, JD, Masters, MBA, MD, and PhD.

Table 3.3: *Glassdoor Gender Wage Gap*

	ln(Wage)				
Female	-0.207***	-0.177***	-0.0761***	-0.0651***	-0.0592***
	(0.00110)	(0.000939)	(0.000764)	(0.000726)	(0.000664)
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	906,971	906,971	906,971	906,971	906,971
Adjusted R^2	0.161	0.407	0.679	0.711	0.770
Note:			*p<0.1; **p<0.05; ***p<0.01.		

Notes: This table displays regressions of the natural log of the hourly wage in 2018 dollars on a female dummy variable and controls, which include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include state and year fixed effects. Robust standard errors are in parentheses. There are 141 industry categories, 1,684 Glassdoor occupation categories, and 40,057 employers in the sample.

Our analyses highlight that controlling for occupational detail accounts for a large portion of the difference in male and female earnings. Occupational detail is useful to more precisely group together individuals who perform similar tasks and thus should be paid comparably, but controlling for very precise occupational information (e.g., job title) could explain away virtually all pay discrepancies, and so could imply an artificially low gender wage gap. The adjusted R-squared from the regression in Column 3 (0.6787) is close to the unadjusted R-squared, 0.6794, indicating that Glassdoor's occupational categories seem to meaningfully control for gender differences in pay.

Another advantage of the Glassdoor data is our ability to examine the top-end of the earnings distribution, which is not possible in publicly available data as the latter is typically top-coded. Table 3.4 reflects the same gender wage gap analysis after we restrict the sample to OES occupations with mean hourly earnings of $35 dollars or more (e.g., physicians and surgeons, chief executives, managers, lawyers). The patterns from this table are similar to those in Table 3.3; females earn approximately 92 cents on the dollar once Glassdoor occupational detail is included in the regression (Column 3). This highlights that females at all parts of the earnings distribution experience a wage penalty (excluding differences in bonus pay and other one-time compensation).

Table 3.4: *Glassdoor Gender Wage Gap for High-Wage Occupations*

	ln(Wage)				
Female	-0.188***	-0.163***	-0.0788***	-0.0673***	-0.0604***
	(0.00115)	(0.000981)	(0.000822)	(0.000780)	(0.000715)
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	807,582	807,582	807,582	807,582	807,582
Adjusted R^2	0.153	0.401	0.654	0.690	0.755
Note:					*p<0.1; **p<0.05; ***p<0.01.

Notes: This table displays regressions of the natural log of the hourly wage in 2018 dollars on a female dummy variable and controls, which include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to OES occupations with a mean wage of $35 or higher. There are 141 industry categories, 1,445 Glassdoor occupation categories, and 39,504 employers in the sample.

3.3.2 Comparison to CPS Female Wage Penalty

The Glassdoor data provide evidence that females receive 0.92-0.94 cents on the dollar, smaller than estimates of the female wage penalty in publicly available data. We can directly compare the female wage penalty observed in the Glassdoor data to the relevant CPS subsample using our manually constructed occupational mapping between Glassdoor and the OES occupations.

Table 3.5 presents the results of this analysis. Column 1 displays the female wage

Table 3.5: *Glassdoor versus CPS Gender Wage Gap*

	ln(Wage)				
	GD			CPS	
Female	-0.0772***	-0.106***	-0.102***	-0.131***	-0.132***
	(0.000764)	(0.000860)	(0.00141)	(0.00215)	(0.00370)
Age, experience, education	X	X	X	X	X
GD Occupation FE	X				
OES Occupation FE		X	X	X	X
Aligned OES			X		X
Observations	906,971	906,971	284,743	286,758	91,697
Adjusted R^2	0.678	0.567	0.596	0.505	0.555
Note:			*p<0.1; **p<0.05; ***p<0.01.		

Notes: This table compares the Glassdoor gender wage gap with those in the CPS. The gender wage gap is calculated by regressing the natural log of the hourly wage in 2018 dollars on a female dummy variable and controls, which include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level (we collapse the JD and MD categories in Glassdoor for comparability with the CPS); experience is not available in the CPS sample, so we substitute with potential experience, calculated as the minimum of (Age - years of education - 6) or (Age - 16). All regressions include state and year fixed effects. Robust standard errors are in parentheses. Column 1 displays the full Glassdoor sample and includes fixed effects for the 1,684 Glassdoor occupations. Column 2 displays the full Glassdoor sample but controls for a broader categorization of occupations using 223 OES occupations. Column 3 further restricts the Glassdoor sample to "aligned" OES occupations (64 categories), defined as Glassdoor occupations where the Glassdoor wage is within 15 percent of the OES wage at the 10th, 25th 50th, 75th and 90th percentiles of the OES occupation's hourly wage distribution. Column 4 reflects the same regression as Column 2 in the relevant CPS subsample, where the CPS subsample reflects individuals with a high-school degree or higher who work in the same occupations represented in the Glassdoor sample. Column 5 further restricts the CPS subsample to the "aligned" OES occupational categories, as previously defined.

penalty in the Glassdoor sample after accounting for age and human capital variables and Glassdoor occupation fixed effects (analogous to Column 3 in Table 3.3).[16] Column 2 replaces the detailed Glassdoor occupations with the 223 OES associated occupations, and the female wage penalty goes up to -0.106 (a ratio of 0.90) as a result of using this coarser level of occupational information. We further restrict the sample in Column 3 to

[16] Note that we collapse the *JD* and *MD* educational categories for comparability with the CPS, which groups these educational levels together.

"aligned" OES occupations, defined as OES occupations where the Glassdoor wage is within 15 percent of the OES occupation's hourly wage distribution at the 10th, 25th 50th, 75th and 90th percentiles.

Column 4 replicates the analysis in Column 2 but uses the CPS subsample with high-school+ degree holders who work in the OES occupations represented in the Glassdoor sample. The magnitude of the female coefficient, -0.131, is slightly larger than the comparable female wage penalty in the Glassdoor sample (-0.106). A similar pattern exists when we restrict to OES aligned occupations in the CPS (Column 6 compared to Column 3). The relatively small difference between the Glassdoor and CPS subsample estimates, on the order of 2 percentage points, provides evidence that Glassdoor's female wage penalty is a useful metric of the gender wage gap, despite the selected nature of the Glassdoor sample.

3.3.3 Do Women Complain about their Salary?

The prior analysis provides evidence that women are paid less than men even after detailed occupation and employer controls are included. Given that Glassdoor's website serves as both a source of self-reported salary information as well as a repository of reviews data, we ask a natural question: do women list pay as a "con" more than men do?

To perform this analysis, we use the portion of Glassdoor's reviews data which have a corresponding salary report, i.e., individuals who submitted a company review and salary review for the same company under the same job title; this is a smaller subsample of the salary data since not everyone who submits a salary report also provides a company review. Information solicited in the company review includes company name, current employment status, job title, overall company rating (1-5 rating), and textual data on pros, cons, and advice to management.

Table 3.6 presents the results of our analysis. The dependent variable is a dummy variable which equals 1 if the individual listed *salary*, *wage*, *pay*, *benefits*, or *compensation* in the "con" section of the company review. Approximately 20 percent of females and males list salary or one of these related words as a con.[17] We control for the natural log of the hourly wage and age and human capital variables, as well as Glassdoor occupation, industry, and employer detail where indicated. State and year fixed effects are included in all specifications.

Table 3.6: *Reviews Data: Do Females Complain about Pay?*

	Complain about Pay				
Female	-0.0216***	-0.0201***	-0.0188***	-0.0202***	-0.0210***
	(0.00498)	(0.00500)	(0.00541)	(0.00546)	(0.00778)
ln(Wage)	X	X	X	X	X
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	31,529	31,529	31,529	31,529	31,529
Adjusted R^2	0.026	0.029	0.043	0.047	0.078
Note:					*p<0.1; **p<0.05; ***p<0.01.

Notes: This table displays regressions of a dummy variable indicating whether the individual listed salary, wage, pay, benefits, or compensation in the "con" section of the company review on a female dummy variable; controls include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include the natural log of the hourly wage, and state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to individuals who provided both a Glassdoor review for their employer and a corresponding salary review while working for the same employer and holding the same job title. There are 135 industry categories, 1,583 Glassdoor occupation categories, and 10,371 employers in the sample.

As can be seen from the *Female* coefficient, women are approximately 2 percentage points less likely to list *salary* or synonymous words as a con in the review.[18] The fact that

[17]Future analysis could control for sentiment in the surrounding text next to the given word; however, given that these words are listed in the "con" column, we think it is a reasonable proxy for having a complaint about pay.

[18]See Appendix Table C.2 for the female wage penalty in this subsample, between -0.085 and -0.061. These results are robust to omitting the word "benefit" from the list of words (see Appendix Table C.3). The *Female*

women are less likely than comparable men to complain about salary on an anonymous website raises questions about whether females receive less information about pay or are simply less vocal about salary concerns. If the latter explanation is true, it would be in line with related gender differences in self-promotion and self-review (Abraham (2020), Exley and Kessler (2019)) and evidence that women are less likely to negotiate about their salary (Leibbrandt and List (2015)).[19] Future work could explore heterogeneity in this result (e.g., specific occupations) to better understand mechanisms for gender differences in being less vocal about salary.

3.4 Conclusion

This paper provides new findings on the extent of gender wage differentials using a unique dataset. Glassdoor data reveals a slightly smaller female wage penalty, on the order of 6 to 8 percent when more specific occupational detail and employer fixed effects are taken into account. Occupational detail, in particular, seems to make a significant difference in explaining earnings differences between males and females. Glassdoor data also provide more detail on top-end earnings relative to publicly available data; we find a similar female-to-male ratio, on the order of 0.92 to 0.94 cents on the dollar, when examining high-wage occupations.

These results must be understood within the context of differences between our data and publicly available data. While Glassdoor data are selected, our comparisons to the CPS indicate that the female wage penalty is roughly 2 percentage points smaller than publicly available data for the given subsample. This difference helps provide bounds on our Glassdoor estimates given data selection concerns.

The Glassdoor reviews data indicate that females are significantly less likely to

coefficient is not negative and significant in all specifications when we do not control for the log of the hourly wage, indicating that salary is an important control variable (see Appendix Table C.4).

[19]Women are not less likely than men to complain in general; as shown in Appendix Table C.5, females are significantly more likely than males to list discrimination-related words in the con text.

complain about pay in the "con" section of the company review. While it is unclear whether females are less aware of salary differences or less likely to be vocal about their salary concerns in an anonymous review, these gender differences in behavior parallel similar findings in the literature on females' willingness to negotiate about pay and self-promote relative to their male counterparts. More work needs to be done to better understand how these behaviors affect labor market outcomes. Additional work using Glassdoor's reviews data could shed light on whether employers with larger female wage penalties are associated with more negative reviews or more discrimination complaints.

Future research on the female wage penalty and associated mechanisms, particularly using novel sources of data, could provide valuable insights for policymakers and researchers interested in designing labor market policies to narrow gender pay gaps. Glassdoor data corroborate the value of occupational detail in explaining part of the female wage penalty, and demonstrate how proprietary data can help quantify gender differences in the labor market.

References

ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. (2017). When should you adjust standard errors for clustering?

—, CHINGOS, M. M. and WEST, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, **100** (4), 567–580.

ABRAHAM, L. (2020). Do women underrate their performance?: Evidence from hr data. *Working Paper*.

— and STEIN, A. (2020). Words matter: Experimental evidence from job applications. *Working Paper*.

AIGNER, D. J. and CAIN, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, **30** (2), 175–187.

APPLE (2019). Different together. Https://www.apple.com/diversity/.

AVILOVA, T. and GOLDIN, C. (2018). What can uwe do for economics? In *AEA Papers and Proceedings*, vol. 108, pp. 186–90.

BABCOCK, L., RECALDE, M. P. and VESTERLUND, L. (2017a). Gender differences in the allocation of low-promotability tasks: The role of backlash. *American Economic Review*, **107** (5), 131–35.

—, —, — and WEINGART, L. (2017b). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, **107** (3), 714–47.

BATCHELDER, L., ELLWOOD, D. T. and WILDE, E. T. (2010). The mommy track divides: The impact of childbearing on wages of women of differing skill levels. *Working Paper*.

BECKER, G. S. (1957). *The economics of discrimination: an economic view of racial discrimination*. University of Chicago.

BENSON, A., LI, D. and SHUE, K. (2020). Promotions and the peter principle. *The Quarterly Journal of Economics*, **134** (4), 2085–2134.

BERTRAND, M., GOLDIN, C. and KATZ, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, **2** (3), 228–55.

BLAU, F. D. and KAHN, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, **55** (3), 789–865.

Bolotnyy, V. and Emanuel, N. (2019). Why do women earn less than men? evidence from bus and train operators. *Working Paper*.

Bowles, H. R., Babcock, L. and Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, **103** (1), 84–103.

Cappelli, P. and Conyon, M. J. (2018). What do performance appraisals do? *ILR Review*, **71** (1), 88–116.

Card, D., Cardoso, A. R. and Kline, P. (2015). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, **131** (2), 633–686.

Castilla, E. J. and Rho, H. J. (2015). Language and gender in the online recruitment process. *Working Paper*, pp. 1–49.

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

—, Collis, M. and Kulkarni, L. (2019). When to apply? *Working Paper*.

Cook, C., Diamond, R., Hall, J., List, J. A. and Oyer, P. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Working Paper*.

Del Carpio, L. and Guadalupe, M. (2018). More women in tech? evidence from a field experiment addressing social identity.

Dunning, D. and Kruger, J. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, **77** (6), 1121.

Exley, C. L. and Kessler, J. B. (2019). The gender gap in self-promotion. *Working Paper*.

Facebook (2019). Facebook diversity update. Https://www.facebook.com/careers/diversity-report.

Flory, J. A., Leibbrandt, A. and List, J. A. (2014). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, **82** (1), 122–155.

Frederiksen, A., Lange, F. and Kriechel, B. (2017). Subjective performance evaluations and employee careers. *Journal of Economic Behavior & Organization*, **134**, 408–429.

Gee, L. K. (2018). The more you know: information effects on job application rates in a large field experiment. *Management Science*.

Ginther, D. and Kahn, S. (2017). Women and stem. *NBER Working Paper*.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American economic review*, **96** (2), 1–21.

— (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104** (4), 1091–1119.

— and Katz, L. F. (2016). A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, **34** (3), 705–746.

—, Kerr, S. P., Olivetti, C. and Barth, E. (2017). The expanding gender earnings gap: Evidence from the lehd-2000 census. *American Economic Review*, **107** (5), 110–14.

Google (2019). Google diversity annual report 2019. Https://diversity.google/annual-report/.

Karabarbounis, M. and Pinto, S. (2018). What can we learn from online wage postings? evidence from glassdoor. *Economic Quarterly*, (4Q), 173–189.

Kleven, H., Landais, C. and Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, **11** (4), 181–209.

Leibbrandt, A. and List, J. A. (2015). Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science*, **61** (9), 2016–2024.

Marinescu, I. and Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, **38** (2), 535–568.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, **107** (12), 3722–59.

Mohr, T. S. (2014). Why women don't apply for jobs unless they're 100% qualified. *Harvard Business Review*, **25**.

Murciano-Goroff, R. (2018). Missing women in tech: The labor market for highly skilled software engineers. *Working Paper*.

National Center for Education Statistics (2019). Status and Trends in the Education of Racial and Ethnic Groups: Indicator 26: STEM Degrees. Https://nces.ed.gov/pubs2019/2019038.pdf.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, **122** (3), 1067–1101.

Rudman, L. A. and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, **57** (4), 743–762.

— and Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, **28**, 61–79.

Sandberg, S. (2013). *Lean in-Women, Work and the Will to Lead*. Random House.

U.S. Department of Labor Bureau of Labor Statistics (2019). Celebrating Women in STEM Occupations. Https://beta.bls.gov/labs/blogs/2019/03/06/celebrating-women-in-stem-occupations/.

Appendix A

Appendix to Chapter 1

A.1 Description of Hiring Process at Uber

Uber's external hiring process begins with the recruiters who review each individual application that comes through the careers website. Recruiters at this stage make an assessment about whether the applicant would be a good fit for the role based on the application information. In a minority of cases, the recruiter will move the applicant to a different job posting (i.e., not the one that the applicant originally applied for) if the applicant seems better suited for another role.

Applicants who are considered suitable for the role will have a phone screen with the recruiter (e.g., "Recruiter Phone Interview"). Only 9 percent of applications in our experiment received a Recruiter Phone Interview. The goal of this initial phone interview is to set expectations with the applicant about the hiring process and to learn more about the applicant's background.

Approximately 3 percent of all applications reached the next stage of the process, which is a phone interview with the hiring manager for the particular role (e.g., "Team Phone Interview"). The purpose of this phone interview is to more deeply evaluate the applicant's background, skills, and qualifications to see if he or she would be a good fit for the specific role. Most applicants who receive a team phone interview are also given a creative or analytical exercise that tests their skills related to the specific role/team they are interviewing for. Once they finish the exercise, it is either a) sent to the hiring manager for review or b) presented by the candidate to the hiring manager and/or broader team.

After the exercise is reviewed, successful candidates reach the on-site interview stage. Approximately 2.5 percent of applications received an on-site interview in our sample. This is an in-person interview with the hiring manager and typically three to six additional interviewers. The on-site interview is designed to ascertain the applicant's fit with the team and job-specific skills.

Applicants who perform successfully at the on-site interview stage are given an offer of employment (0.7 percent of applications in our sample). The entire hiring process typically takes two to four weeks.

A.2 Experiment Execution

A. Technical Execution

Our experiment was executed with the A/B testing software Optimizely and custom-made technology from Cro Metrics, the company that helped us technically implement the experiment. Randomization occurred at the browser-device level by saving pertinent IDs into local storage. This is similar to identifying by IP address, with the difference being that if an individual opened a different browser (even on the same device), or used a different device (even on the same IP address), Optimizely would consider that individual to be a different viewer. Clearing local storage (which occurs when clearing history or browsing data) would also induce Optimizely to treat the individual as a new viewer. An individual using the same browser-device on two different IP addresses would still be treated as the same viewer. Individuals with ad-blocking browser extensions were excluded from the experiment since ad-blocking software entirely blocks Optimizely.

Randomization occurred when an individual clicked on a job posting with the "www.uber.com/careers/list/" structure (regardless of whether or not that particular job posting was included in the experiment). This URL format automatically excluded individuals who accessed the non-US versions of the careers website (e.g., "www.uber.com/es-US/careers/list/").[1] Once an individual was randomized, he or she saw the Treated version for *all* job postings in the experiment or the Control version for *all* job postings in the experiment. Each newly identified viewer was equally likely to be bucketed into the Treated or Control condition. There were 203,232 individuals who were randomized over the duration of our experiment. Of these 203,232 individuals, 59,140 viewed a job posting included in our experiment, in-line with the approximately 30% of job postings that were treated on the careers website at any given point in time.

Our experiment linked treatment status from randomization with individual ap-

[1] On April 27, 2019, Uber redesigned their careers website and the original URL migrated to "www.uber.com/us/en/careers/".

plications through the use of cookies. After page load for all job postings with the "www.uber.com/careers/list/", iis and iisp cookies were cleared for the given individual (iis cookies denoted the "uber.com/careers" domain and iisp cookies denoted the Treated or Control tag, "he-6594079" or "he-6584079", respectively). Then the software that replaced the Control qualifications text with the Treated qualifications text also re-added the iis cookies and iisp cookies for the job postings included in the experiment. The "he-6594079" or "he-6584079" parameters were simultaneously added to the webpage URL when the iisp cookies were added (i.e., job postings in the experiment changed from *https://www.uber.com/careers/list/45142/* to *https://www.uber.com/careers/list/45142/ ?iis=uber.com/careers&iisp=he-6594079*, which was the only visible difference between job postings in the experiment and those not in the experiment to the public viewer).

Individuals that chose to click on the "Apply Now" button at the top or bottom of the job posting would be taken to an interstitial page (see Appendix Figure A.2) where the applicant chose a method for applying (e.g., LinkedIn, Manual, Resume, or SignIn), filled out the necessary information, and then submitted the application by clicking "Apply Now" at the very bottom of the page. Because iisp cookies were preserved at this stage, treatment status flowed into Uber's recruiting database thereby attributing Treated or Control assignment to the given application for use in our analysis.

B. Application Information

The resume text is provided for all applications regardless of application method (LinkedIn, Resume, Manual, or SignIn). Though all methods ask the same information, there are differences in terms of what is stored in Uber's recruiting system. In particular, the Resume method stores the full text of the resumes, while the LinkedIn and Manual methods only pull the information associated with the application form in Appendix Figure A.3. The SignIn method represents the last method an individual used to apply through Uber's careers website. In practice, both the Resume and SignIn textual data provide more information in our analysis than the other methods.

Because all applications store the latest version of what the candidate last submitted to the Uber careers website, there is the possibility for mismeasurement in the application method variable. More specifically, if the same person (as identified by email address) later applied to another Uber job posting using a different method or made other updates to their resume information, the new information would be stored in their profile and assigned to all prior applications. Thus the application method assigned to a given candidate-application is "overwritten" by later use of a different method. Given this mismeasurement, we do not control for application method in our regressions. However, controlling for application method does not materially change our results.

C. Implementation

Daily and weekly experimental implementation consisted of several steps. First, we scraped the Uber careers website once per day over the experiment duration to determine when U.S. job postings were opened or expired (to the nearest day) and when long-digit-id job postings (described in detail below) were opened (to the nearest day).[2] We also scraped the text and bolded formatting of the U.S. job postings so that we could record (to the nearest day) when recruiters or hiring managers made edits to the qualifications section.

The next step of the experiment consisted of manually reading the job postings that had been opened that week and determining the subsample to include in the experiment. We were constrained by both the sheer volume of job postings opened each week (75-200), as well as heterogeneity in treatment (e.g., some job postings did not have as many or any margins to execute the treatment). Each week a batch of new job postings was then included in the experiment. There was typically at least a week delay between a job posting being opened on the careers website and then included in our experiment. This was intentional since we anticipated that most recruiter edits would come shortly after opening the job, and we wanted to initialize the job posting into the A/B test after those edits had been reflected.

[2] We scraped both U.S. and international job postings. For the U.S. job postings, we scraped several different subpages of the careers website for comprehensiveness (e.g., subpages listing U.S. job postings, subpages listing U.S. job postings by city, and subpages listing team assignment for U.S. job postings). This was done because job postings were sometimes initially listed without the appropriate tags, causing them to not show up on a particular list. We did not scrape by U.S. subteam page (one level below team assignment), but we do not anticipate that this materially impacts the number of job postings we record that opened or expired.

We also determined which job postings included in our experiment had been subsequently edited by recruiters or hiring managers, whether any long-digit-id job postings had been opened which had a corresponding 5-digit-id version in our experiment (discussed below), and whether any job postings in our experiment had expired (i.e., had been removed from the website by the recruiter/hiring manager); these job postings were then tagged on a weekly basis for removal from the experiment.[3] There was typically at least a week delay between being tagged and actually being removed from the experiment. Note that recruiters and hiring managers could not view our treatment to the job postings from the internal system they use.

Finally, we recorded the specifics of the treatment as the experiment was running, and reflected all of the execution details (a job posting's initialization into the experiment, recruiter edits, and removal date) on our treatment tracker, which was continually uploaded throughout the experiment's duration to the AEA RCT registry.

D. Potential Sources of Slippage

There is the potential for slippage in our experiment. We discuss each below. In general, slippage would tend to dampen the treatment effect.

(i.) Slippage in Assignment to Treatment. The first type of slippage represents inconsistent assignment of a viewer to Treatment status over the experiment's duration (e.g., a unique viewer being assigned to the Treated group at one point in time and the Control group at a different point in time). While this type of slippage does not introduce bias in our regressions, seeing differing language might have been confusing for viewers if they saw different versions for the same job posting.

We believe any confusion is mitigated by several factors: (1) not all individuals

[3]We record the day that a job posting was edited, and we use the latest date that the job posting was not edited (typically the day before) as the date cutoff for including applications in our analysis. Less than 10 percent of job postings were edited by recruiters / hiring managers after they were included in our experiment. Some job postings expired and re-opened. We did not re-include a job posting that had been expired and removed from the experiment if it was later re-opened. There are some job postings in the experiment that expired and re-opened before they could be removed, and so were kept in the experiment until their next expiration.

are likely to remember the exact specifics of any given job posting; (2) there is already heterogeneity in Uber's job postings, so individuals are already exposed to a range of language between viewing the job postings included in our experiment and those that are not; and (3) the relative ease and shortness of the application process are likely to make listed qualifications most salient at the point of application.

The only way this type of slippage would occur is if the viewer cleared the local storage on his or her device or if the viewer accessed Uber's careers website from more than one browser or device over the duration of the experiment. For the subsample of individuals that applied to more than one job posting in our experiment (847), only 46, or approximately 5 percent, received both a Treated and Control tag over the duration of the experiment; while this is a selected subsample of only those who chose to apply, the relative infrequency of multiple assignment to Treated and Control groups is reassuring.

(ii.) Slippage from Accessing a Third-Party Website. The second type of slippage comes from accessing a third-party website (e.g., LinkedIn, Glassdoor) over the experiment duration. All Uber job postings on third-party websites showed the Control version of the job advertisement, so accessing a third-party website would only present inconsistent versions of the job postings for a viewer randomized into the Treated group (i.e., the slippage goes in one direction).[4] We discuss both contemporaneous and delayed viewing of a third-party website below.

Contemporaneous Viewing. This type of slippage only occurred for viewers who *first* visited Uber's careers website and *then* visited a third-party website in another tab or window *during the same session*.[5] For the viewers that chose to apply through the careers website, we can eliminate their applications from our analysis (because we can detect the presence of the inherited third-party website cookie); the relatively small number of these

[4] Applying via a third-party website involves clicking a link on the third-party website that redirects applicants to Uber's careers website. Once on Uber's careers website, these applicants follow the same application process, but are tagged as coming from a third-party website.

[5] Note that viewers who *first* visited the third-party website and *then* visited Uber's careers website in another tab or window of the same browser-device during the same session are automatically excluded from randomization in our experiment.

applications (92 in the Control group and 73 in the Treated group) is reassuring. We cannot track the viewers in this group that chose not to apply through the careers website (i.e., those who chose not to apply at all or chose to apply through the third-party website process). However, in order to bias upward our estimates of the treatment effect, we would have to assume that seeing different versions of the job posting for those in the Treated group would cause confusion that disproportionately induces these viewers to apply through the careers website; while some degree of confusion is possible, it is unlikely that we would observe a significant 7 percent increase in Treated applications or our observed patterns in application attributes due solely to confusion.

Delayed Viewing. This type of slippage occurs for all viewers that visit a third-party website at any other point in time (either before or after being randomized into the Treated or Control group on the careers website). We cannot observe the number of these viewers in our experiment. While it possible that there might be confusion due to seeing both Treated and Control versions of the job posting for viewers randomized into the Treated group, in order to upwardly bias our estimates of the treatment effect, we would again have to assume that this confusion disproportionately causes Treated viewers to apply through the careers website, which seems unlikely. We also believe that this type of slippage is mitigated by the fact that not all individuals are likely to remember the exact specifics of any given job posting – especially if there is a significant time delay between viewing the job posting on the third-party website and the careers website – and the fact that listed qualifications are likely to be most salient at the point of application.

(iii.) Slippage from Delayed Updating of Website with Treatment. Slippage could also occur due to the split second delay that sometimes occurs when the webpage content is loading in a given browser. For viewers in the Treated group, the Control version of a job would have been displayed for a split second before the Treated version appeared in its place. We believe that this issue is mitigated by the fact that viewers are accustomed to the dynamic nature of websites, and that updates to the "What You'll Need" section occur below the fold (i.e., to the lower half of the webpage which individuals must scroll down to view). Moreover, this slippage would tend to dampen our estimates of the treatment, as only the Treated group would have experienced slippage in viewing.

(iv.) Slippage from Delayed Refreshing of Website. Slippage could also be present due to delayed refreshing of the careers website and imperfect information about when an individual views a job. To understand this scenario, consider a viewer assigned to the Treated group who saw the Control version of a job posting in our experiment minutes before the job posting was entered into the A/B test, and then applied just after the A/B test for the job posting had been initiated. If the website did not refresh between viewing and applying for the job posting, then the applicant would be counted as Treated, but would never have viewed the Treated version of the job.[6] To account for this slippage, we perform sensitivity analyses that exclude applications which came in at varying intervals after the A/B test was initiated (see Appendix Document A.3). To the extent that this slippage is present, it would only downward bias our estimates of the treatment.

(v.) Slippage from Long-Digit-Id Job Postings. There is also the potential for slippage due to the presence of long-digit-id job postings on the careers website, representing approximately 10% of all U.S. job postings. These long-digit-id job postings are effective "duplicates" of an original job posting on Uber's website (i.e., the posting reflects the exact same language but is typically for an opening in a different city) and is usually listed right below or above the original job posting. We were not able to include any of these long-digit-id jobs in the experiment due to technical constraints. As a result, when a 5-digit-id job posting was included in the experiment, and a long-digit-id job posting was subsequently opened on the website, we removed the original 5-digit-id job posting from the experiment. This was done to minimize confusion by not having Treated and Control versions of effectively the same job posting in close proximity on the careers website.

These long-digit-id job postings also exhibited a particular technical issue (independent of our experiment): when a user clicked on a 5-digit-id job posting, hit the back button,

[6] Refreshes to the website only occur when an individual manually refreshes the website or clicks on something that takes them away from that page before they revisit the webpage.

and then clicked on a long-digit-id job posting, the long-digit-id job displayed the cached Control version of the 5-digit-id job posting just viewed. This technical glitch could cause confusion for Treated viewers, who would be exposed to the Control version of the 5-digit-id job posting after having just seen the Treated version. However, none of the Treated or Control applications from this process would be included in the experiment (as the relevant treatment tag cookie was erased when a viewer initially clicked on the long-digit-id).[7] The exception to this rule concerns viewers who inherited a treatment tag from another tab in the same browser-device-session; in this case, the application (from both Treated or Control viewers) would flow into our experimental data.[8] Given the very detailed process that would need to occur here, we think it is unlikely that significant slippage from this issue is occurring. Moreover, the slippage would only dampen the treatment effect.

(vi.) Multiple Tab Issue. Since we assigned treatment status through the use of cookies, there are some applicants that could have received the treatment status cookie if they had two or more Uber job postings open on multiple tabs or windows of the same browser during a single session (i.e., at the same moment in time). The cookie that would have been saved would reflect whatever the viewer did last on either of the open tabs/windows in the session before hitting the "Apply Now" button on the job posting.

Viewers that clicked on a long-digit-id link or a job posting not in the experiment as the last thing they did when both tabs/windows were open, and then subsequently applied for a job posting in the experiment on the other tab/window would not be included in our data (because the treatment status cookie would be removed). This would have reduced our sample size but would not cause bias as it is reasonable to assume this is happening equally between the Treated and Control groups. Conversely, if a viewer clicked on a job posting in the experiment last, they would inherit the treatment status cookie when they subsequently applied to the job posting on the other tab. We eliminate applications for jobs not in our

[7] Note that hitting the "see job post" link on the upper left corner of the interstitial page would take viewers to their original Treated version of the 5-digit job posting-job posting and the treatment tag cookies would be re-added properly.

[8] See *Multiple Tab Issue* discussed below.

experiment or before the A/B test was initialized.

The only way for the multiple tab issue to bias our results is if this issue simultaneously interacted with the long-digit-id issue (detailed above).

A.3 Technical Details and Robustness Checks

There were times over the course of the experiment where there was a technical issue. We document these issues for comprehensiveness, but none of these significantly affect the conclusions from our experiment. We also describe additional robustness checks that confirm our results.

A. Experiment Glitch In September 2018

In mid-September 2018, the experiment stopped working for approximately three days. The experiment stopped working because of two updates that were made to the A/B test code when we were initializing a new set of job postings. The first update occurred on September 11, 2018 at 5:49pm PST and the second update occurred on September 12, 2018 at 8:47am PST. The experiment was back online and working properly on September 14, 2018 at 2:53pm PST.

To account for this technical issue, we deleted applications that came in between September 11, 2018 at 5:49pm PST and September 14, 2018 at 2:53pm PST. As a robustness check, we instead deleted applications that came in for a shorter interval, between September 12, 2018 at 8:47am PST and September 14, 2018 at 2:53pm PST; using the shorter interval does not materially affect our results.

B. Experiment Glitch In October 2018

On October 31, 2018 at 4:30pm PST, the experiment stopped working due to a macro change to the Optimizely system. The experiment was back online and working properly on November 7, 2018 at 2:20pm PST. To account for this technical issue, we deleted applications that came in during this time period.

C. Experiment Glitches In November 2018

Throughout November 2018, the experiment experienced glitches on different browsers, whereby the cookies would populate but the Treated version of the job posting would not show up for Treated viewers. This issue was primarily attributable to browser-specific issues interacting with server demand for the software we used to configure the test implementation, and browser-specific issues interacting with changes to Uber's careers website.

We made three discrete changes to the experiment over this time period to account for these issues: 1) On November 8, 2018 3:30pm PST, we removed Microsoft Edge from the experiment; 2) on November 16, 2018 7:02am PST, we changed the way our experimental cookies were caching; and 3) on November 26, 2018 12:43pm PST, we removed all browsers from the experiment except for Chrome and Safari (the two most heavily used browsers).

We use these discrete changes as additional robustness checks by deleting applications from the experiment breakdown on October 31, 2018 at 4:30pm PST to each of the aforementioned dates. Using these three end dates instead of the November 7, 2018 2:20pm PST end date does not materially affect our results.

D. Additional Robustness Checks

In this section, we describe robustness checks related to implementation.

i. September 11 Correct Removal of Job Postings. On September 11, 2018, there were 25 job postings that had expired/been edited and were intended for removal, but due to a technical error were not deleted from the experiment. Several of these expired jobs subsequently reopened on the Uber website. Our base specification records these job postings as removed from the experiment on September 11, 2018 at 5:49pm PST, but, as a robustness check, we include applications that came in after September 11 from the reopened job postings. This robustness check does not materially affect our results.

ii. Job Postings Artificially Expired. Over the course of the experiment, there were 22 job postings that appeared as if they expired in our daily scraping, but were in fact still alive. For these "artificially expired" job postings, we manually checked that no edits had been made to the original qualifications section (so that there was no discrepancy between

Treatment and Control). However, given that there was often a week plus lag between the first indication of artificial expiry and when we manually checked them, we conduct a robustness check that eliminates applications for these job postings which accrued after the last date the job posting was able to be scraped and matched to its original Control version. This robustness check does not materially affect our results.

iii. Adjustments for Time Difference in Scraping. We scraped the text of the job postings every day over the course of our experiment duration to determine if a recruiter or hiring manager had edited any job postings. Scraping all of the postings on the Uber careers website took approximately two hours, so we recorded the time that the scraping started as the last date and time the posting matched, since it is theoretically possible that a job posting was edited over the two hours that the scraping occurred. However, between August 14, 2018 (the experiment start time) and August 25, 2018, we recorded the scraping end time instead of the start time. As a result, for job postings that were edited in this time period, we adjust the latest date that the job posting matched by two hours to account for this issue.

iv. Precise Time of A/B Test. Each time we entered a new set of job postings into the experiment, we recorded the precise time that the A/B Test was initiated for the job postings. The exception to this were the August 16, 2018 and August 22, 2018 job entry batches in which the exact minute was not recorded (the time delay was by a few hours). However, there were no applications that are affected by this, so it is not an issue for our analysis.

As discussed in Appendix Document A.2, given non-instantaneous refreshing of the website, there is the potential that some individuals assigned to the Treated group saw the Control version of the job posting. For example, if an individual assigned to the Treated group viewed the Control version of the job posting at 11:59am PST, the A/B test for the job went into effect at 12:03pm PST, and the individual applied at 12:04pm PST, the individual would have been counted in the experiment even though s/he never actually saw the Treated version of the job. Since we cannot precisely know when an individual viewed a job relative to when s/he applied, we remove applications that came in 30 minutes after the A/B test started (30 minutes is also the standard definition for a "session" length in Google Analytics). We conduct a robustness check with a shorter time period (15 instead of 30 minutes), and the results of our analyses are unaffected.

E. Formatting Changes

Due to technical constraints, between September 13, 2018 and October 2, 2018 at 9:00am PST, the bolding of the "What You'll Need" title (or analogous title) for the qualifications section did not match between the Treated and Control versions for some of the job postings.[9] The bolding inconsistency between Treated and Control titles was fixed on October 2, 2018 at 9:01am PST, though there were still some job postings even after this fix that could not be adjusted due to technical issues; for these edge cases, the "What You'll Need" title would default to the unbolded version, in line with our treatment which de-emphasizes the qualifications. We do not believe bolding discrepancies to the title materially affected our results.

There were some job postings for which the swapped in Treated version of the qualifications section appeared with a slightly different formatting (e.g., font, bullet indentation) relative to the Control qualifications text. However, given that several of Uber's job postings display inconsistent font type and indenting within a posting, this does not seem like a significant departure from what an individual might expect to see.

Two other job postings had formatting changes in the Treated version: 40893 and 43785 both had a hyperlinked website URL listed in the Control, and while the website URL was still present in the Treated version, it was not possible to be hyperlinked. We do not believe this change materially affected this job posting since the link was still listed.

[9]From September 13, 2018 through September 23, 2018, the Treated "What You'll Need" title was unbolded for some job postings (and bolded for other job postings); from September 23, 2018 at 2:06pm PST through October 1, 2018 at 2:49pm PST, the Treated "What You'll Need" title was bolded for all job postings; from October 1, 2018 at 2:50pm PST to October 2, 2018 at 9:00am PST, the Treated "What You'll Need" title was unbolded for all job postings.

F. Miscellaneous Details on Sample Limitations and Specific Job Postings

In this section, we detail miscellaneous details on specific job postings in our experiment.

i. Job Postings Accidentally Deleted. On October 1, 2018 at 2:50pm PST, there were three job postings - 41097, 41067, and 39625 - that were deleted from the experiment due to a technical issue. They were successfully put back into the experiment on October 4, 2018 at 7:18am PST.

ii. Typos in Job Postings. We manually created the Treated versions of all 616 job postings in our experiment. After the experiment concluded, we found 5 typos in the Treated version for 4 job postings: in job posting 42244, there was no space between "have" and "Javascript"; in job posting 43946, the word "handle" was mispelled as "hande"; in job posting 38002 there was an extra space between "A" and "desire" and the word "to" was repeated twice; and in job posting 37591, the word "to" was repeated twice. We do not believe these typos materially affect our results.

There were occasionally existing typos in the Control version of the job posting which we chose not to correct in the Treated version (unless it directly conflicted with the language we were editing), so as to preserve comparability between the two versions.

iii. Other Sample Limitations. Between August 13, 2018 at 1:59pm PST and August 14, 2018 at 11:30am PST, the software used in our experiment did not initially block randomization for viewers who went to a third-party website; we eliminate these applications.

As discussed in Appendix Document A.2, due to the multiple tab issue, there was the possibility of individuals accruing Treated or Control tags even if they were not part of the experiment. As such, we remove applications which were not part of our experiment (e.g., from outside the careers website, for job postings not in our experiment, and applications for those that came in before the A/B test for the job posting was initialized). We also deleted one application that absorbed both the Treated and Control tag during the same session (likely attributable to a bot that applied for both versions of the job posting during the same session).

iv. Job Postings/Applications Removed for Other Reasons Besides Expiry and Editing. There are certain job postings in our experiment that were not included or removed for reasons other than it being edited or expiring. Job posting 28620 was removed from the experiment on November 11, 2018 at 2:48pm PST because it started accruing several applications with both a Treated and Control tag. Applications from job posting 40416 are not included in our experiment because a significant fraction (3 out of 9, or 33 percent) of its applications accrued both a Treated and Control tag. Three applications from job posting 42021 were not included in our sample since we were not able to confirm that 42021's A/B test was initialized properly. Job posting 42314 was removed on September 12, 2018 at 8:47am PST (a day after it was entered into the A/B test on September 11, 2018 at 5:50pm PST) because we realized that the "Who You Are" section (outside of the "What You'll Need" section) contained especially strong language on qualifications that we could not change, and so was not in line with our treatment.

v. Job Postings that Remained in Sample Despite Expiry or Long-Digit-Id. There were 7 jobs that were included in the experiment even after initial expiration (job postings 43252, 42781, 40312, 40953, 35756, 39035, and 38774). This is because they initially expired and then re-opened before they were set to be removed.

There were 2 jobs postings (37673 and 41124) that we chose not to remove even though it had a long-digit-id open. This is because these long-digit-id postings were not tagged to a particular location and did not show up in the U.S. team or subteam pages, and so there was a low likelihood that a viewer searching in the U.S. domain would see both the original 5-digit-id version and the long-digit-id version.

A.4 Model: Robustness & Continuous Case

I. Alternative Model Formulations

i. Changing the form of the job seeker's application decision to $(p_g(s,R) \cdot n_g(s,R))V_g(s,R) - c > 0$ does not change the model's predictions. The job seeker's decision to apply is given by:

$$\gamma_g(s,R)V_g(s,R) - c = (p_g(s,R) \cdot n_g(s,R))V_g(s,R) - c > 0$$

Taking the derivative of c^* with respect to R yields:

$$\frac{\partial c^*}{\partial R}\Big|_s = \Big(\frac{\partial p_g(s,R)}{\partial R} n_g(s,R) + \frac{\partial n_g(s,R)}{\partial R} p_g(s,R)\Big) V_g(s,R) + \frac{\partial V_g(s,R)}{\partial R}\Big(p_g(s,R) \cdot n_g(s,R)\Big)$$

This is analogous to the main equation in the current model, except $\frac{\partial p_g(s,R)}{\partial R}$ is multiplied by $n_g(s,R)$, $\frac{\partial n_g(s,R)}{\partial R}$ is multiplied by $p_g(s,R)$, and $\frac{\partial V_g(s,R)}{\partial R}$ is multiplied by $p_g(s,R) \cdot n_g(s,R)$. This suggests that changes in the confidence channel from changing R are magnified when the job seeker perceives less competition, and changes in the competition channel from changing R are magnified when the job seeker is more confident about being qualified. Importantly, while the three channels of the model are each scaled by a different factor relative to the baseline model, this only affects the magnitude and not the directional effects, and so the predictions from the original model as a result of changing R remain the same.

ii. If the job seeker believes that the employer hires the best fit for the position, and so penalizes applicants who are overqualified, there would be another term in the job seeker's application decision, $\frac{\partial f_g(s,R)}{\partial R}$, which represents the job seeker's perceived probability of being a good fit for the role; including this term does not materially impact the model's predictions. The job seeker's decision to apply is given by:

$$\gamma_g(s,R)V_g(s,R) - c = (p_g(s,R) + f_g(s,R) + n_g(s,R))V_g(s,R) - c > 0$$

Taking the derivative of c^* with respect to R yields:

$$\frac{\partial c^*}{\partial R}\Big|_s = \Big(\frac{\partial p_g(s,R)}{\partial R} + \frac{\partial f_g(s,R)}{\partial R} + \frac{\partial n_g(s,R)}{\partial R}\Big)V_g(s,R) + \frac{\partial V_g(s,R)}{\partial R}\Big(p_g(s,R) + f_g(s,R) + n_g(s,R)\Big)$$

For low s, $\frac{\partial f_g(s,R)}{\partial R} < 0$, since reducing R makes the role more in-line with the job seeker's skills, and so increases the job seeker's perception of job fit. For high s, $\frac{\partial f_g(s,R)}{\partial R} > 0$, since reducing R makes the role less in-line with the job seeker's skills, and so decreases the job seeker's perception of job fit. This additional channel reflects the same directional effects as the *Job Match Quality* channel, and so does not alter the model's predictions as a result of changing R.

II. Continuous s case: Change to the Number of Applications and Gender Composition

The number of job seekers that apply for the job is given by:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{1}\{(p_g(s,R)+n_g(s,R))V_g(s,R)-c=0\}f(s)j(c)dsdc =$$
$$\int_{-\infty}^{\infty} \Phi_c\{(p_g(s,R)+n_g(s,R))V_g(s,R)\}f(s)ds =$$
$$\int_{-\infty}^{\infty} \Phi_c(c^*(s,R))f(s)ds$$

We can characterize changes to the number of applications as a result of changing R for each gender g in terms of our original continuous measure of s. The predicted change in the number of applications is a function of the relative magnitudes of the *Confidence*, *Competition*, and *Job Match Quality* channels, analogous to the 2x2 skill case:

$$\frac{\partial}{\partial R}\int_{-\infty}^{\infty} \Phi_c(c^*(s,R))f(s)ds =$$
$$\int_{-\infty}^{\infty} \phi_c(c^*(s,R))\frac{\partial c^*(s,R)}{\partial R}f(s)ds$$

where

$$\frac{\partial c^*(s,R)}{\partial R} = \Big(\frac{\partial p_g(s,R)}{\partial R}+\frac{\partial n_g(s,R)}{\partial R}\Big)V_g(s,R)+\frac{\partial V_g(s,R)}{\partial R}\Big(p_g(s,R)+n_g(s,R)\Big)$$

As can be seen in the above equation, the predicted change in applications for a given gender g increases if confidence about being qualified increases, perception of competition decreases, or perceived job match quality increases. The relative change for each gender determines how the fraction of female applications changes.

III. Continuous s case: Change to the Skill Distribution of Applications

We can also characterize changes to the skill distribution, by examining how responsiveness to changes in the probability of applying changes in s. To do this, we take the following derivative:

$$\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial s \partial R}$$

which is the continuous version of the analysis in Figure 3. Thus we have:

$$\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial s \partial R} = \frac{\partial}{\partial s}\left[\phi_c(c^*)\frac{\partial c^*}{\partial R}\right] = \phi_c'(c^*)\frac{\partial c^*}{\partial s}\frac{\partial c^*}{\partial R} + \phi_c(c^*)\frac{\partial^2 c^*}{\partial s \partial R}$$

Now, we use that $\phi_c(c) = \frac{1}{\sqrt{2\pi}\sigma_c}e^{-\frac{1}{2}\frac{(c-\mu_c)^2}{\sigma_c^2}}$ (normal pdf) to get $\phi_c'(c) = -\frac{(c-\mu_c)}{\sigma_c^3}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(c-\mu_c)^2}{\sigma_c^2}} = -\frac{1}{\sigma_c^2}(c-\mu_c)\phi_c(c)$. Substituting, we get:

$$\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial s \partial R} = \phi_c(c^*)\left[\frac{\partial^2 c^*}{\partial s \partial R} - \frac{\partial c^*}{\partial s}\frac{\partial c^*}{\partial R}\frac{1}{\sigma_c^2}(c^* - \mu_c)\right]$$

Assume that $(c^*(s,R) - \mu_c) > 0$. Note that $(c^*(s,R) - \mu_c)$ is independent on σ_c^2. In the limit, $\sigma_c^2 \to 0$. Then, the sign of $\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial R \partial s}$ is determined by the second term, and we have:

$$\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial s \partial R} \propto -\frac{\partial c^*}{\partial s}\frac{\partial c^*}{\partial R}$$

We can use the above expression to understand how changing R changes the skill distribution. Assume that $\frac{\partial c^*}{\partial s} > 0$ (higher skills increase the job application threshold). Suppose that $\frac{\partial c^*}{\partial R} < 0$, so that higher requirements lower the job application threshold. Then, we have $\frac{\partial^2 \Phi_c(c^*(s,R))}{\partial s \partial R} > 0$. This means that $\frac{\partial \Phi_c(c^*(s,R))}{\partial R}$ is higher at higher s. Since $\frac{\partial c^*}{\partial R}$ was assumed to be negative, this means that lowering requirements has more of an impact on bringing in low skilled applicants than high skilled ones, in line with the predictions from our 2x2 skill s case. Note that this result could be reversed if we make different assumptions (e.g., $(c^*(s,R) - \mu_c) < 0$ or $\frac{\partial c^*}{\partial s} < 0$).

A.5 Power Calculation

The power calculations for changes in the fraction of female applications are based off of the individual-level regression:

$$Female_i = \alpha + \beta Treated_i + \epsilon_i$$

Given that the variable *Female* takes a value of either 0 or 1, *Female* is a Bernoulli random variable with mean p, where p is the fraction of female applications in the sample. The standard deviation of *Female* is the standard deviation of a Bernoulli random variable with mean p: $\sqrt{p(1-p)}$.

Since we know the mean and standard deviation of *Female* for both the Treated and Control groups, we can bound the percentage point change in the fraction of female applications between Treated and Control given the number of applications in our experiment. More specifically, given a significance level of α=0.5, a power level of 80%, 0.382 fraction female applications in the Control group, 0.377 fraction female applications in the Treated Group, 2,996 applications in the Control group, and 3,189 applications in the Treated group, we know that the change in fraction female applications between Treated and Control groups can be bound under approximately 3.5 percentage points.[10] On an individual applicant basis, the change in the fraction of female applicants between Treated and Control groups can be bound under approximately 4 percentage points.

We use a similar procedure to bound the null effects at the subsequent hiring stages.

[10]These calculations are obtained using Stata's *power twomeans* command as well as the standard formula, $(Z_{power} + Z_{\alpha/2}) * standarderror(teststatistic)$, where Z_{power}=0.84 (corresponding to 80% power), $Z_{\alpha/2}$=1.96, and $standarderror(teststatistic) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$.

A.6 Resume Analysis Definitions

I. Educational Attainment

The below lists the words we used to categorize educational attainment. We include variants of the acronyms listed below to account for differences in punctuation and capitalization in the resume text. We use the *Base Definition* in our analyses but construct all 81 combinations using the below definitions; when we conduct robustness to Table 1.6 and Appendix Table A.3 using these alternative definitions, our results remain the same.

i. Bachelors

a. Base Definition: Bachelor(s), Bachelor (e.g., A.B., S.B.), Bachelor of Arts (e.g., B.A.), Bachelor of Fine Arts (e.g., B.F.A.), Bachelor of Business Management (e.g., B.B.M.), Bachelor of Business Administration (e.g., B.B.A.), Bachelor of Science (e.g., B.S., B.Sc., Sc.B.), Bachelor of Engineering (e.g., B.E., B.Eng), Bachelor of Science in Engineering (e.g., B.S.E.), Bachelor of Computer Science (e.g., B.CompSc, B.C.S.), Bachelor of Applied Science (e.g., B.AS., BSAS, B.ASc., BAppSc)

b. Narrow Definition: Bachelor(s), Bachelor (e.g., A.B., S.B.), Bachelor of Arts (e.g., B.A.), Bachelor of Science (e.g., B.S., B.Sc., Sc.B.)

c. Broad Definition: Base Definition (above) plus proxies for expected graduation (e.g., Expected Graduation, College, University)

ii. Masters

a. Base Definition: Master(s), Master of Arts (e.g., M.A., A.M.), Master of Fine Arts (e.g., M.F.A.), Master of Philosophy (e.g., M.Phil), Master of Science (M.S., M.Sc., S.M., Sc.M, M.Sci), Master of Engineering (M.Eng), Master of Computer Science (M.C.S.), Master of Chemistry (M.Chem), Master of Math (M.Math), Master of Physics (M.Phys), Master of Psychology (M.Psych), Master of Public Administration (M.P.A.), Master of Public Policy (M.P.P.)

b. Narrow Definition: Master(s), Master of Arts (e.g., M.A., A.M.), Master of Philosophy (e.g., M.Phil), Master of Science (M.S., M.Sc., S.M., Sc.M)

c. Broad Definition: Base Definition (above) plus miscellaneous abbreviations (M.E., ME for Master of Engineering; MS for Master of Science; MA for Master of Arts), and *graduate student*.

iii. MBA

a. Base Definition: Master of Business Administration, M.B.A.

iv. JD

a. Base Definition: J.D., Juris Doctor, Doctor of Jurisprudence, L.L.M., S.J.D.

b. Narrow Definition: J.D., Juris Doctor, Doctor of Jurisprudence

c. Broad Definition: Base definition (above) plus *Attorney*

v. PhD

a. Base Definition: PhD, DPhil, Doctor of Science

b. Narrow Definition: PhD

c. Broad Definition: Base definition (above) plus *Doctorate*, Sc.D., D.Sc., D.B.A., and *graduate student*.

II. Educational Degrees

The below lists the words we used to categorize educational degrees. We include variants of these to account for differences in capitalization in the resume text. We also classify degrees if the relevant information is indicated in educational attainment (e.g., Bachelor of Science would be categorized as STEM).

a. STEM: Aeronautical, Algorithm, Artificial Intelligence, Biology, Biophysics, Computer, Computer Science, CompSci, Chemistry, Deploy, Engineering, Information Science, Machine Learning, Math, Mathematics, Neuroscience, Signal Processing, Statistics, STEM, Physics *b.* *Social Science:* Econ, Economics, Linguistics, Organizational Behavior, Organizational Development, Psychology, Psych, Social Science, Sociology

c. *Business, Finance, or Operations:* Accounting, Administration, Business, Business Operations, Commerce, Finance, Information Science, Information Systems, Information Technology, Logistics, Manufacturing, Operations, Supply Chain, Taxation

d. *Architecture, Environment, Urban Planning, or Geography:* Architecture, Environment, Geography, Geosciences, Geo Sciences, Geo, Urban, Urban Planning

e. *Advertising, Communication, Design, HR, or Marketing:* Advertising, Communication(s), Design, Human Resource(s), Marketing, Media

III. Educational Institutions

The below lists the words we used to classify an elite educational institution. We include variants of these to account for differences in capitalization in the resume text. All institutions are from 2020 U.S. News & World Report Rankings. *Elite Schools* reflects the top 6 by rank plus ivy league institutions. *Elite U.S. Business Schools* reflects the top 10 business schools by rank.

a. *Elite Schools:* Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, Massachusetts Institute of Technology, Princeton University, Stanford University, University of Chicago, University of Pennsylvania, Wharton, Yale University

b. *Elite Business Schools:* Booth School of Business, Columbia Business School, Fuqua School of Business, Haas School of Business, Harvard Business School, Kellogg School of Management, Ross School of Business, Sloan School of Management, Stanford Graduate School of Business, University of Pennsylvania, Wharton, Yale School of Management

IV. Technical Skills

The below lists the words we used to categorize top, broad, and general technical skills. (We use *RStudio* and *R Shiny* as proxies for *R*.) *Top* refers to listing one of the most commonly listed technical skills in our sample's job postings, *Broad* refers to listing one of a broad set of over 40 technical skills sourced from our sample's job postings, and *Language* refers to listing technically inclined language sourced from our sample's job postings.

a. Top: C++, C#, Golang, Java, Python, RStudio, R Shiny, SQL

b. Broad: Bash, C++, C#, CSS, Elixir, Golang, Git, Hadoop, HDFS, Hive, HTML, Java, Javascript, Kotlin, Linux, Matlab, NodeJS, Node, Objective C, Oracle, Perl, Pig, PHP, Presto, Python, RStudio, R Shiny, Ruby, Rust, Scala, SAS, Spark, SPSS, STATA, SQL, MySQL, NOSQL, pgSQL, PostgreSQL, Swift, TypeScript, VB .NET, Visual Basic .NET, Visual Basic

c. Language: algorithm(s), algorithmic, Android, iOS, web development, programming, machine learning, modeling, data analysis, data analytics, quantitative analysis, big data, data science

V. Work Experience

The below lists the words we used to categorize elite tech, elite finance, consulting, or accounting, and elite law firms. An elite tech company is defined as one of the top 10 internet/social media companies, one of the top 8 hardware/equipment companies, or one of the top 10 software/computer services companies in *Vault's* 2019 rankings, supplemented by a list of additional elite tech firms (the additional firms are sourced from a 2019 *Hired* survey of approximately 4,000 tech workers asking about prestigious tech companies.[11] An elite finance, consulting, or accounting company is defined as one of the 3 most prestigious finance companies, 3 most prestigious consulting companies, or 4 most prestigious accounting companies in *Vault's* 2020 rankings. An elite law firm is defined as one of 10 most prestigious law firms to work for in *Vault's* 2020 rankings. All the above firms from *Vault's* lists are chosen because they have a score of 7.5 or higher (out of 10). We include

[11]The survey link is here: https://hired.com/page/brand-health-report/top-global-employer-brands. More information about the survey is here: https://www.businessinsider.com/best-tech-companies-to-work-at-hired-google-netflix-apple-2019-9. We included the top 15 companies in this list.

variants of company names to account for differences in capitalization in the resume text.

a. Elite Tech: Adobe, AirBnB, Amazon, Apple, Avaya, CA Technologies, Cisco, Dell, Dropbox, Ebay, Electronic Arts, Facebook, Google (Alphabet), Hewlett Packard, Hulu, IBM, Intel, Intuit, Instagram, Lyft, Netflix, Oracle, Qualcomm, Salesforce, SAP, Slack, Snap, SpaceX, Symantec, Tesla, Twitter, Yelp, YouTube (Loon), Xerox

b. Elite Finance, Consulting, or Accounting: Bain, Boston Consulting Group, Deloitte, Ernst & Young, Goldman Sachs, J.P. Morgan, KPMG, McKinsey, Morgan Stanley, PriceWaterhouseCoopers

c. Elite Law: Cravath, Swaine & Moore, Davis Polk & Wardwell, Gibson, Dunn & Crutcher, Kirkland & Ellis, Latham & Watkins, Paul, Weiss, Rifkind, Wharton & Garrison, Simpson Thacher & Bartlett, Skadden, Arps, Slate, Meagher & Flom, Sullivan & Cromwell, Wachtell, Lipton, Rosen & Katz

VI. Social Skills

The below lists the words we used to categorize management / leadership skills, communication skills, teamwork / interpersonal skills, and having proactive, self-promoting or praising language. The words are sourced from our sample's job postings.

a. Management / leadership skills: leadership, leader, leading, lead, led, manager, managerial, manage, managing, managed, management, oversee, overseeing, oversaw, oversight, supervisor, supervise, supervising, supervised, supervision, supervisory

b. Communication skills: communicator, communication, communicate, communicating, communicated, public speaker, public speaking

c. Teamwork / interpersonal skills: collaborator, collaborate, collaborating, collaborated, collaboration, collaborative, collaboratively, cooperate, cooperating, cooperated, cooperation, interpersonal, teamwork, team(s)

d. Proactive, self-promoting, praising language: accomplish, accomplished, accomplishing, achieve, achieved, achieving, challenge, challenged, challenging, confidence, confident, confidently, creative, creatively, deep, demonstrated, dynamic, driven, engage, engaged, engaging, enthusiasm, enthusiastic, entrepreneurial, excellent, excellence, exceptional, effective, effectively, expertise, expert, expertly, extraordinary, highly, initiative, innovative, insightful, motivate, motivational, motivated, obsess, obsessed, obsessive, outstanding, passion, passionate, passionately, persuasive, proactive, proactive, proactively, proven, self starter, skilled, solid, strong, strongly, successful, successfully, superb, superior

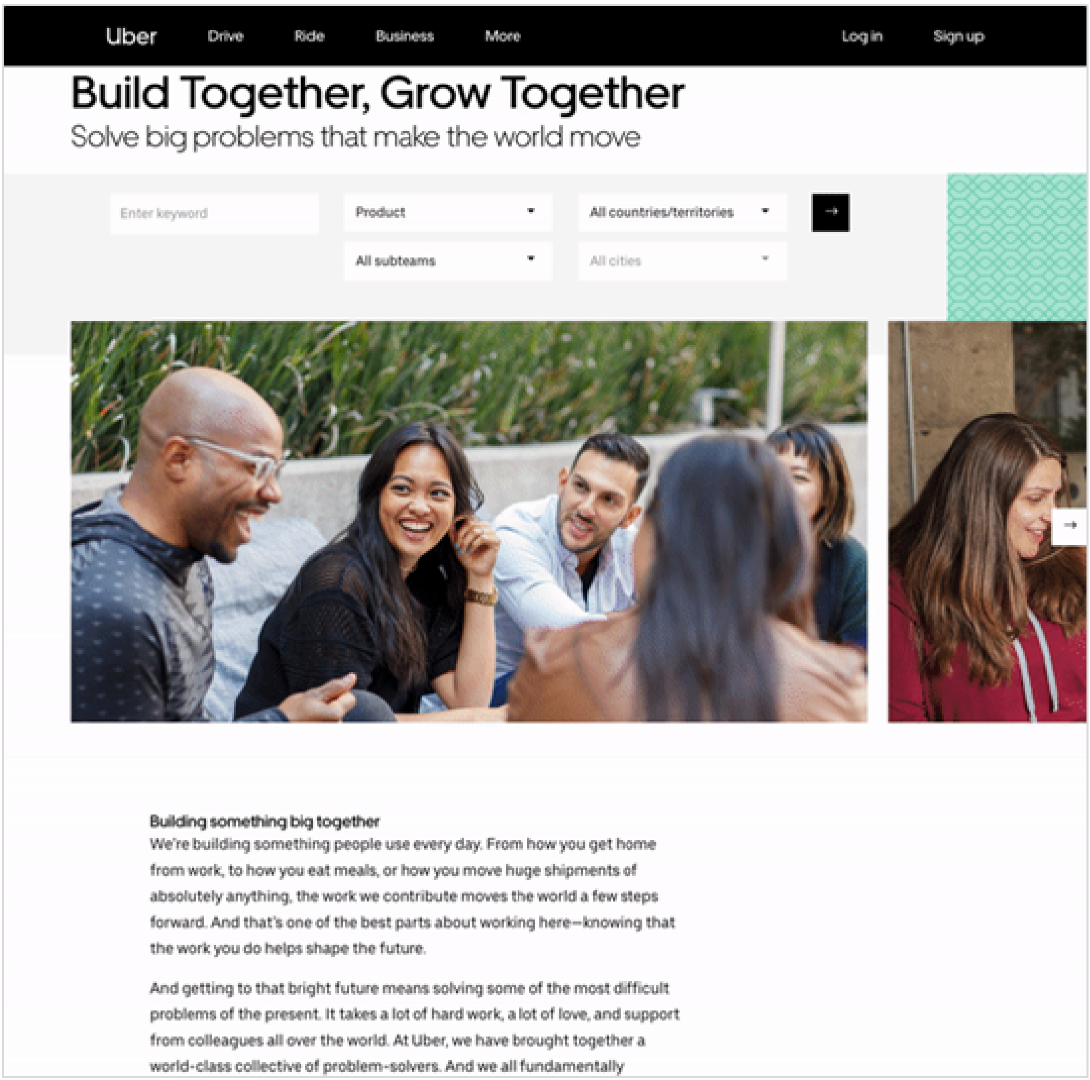
VII. Language

The below lists the languages we categorize as a foreign language. An application is classified as listing a foreign language if at least one of the 33 below are listed in the resume text.

a. Languages: Arabic, Bengali, Cantonese, Chinese, Danish, Dutch, Farsi, Finnish, French, German, Greek, Hebrew, Hindi, Hindustani, Icelandic, Indonesian, Italian, Japanese, Korean, Latin, Malay, Malaysian, Mandarin, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Thai, Turkish, Ukranian, Vietnamese

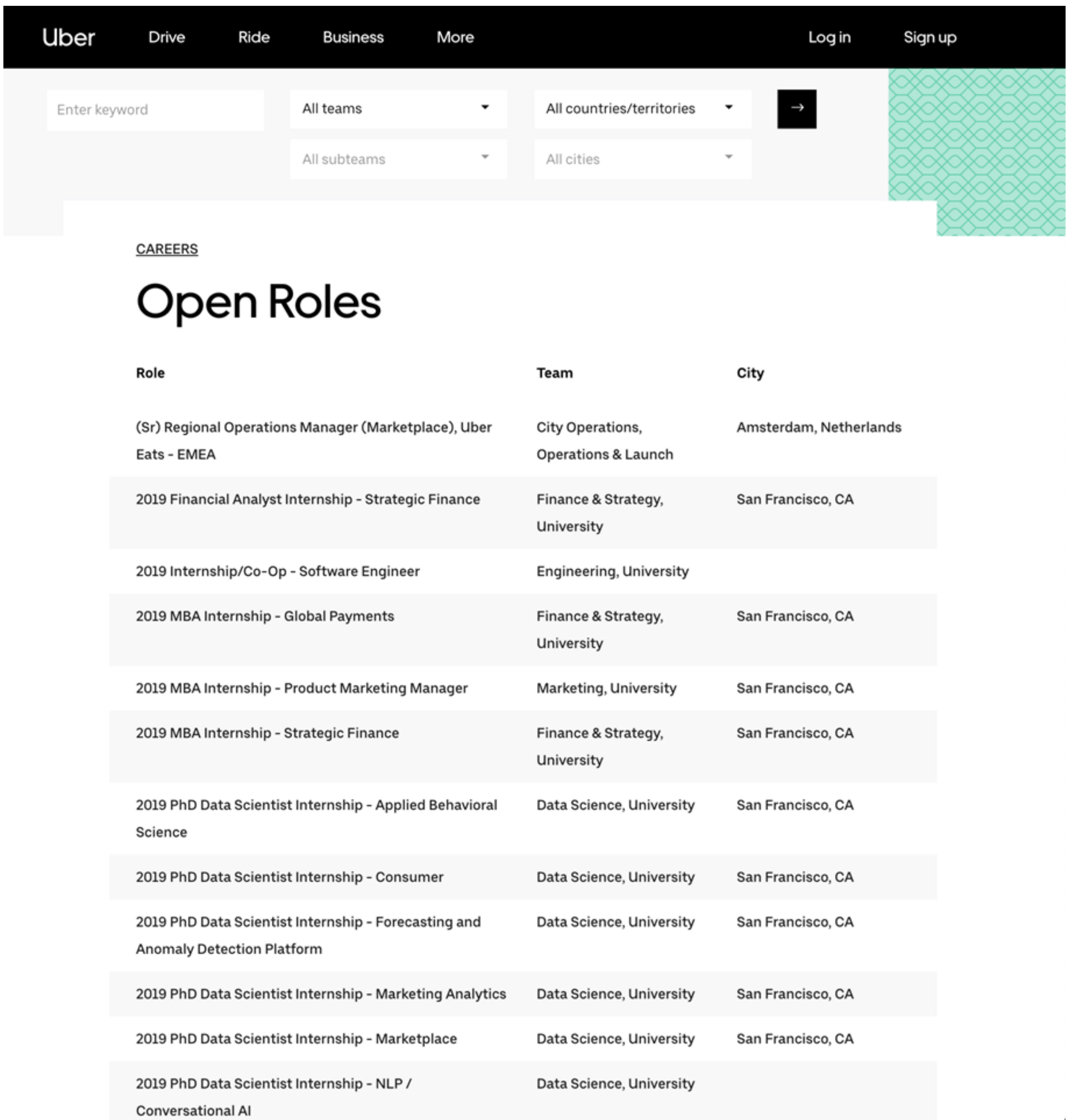
A.7 Supplementary Tables and Figures

Figure A.1: *Job Search on Uber Careers Website*



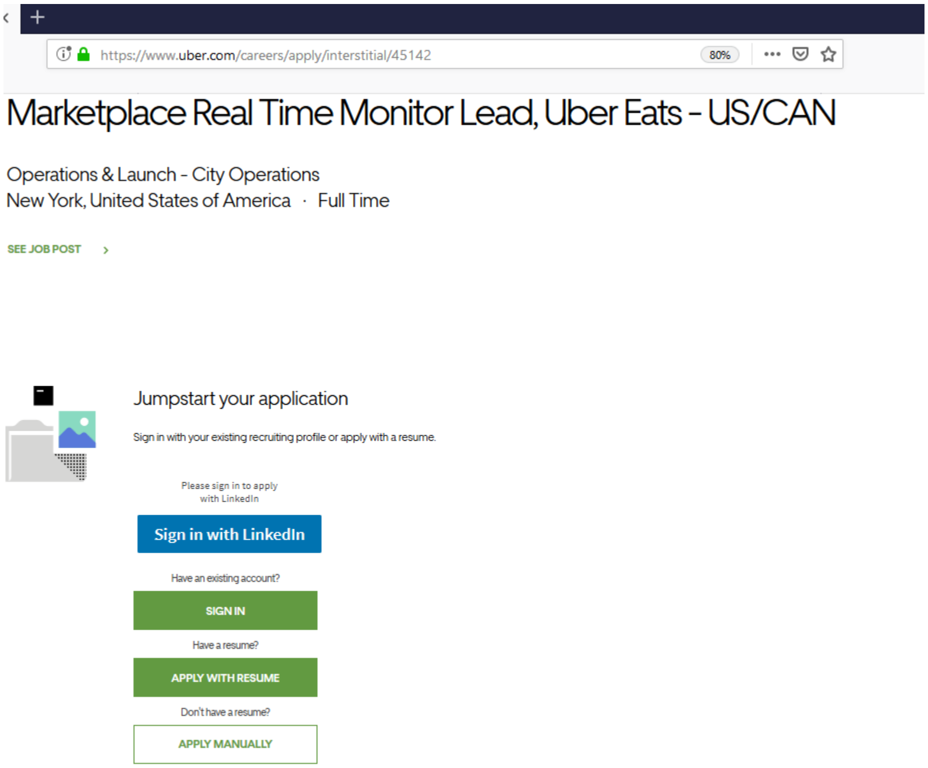
Notes: This figure depicts the initial main page of the Uber careers website over the duration of our experiment. Viewers had the ability to search by keyword, Uber team, subteam, and location (country and/or city).

Figure A.1: *Job Search on Uber Careers Website (Continued)*



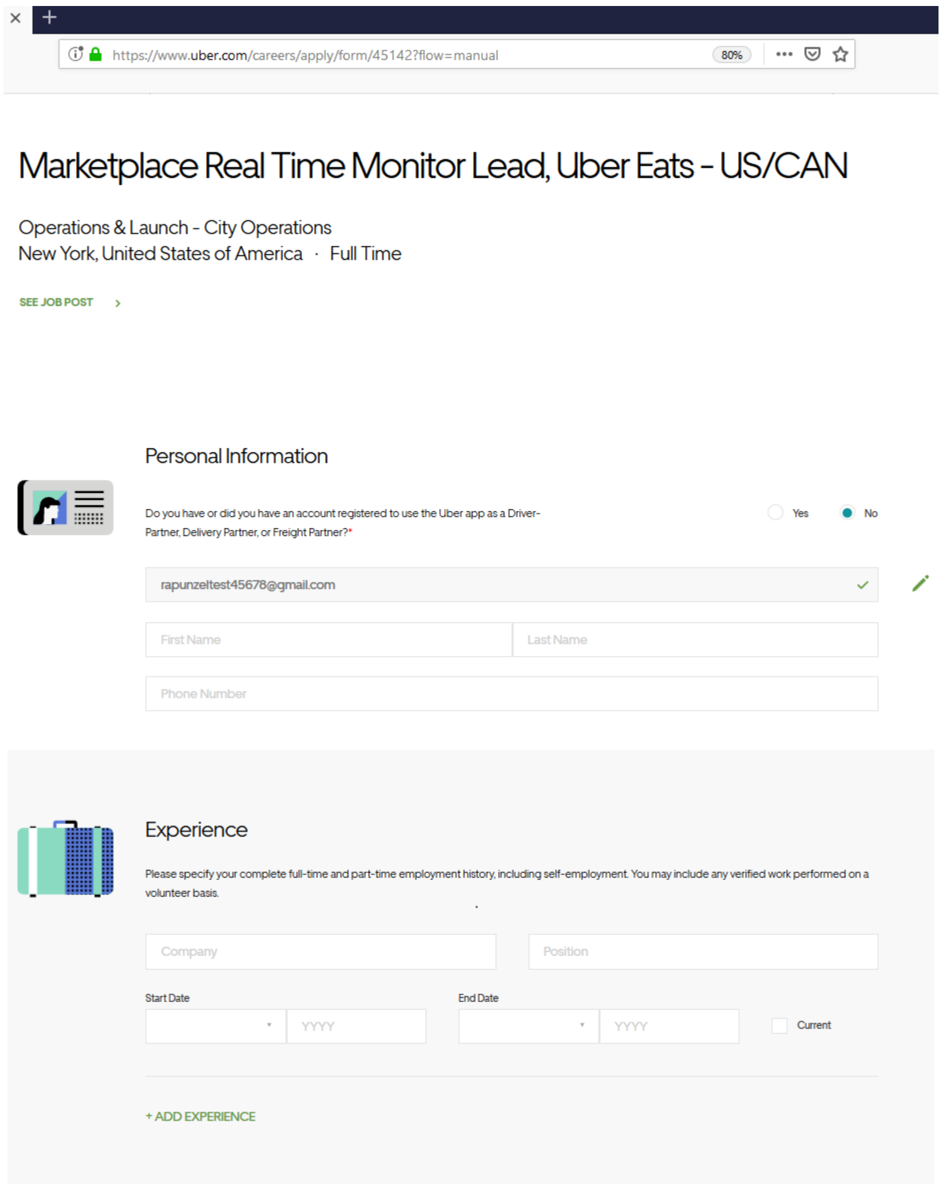
Notes: This figure depicts the list of jobs conditional on search filters on the Uber careers website; depending on the length of the list, job postings continued onto subsequent pages. Nothing about job search changed as a result of our experiment; in particular, searching by keyword (e.g. "PhD preferred") did not result in differential search results for Treated and Control viewers even if certain words were deleted in the Treated version of the job postings.

Figure A.2: *Uber Application Process*



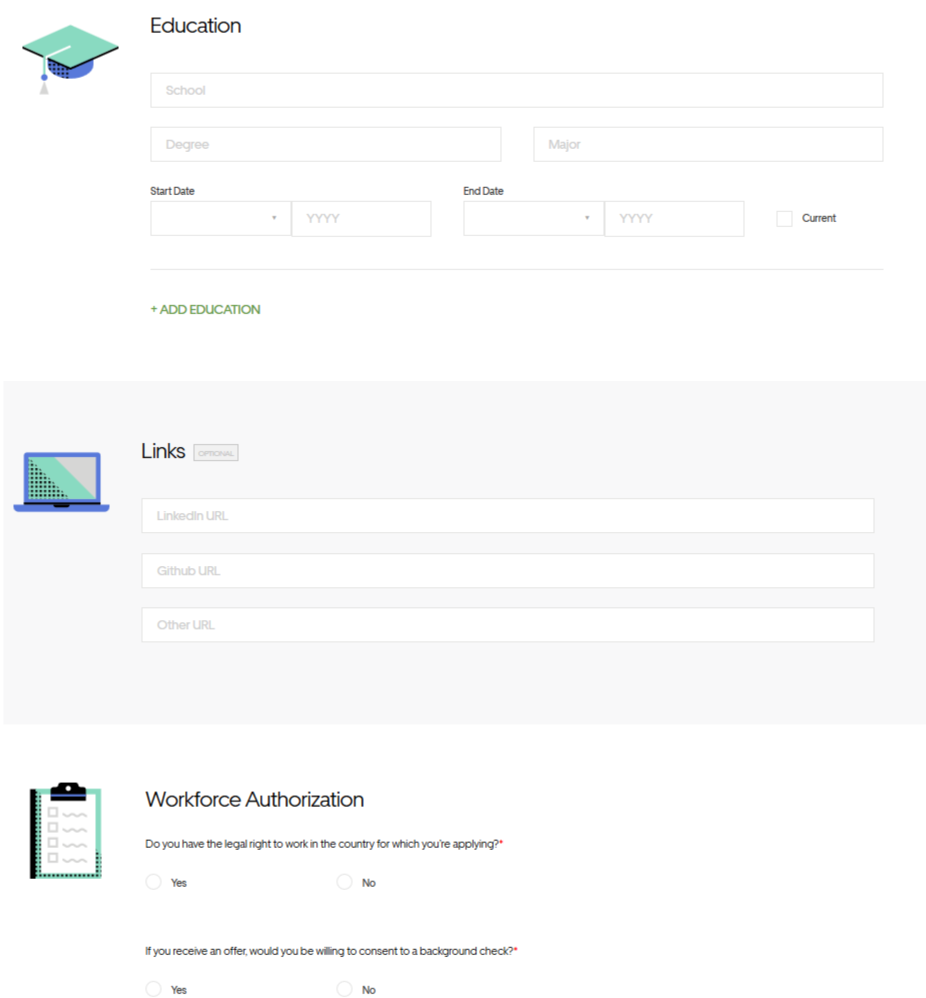
Notes: This figure depicts the interstitial page that candidates reach upon clicking the "Apply Now" button on any given job posting. There are four methods for applying: "LinkedIn", "SignIn", "Apply with Resume", and "Apply Manually". Applying with LinkedIn imports the candidates LinkedIn data into the application form, applying manually asks the candidate to fill out the application form directly, and applying with resume imports the data from the candidate's resume into the application form. Applying with the SignIn option pulls the latest existing information in the Uber system (entered from a previous application) and so could reflect any of the three application methods. While there is variation in the information stored in Uber's recruiting database as a result of these different methods, controlling for application method in our analyses does not materially change our results.

Figure A.3: *Uber Application Form*



Notes: This set of figures depicts the application form. Candidates are asked about work experience, education, additional links (e.g., LinkedIn, GitHub), workforce authorization, and demographic information. Sections with a red asterisk are required, and basic personal, education, and experience information must also be populated before submitting the application.

Figure A.3: *Uber Application Form (Continued)*



Notes: This set of figures depicts the application form. Candidates are asked about work experience, education, additional links (e.g., LinkedIn, GitHub), workforce authorization, and demographic information. Sections with a red asterisk are required, and basic personal, education, and experience information must also be populated before submitting the application.

Figure A.3: *Uber Application Form (Continued)*

Demographic Information

Uber is an equal opportunity employer. These questions are optional, but if answered will help us determine whether our efforts are consistent with equal employment opportunity best practices. Responses to these questions are completely voluntary and are not used in the hiring decision. Learn more about Equal Employment Opportunity.

Sex*

- Male
- Female
- Decline to Self Identify

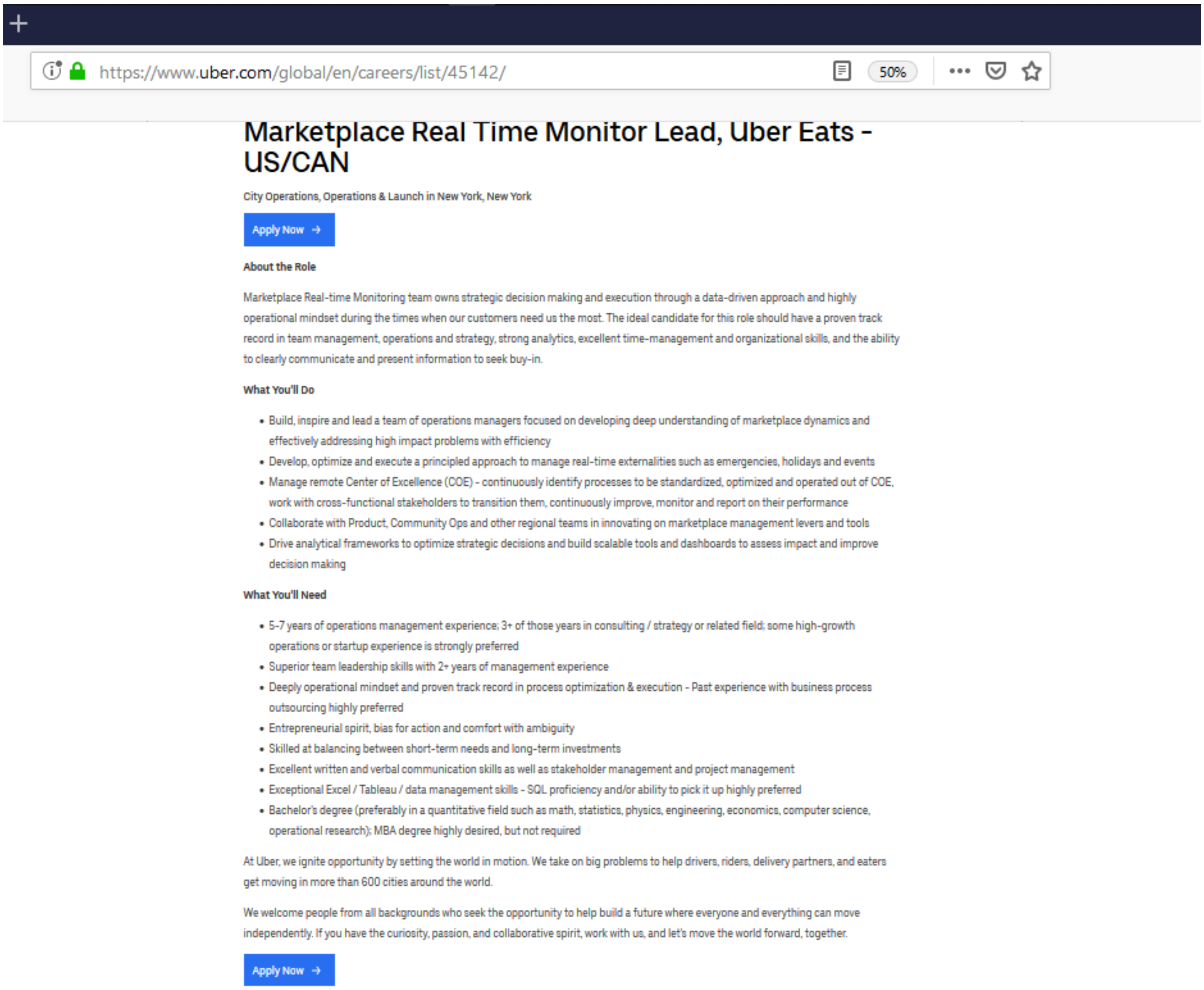
Race / Ethnicity (check one)*

- Hispanic or Latino
  A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.
- White (Not Hispanic or Latino)
  A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- Black or African American (Not Hispanic or Latino)
  A person having origins in any of the black racial groups of Africa.
- Native Hawaiian or Pacific Islander (Not Hispanic or Latino)
  A person having origins in any of the peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
- Asian (Not Hispanic or Latino)
  A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian Subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- Native American or Alaska Native (Not Hispanic or Latino)
  A person having origins in any of the original peoples of North and South America (including Central America), and who maintain tribal affiliation or community attachment.
- Two or More Races (Not Hispanic or Latino)
  All persons who identify with more than one of the above six races.
- Decline to Self Identify

APPLY NOW →

Notes: This set of figures depicts the application form. Candidates are asked about work experience, education, additional links (e.g., LinkedIn, GitHub), workforce authorization, and demographic information. Sections with a red asterisk are required, and basic personal, education, and experience information must also be populated before submitting the application.

Figure A.4: *Example Uber Job Posting*



https://www.uber.com/global/en/careers/list/45142/

Marketplace Real Time Monitor Lead, Uber Eats - US/CAN

City Operations, Operations & Launch in New York, New York

Apply Now →

About the Role

Marketplace Real-time Monitoring team owns strategic decision making and execution through a data-driven approach and highly operational mindset during the times when our customers need us the most. The ideal candidate for this role should have a proven track record in team management, operations and strategy, strong analytics, excellent time-management and organizational skills, and the ability to clearly communicate and present information to seek buy-in.

What You'll Do

- Build, inspire and lead a team of operations managers focused on developing deep understanding of marketplace dynamics and effectively addressing high impact problems with efficiency
- Develop, optimize and execute a principled approach to manage real-time externalities such as emergencies, holidays and events
- Manage remote Center of Excellence (COE) - continuously identify processes to be standardized, optimized and operated out of COE, work with cross-functional stakeholders to transition them, continuously improve, monitor and report on their performance
- Collaborate with Product, Community Ops and other regional teams in innovating on marketplace management levers and tools
- Drive analytical frameworks to optimize strategic decisions and build scalable tools and dashboards to assess impact and improve decision making

What You'll Need

- 5-7 years of operations management experience; 3+ of those years in consulting / strategy or related field; some high-growth operations or startup experience is strongly preferred
- Superior team leadership skills with 2+ years of management experience
- Deeply operational mindset and proven track record in process optimization & execution - Past experience with business process outsourcing highly preferred
- Entrepreneurial spirit, bias for action and comfort with ambiguity
- Skilled at balancing between short-term needs and long-term investments
- Excellent written and verbal communication skills as well as stakeholder management and project management
- Exceptional Excel / Tableau / data management skills - SQL proficiency and/or ability to pick it up highly preferred
- Bachelor's degree (preferably in a quantitative field such as math, statistics, physics, engineering, economics, computer science, operational research); MBA degree highly desired, but not required

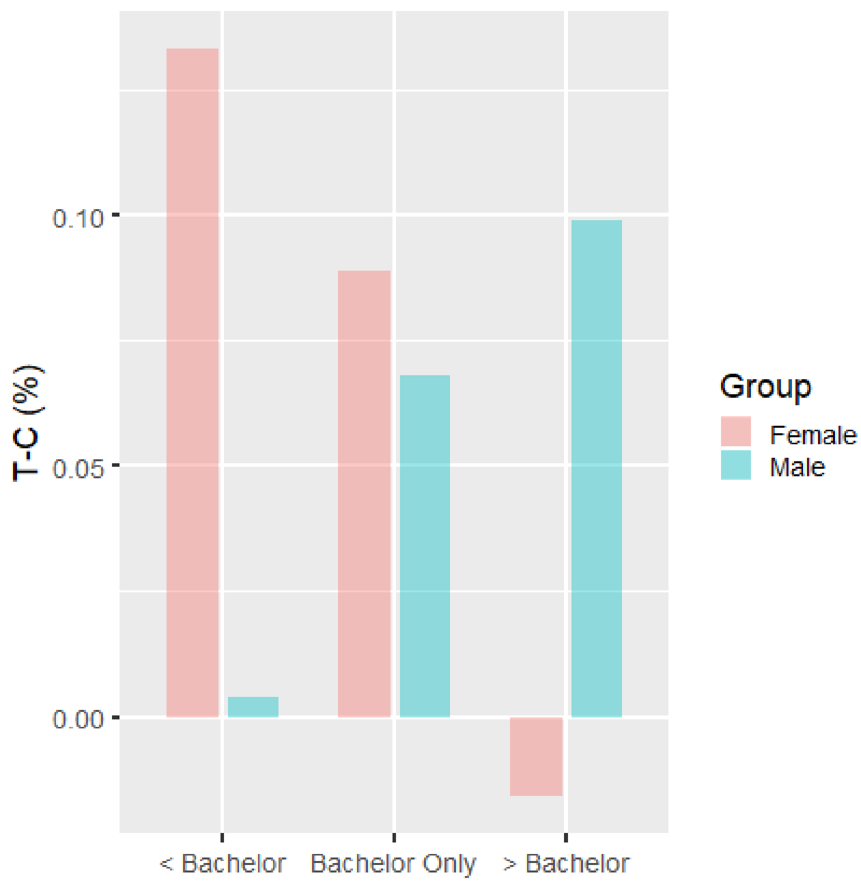
At Uber, we ignite opportunity by setting the world in motion. We take on big problems to help drivers, riders, delivery partners, and eaters get moving in more than 600 cities around the world.

We welcome people from all backgrounds who seek the opportunity to help build a future where everyone and everything can move independently. If you have the curiosity, passion, and collaborative spirit, work with us, and let's move the world forward, together.

Apply Now →

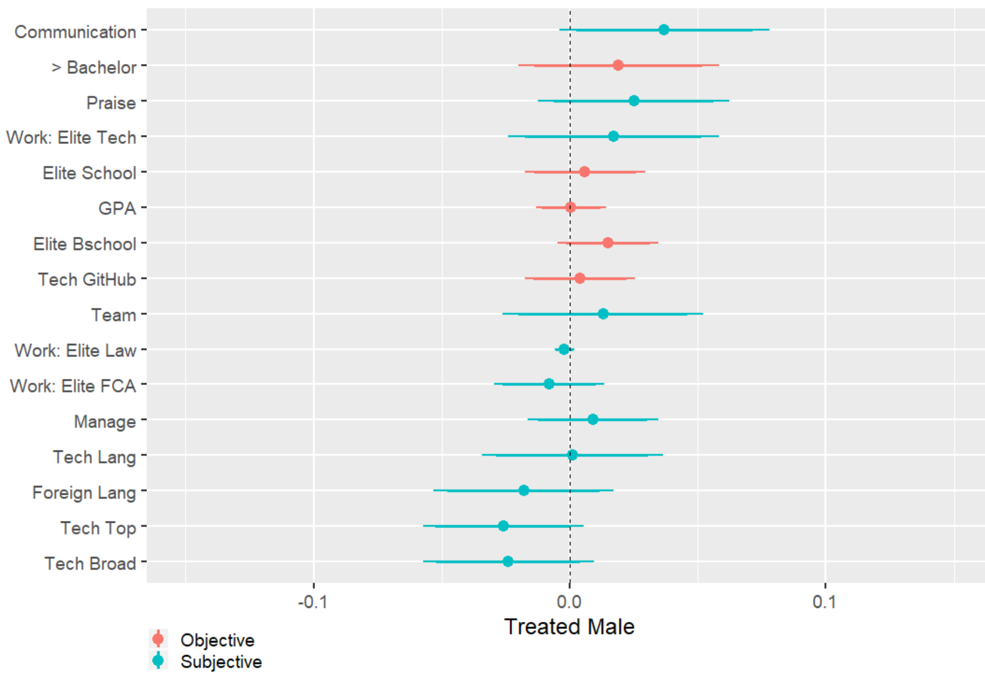
Notes: This figure depicts a job posting on Uber's careers website. All job postings on Uber's careers website exhibit the same general format and structure: job title, job location, "About the Role", "What You'll Do", and a "What You'll Need" section. Some job postings also contained an "About Uber" section and/or a team-specific section.

Figure A.5: *Percent Change in Levels to Educational Distribution*



Notes: This figure displays how the number of applications between the Treated and Control groups (as a percent of the Control group) changed within each educational attainment category. For females, there is a 13 percent and 9 percent increase in the less than bachelor and bachelor only categories, respectively, and a 2 percent decrease in the higher than bachelor category. For males, the opposite pattern occurs: there is a 0.4 percent, 7 percent, and 10 percent increase in the less than bachelor, bachelor only, and higher than bachelor categories, respectively. These differences reflect changes to the raw totals and do not control for job fixed-effects.

Figure A.6: *Distributional Changes for Application Attributes: Males*



Notes: This figure depicts the changes to application attributes for males as a result of our treatment. These estimates reflect individual regressions of the given attribute on a dummy variable for being Treated, being Female, and being Treated and Female, controlling for job fixed-effects and robust standard errors clustered by applicant. The graph displays the coefficient on the Treated variable, which represents the Treated Male premium (deficit) relative to Control Males (along with 90 and 95% confidence intervals). > Bachelor refers to listing an advanced degree. Elite School refers to listing an elite educational institution, Bschool refers to listing an elite business school, and GPA refers to the fractional form of the listed GPA, ranging between 0 and 1. Tech Top refers to listing one of the most commonly listed technical skills in our sample's job postings, Tech Broad refers to listing one of a broad set of over 40 technical skills sourced from our sample's job postings, Tech Lang refers to listing technically inclined language sourced from our sample's job postings, and Tech GitHub refers to listing "github." Work: Elite Tech refers to listing an elite tech company, Work: Elite FCA refers to listing an elite finance, consulting, or accounting company, and Work: Elite Law refers to listing an elite law firm. Social skills (Communication, Team, and Manage) refer to listing words describing these attributes in the resume text, and Praise refers to listing proactive, self-promoting, and praising language in the resume text. Foreign Language refers to listing one of 33 foreign languages in the resume text. See Appendix Document A.6 and Appendix Tables A.5 through A.8 for more detail.

Table A.1: *Browser & Device Detail*

Browser	Percent of Viewers
Chrome	61.7
Safari	29.2
Firefox	3.4
Microsoft Edge	1.6
Internet Explorer	1.4
Opera	0.1
UC Browser	< 0.0
Unidentified	2.5
Device	**Percent**
Desktop/Laptop	66.9
Phone	31.2
Tablet	1.8
Unidentified	< 0.0

Notes: This table shows the browser (top panel) and device (bottom panel) breakdown for the 59,140 viewers (as identified by browser-device) in the experiment. Unidentified browsers represent those that could not be identified by Optimizely (our A/B testing software), and likely reflect mobile-specific browsers. Unidentified devices represent those that were identified by Optimizely as being of more than one type (e.g., desktop and tablet).

Table A.2: *Effect of Treatment on Fraction Female Applications for Particular Subgroups*

	Female			
	Entry	Non-Entry	Technical Jobs	Legal group
Treated	−0.014	0.013	0.009	0.188**
	(0.020)	(0.023)	(0.036)	(0.085)
Job FE	X	X	X	X
Control Mean	0.442	0.295	0.292	0.388
Observations	3,682	2,503	1,337	160
R^2	0.115	0.226	0.148	0.312
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of the treatment on the fraction of female applications for entry level jobs, non-entry level jobs, technical jobs, and for job postings in the Legal subgroup. The dependent variable is a dummy variable equaling 1 if the individual reported being a female, and 0 if male; the sample is restricted to those who report their gender. The independent variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. Robust standard errors are clustered by applicant.

Table A.3: *Educational Attainment Identified by Education Substring*

	Educational Attainment		
	< Bachelor	Bachelor Only	> Bachelor
Treated	−0.018	−0.004	0.021
	(0.014)	(0.020)	(0.020)
Female	−0.010	-0.052^{**}	0.062^{***}
	(0.017)	(0.024)	(0.023)
Treated X Female	0.029	0.029	-0.059^{*}
	(0.022)	(0.031)	(0.030)
Job FE	X	X	X
Control Male Mean	0.157	0.402	0.441
Observations	6,185	6,185	6,185
R^2	0.282	0.187	0.295
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on educational attainment. Educational attainment is identified within the education substring of the resume text. The dependent variable in the first column is a dummy variable equaling 1 if the application is associated less than a bachelor's degree. The dependent variable in the second column is a dummy variable equaling 1 if the application is associated with only a bachelor's degree. The dependent variable in the third column is a dummy variable equaling 1 if the application is associated with having higher than a bachelor degree – defined as having a Master's, MBA, JD, or PhD. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.4: *Effect of Treatment on Having > Bachelor, by Degree Type & Technical Role*

	> Bachelor by Degree Type		
	Masters or MBA	JD	PhD
Treated	0.024	−0.007	0.003
	(0.020)	(0.005)	(0.005)
Female	0.070***	−0.010**	−0.001
	(0.023)	(0.005)	(0.005)
Treated X Female	−0.073**	0.014*	−0.001
	(0.031)	(0.007)	(0.008)
Job FE	X	X	X
Control Male Mean	0.405	0.037	0.021
Observations	6,185	6,185	6,185
R^2	0.267	0.629	0.187
	> Bachelor by Technical Role		
	Technical Job	ATG	ENG
Treated	0.001	0.067	0.002
	(0.039)	(0.059)	(0.044)
Female	0.157***	0.253***	0.169***
	(0.048)	(0.078)	(0.058)
Treated X Female	−0.030	−0.133	0.006
	(0.067)	(0.121)	(0.076)
Job FE	X	X	X
Control Male Mean	0.637	0.503	0.595
Observations	1,337	378	949
R^2	0.167	0.259	0.164
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on educational attainment by degree type and for technical jobs. The dependent variable in the top panel is a dummy variable equaling 1 if the application is associated with the educational degree listed in the column. Masters or MBA refers to listing a Masters or MBA (and no JD or PhD), JD refers to listing a JD, and PhD refer to listing a PhD. The dependent variable in the bottom panel is a dummy variable equaling 1 if the application is associated with a higher than bachelor's degree. Technical Job is comprised of job postings which have "Engineer","Data", or "Tech" in the job title, ATG reflects the Advanced Technologies Group job postings, and ENG reflects the Engineering group job postings. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.5: *Resume Analysis: Elite Education & GPA*

	Elite Education & GPA		
	Elite School	Elite BSchool	GPA
Treated	0.006	0.015	0.0005
	(0.012)	(0.010)	(0.007)
Female	0.018	0.014	0.015^{**}
	(0.014)	(0.012)	(0.007)
Treated X Female	-0.026	-0.029^{*}	-0.003
	(0.018)	(0.015)	(0.010)
Job FE	X	X	X
Control Male Mean	0.079	0.047	0.901
Observations	6,185	6,185	1,370
R^2	0.153	0.171	0.309
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on listing an elite educational institution and listed GPA. The dependent variable in the first column is a dummy variable equaling 1 if the application lists an elite educational institution (defined as listing one of the top 6 national universities in the 2020 US News & World Report rankings or an Ivy league institution). The dependent variable in the second column is a dummy variable equaling 1 if the application lists an elite business school (defined as one of the top 10 business schools in the 2020 US News & World Report rankings). The dependent variable in the third column is the fractional form of the listed GPA, ranging between 0 and 1; we trim values greater than 1 and less than 0.3, and we calculate the average if two or more GPAs are listed in the same application. See Appendix Document A.6 for more detail. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.6: *Resume Analysis: Technical Skills*

	Technical Skills			
	Top	Broad	Language	GitHub
Treated	−0.026	−0.024	0.001	0.004
	(0.016)	(0.017)	(0.018)	(0.011)
Female	−0.033*	−0.034*	−0.011	0.009
	(0.018)	(0.019)	(0.020)	(0.014)
Treated X Female	0.015	0.039	−0.033	−0.022
	(0.023)	(0.025)	(0.027)	(0.019)
Job FE	X	X	X	X
Control Male Mean	0.407	0.476	0.453	0.086
Observations	6,185	6,185	6,185	6,185
R^2	0.514	0.449	0.387	0.326
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on listing technical skills. The dependent variable in the first column is a dummy variable equaling 1 if the application lists one of the most commonly listed technical skills in our sample's job postings (C++, C#, Golang, Java, Python, R (proxied by RStudio and R Shiny), or SQL. The dependent variable in the second column is a dummy variable equaling 1 if the application lists one of a broad set of over 40 technical skills sourced from our sample's job postings. The dependent variable in the third column is a dummy variable equaling 1 if the application lists technically inclined language sourced from our sample's job postings (specifically, Android, iOS, web development, programming, machine learning, modeling, algorithm(s), algorithmic, data analysis, data analytics, quantitative analysis, big data, or data science). The dependent variable in the fourth column is a dummy variable equaling 1 if the application lists github, indicative of a GitHub account. See Appendix Document A.6 for more detail. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.7: *Resume Analysis: Work Experience*

	Elite Work Experience		
	Tech	FCA	Law
Treated	0.017	−0.008	−0.002
	(0.021)	(0.011)	(0.002)
Female	0.019	−0.002	0.001
	(0.024)	(0.014)	(0.002)
Treated X Female	−0.015	0.026	−0.001
	(0.033)	(0.017)	(0.003)
Job FE	X	X	X
Control Male Mean	0.496	0.079	0.005
Observations	6,185	6,185	6,185
R^2	0.192	0.185	0.282
Note:	*p<0.1; **p<0.05; ***p<0.01		

Notes: This table shows the effect of treatment and female status on work experience. The dependent variable in the first column is a dummy variable equaling 1 if the application lists an elite tech company (defined as one of the top 10 internet/social media companies, one of the top 8 hardware/equipment companies, or one of the top 10 software/computer services companies in Vault's 2019 rankings, supplemented by an additional list of elite tech firms identified in a survey of approximately 4,000 tech workers); note that this may not be a perfect measure of work experience (as some tech company names are also likely to be work skills, e.g.,"proficient in Google suite"). The dependent variable in the second column is a dummy variable equaling 1 if the application lists an elite finance, consulting, or accounting (FCA) company (defined as one of the 3 most prestigious finance companies, 3 most prestigious consulting companies, or 4 most prestigious accounting companies in Vault's 2020 rankings). The dependent variable in the third column is a dummy variable equaling 1 if the application lists an elite law firm (defined as one of 10 most prestigious law firms to work for in Vault's 2020 rankings). See Appendix Document A.6 for more detail. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.8: *Resume Analysis: Social Skills & Foreign Language*

	Social Skills & Foreign Language				
	Praise	Comm.	Team	Manage	Foreign Lang.
Treated	0.025	0.037*	0.013	0.009	−0.018
	(0.019)	(0.021)	(0.020)	(0.013)	(0.018)
Female	0.025	0.073***	0.008	−0.009	−0.024
	(0.023)	(0.026)	(0.024)	(0.014)	(0.021)
Treated X Female	−0.022	−0.039	−0.022	0.008	0.053*
	(0.030)	(0.034)	(0.032)	(0.020)	(0.028)
Job FE	X	X	X	X	X
Control Male Mean	0.702	0.337	0.687	0.892	0.224
Observations	6,185	6,185	6,185	6,185	6,185
R^2	0.114	0.115	0.112	0.139	0.129
Note:					*p<0.1; **p<0.05; ***p<0.01

Notes: This table shows the effect of treatment and female status on social skills and foreign language skills. The dependent variable in the first four columns is a dummy variable equaling 1 if the application lists words that correspond with the given attribute: Praise refers to proactive, self-promoting, and praising language, Comm refers to communication skills, Team refers to teamwork skills, and Manage refers to managerial skills. The dependent variable in column 5 is a dummy variable equaling 1 if the application lists at least one of 33 foreign languages. See Appendix Document A.6 for more detail. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Table A.9: *Resume Analysis: Any Bonus*

	Has Bonus			
	Tech	Educ	Cert.	Any
Treated	0.002	−0.031	0.010	−0.007
	(0.023)	(0.029)	(0.045)	(0.019)
Female	0.021	−0.031	0.011	−0.010
	(0.028)	(0.032)	(0.052)	(0.023)
Treated X Female	−0.032	0.005	−0.013	−0.005
	(0.037)	(0.044)	(0.068)	(0.030)
Job FE	X	X	X	X
Control Male Mean	0.385	0.558	0.242	0.498
Observations	2,744	2,251	687	4,578
R^2	0.346	0.249	0.311	0.319
Note:		*p<0.1; **p<0.05; ***p<0.01		

Notes: This table shows the effect of treatment and female status on having bonus points associated with the application's specific job posting. The sample is restricted to job postings which had bonus points that could be coded and matched on. The dependent variable in the first column is a dummy variable equaling 1 if the application lists any of the job posting's technical bonus points. The dependent variable in the second column is a dummy variable equaling 1 if the application lists any of the job posting's educational bonus points. The dependent variable in the third column is a dummy variable equaling 1 if the application lists any of the job posting's certification bonus points. The dependent variable in the fourth column is a dummy variable equaling 1 if the application lists any of the job posting's bonus points. The Treated variable is a dummy variable equaling 1 if the application is associated with a Treated viewer. The Female variable is a dummy variable equaling 1 if the individual reported being a female; the sample is restricted to those who report their gender. Robust standard errors are clustered by applicant.

Appendix B

Appendix to Chapter 2

B.1 Data Cleaning Procedures

The final sample of performance evaluation data used in the analysis reflects certain data-cleaning procedures, detailed below.

Sample Cleaning: I cleaned the sample by dropping performance reviews with "linked goals", a feature of the platform where workers can list goals they have for themselves and later review themselves on these goals (approximately 5 percent of the original sample); I chose not to include these reviews because workers might strategically choose goals that they then review themselves more highly on. I eliminated reviews with no reported questions underlying the review (less than approximately 1 percent of the original sample). I also dropped worker-reviewers with more than one observation per month (approximately 3 percent of the original sample). Workers and reviewers with tenure lower than 30 days (approximately 2 percent of the sample) were eliminated. The final sample contains non-missing data for worker/reviewer gender (described below), worker/reviewer tenure, and company identifier; the only variable which is missing for part of the sample is industry (information is available for 73 percent of the final sample). I also keep workers with at least one self score and one manager score (or other type of reviewer for the regressions that employ peers or direct reports).

Gender: In order to ascertain gender, the employee performance management software company employed standard gender reclassification software which provides a probability of being female or male based upon first name. The company then classified workers and reviewers as either female, male, or uncertain (i.e., androgynous or first name not able to be classified). I dropped reviews where gender of the worker or reviewer was uncertain (approximately 24 percent of the original sample).

Reviewer Score: I removed reviews with individual competency scores above 1 or below 0 (less than 1 percent of the sample). Note that I do not remove individual competency scores equal to 0 or equal to 1. Noise in the reviewer score represents measurement error in the dependent variable, and so would make it harder to find statistically significant coefficients for the independent variables, but would not lead to bias.

B.2 Creating Competency Categories

I constructed five skill categories in the dataset: *communication & collaboration*, *management & leadership*, *technical skills*, *work ethic*, and *proactive attitude*. I constructed these categories by extracting the competency scores in the data where the corresponding competency label contains words related to the skill (averaging if two or more competency labels in a given review are associated with the same skill). I list below the competency labels that are affiliated with each skill (note that some words are truncated to allow for matching across different forms of the word and parts of speech):

Communication & collaboration: collab*, communicat*, cooperat*, team*

Management & leadership: lead*, manag*

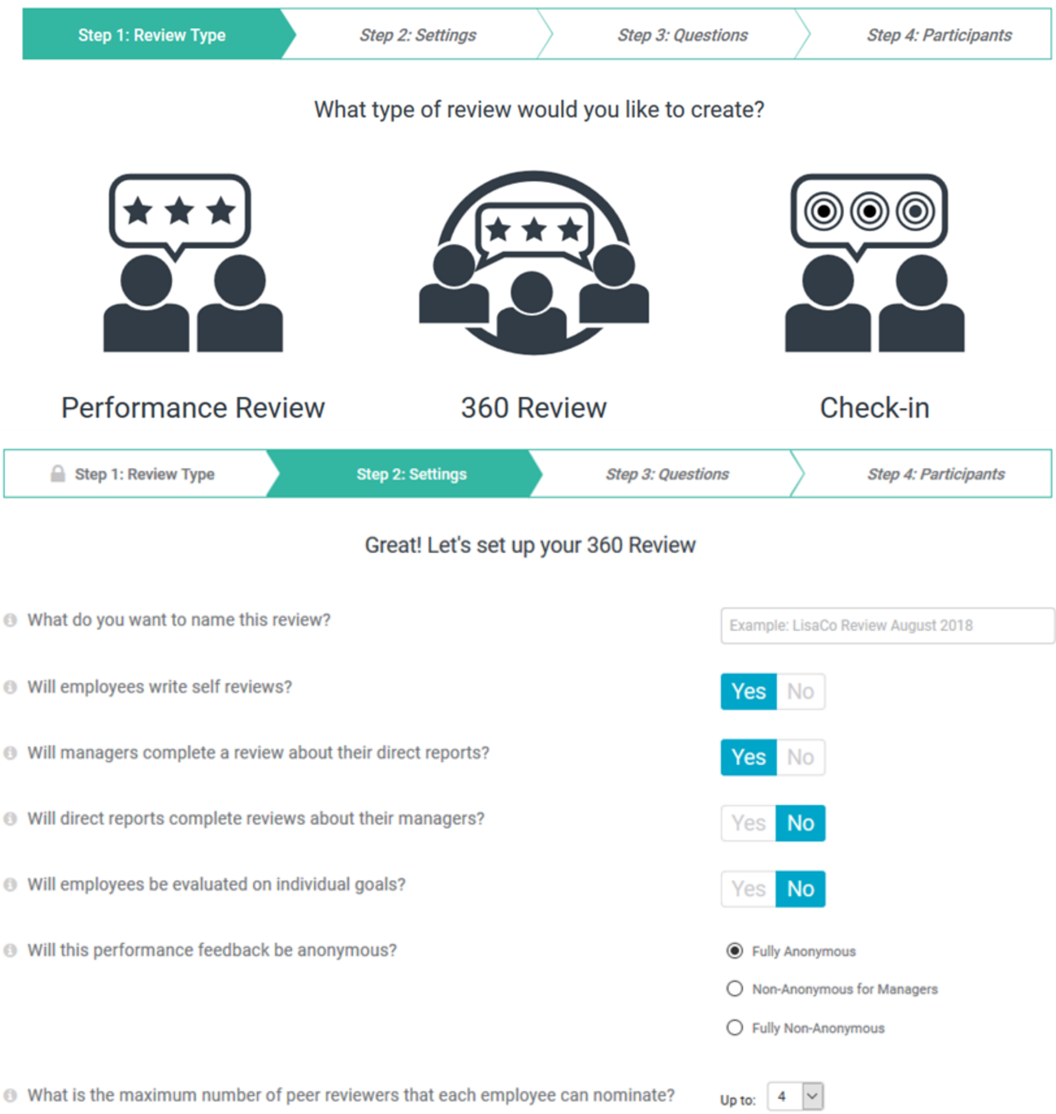
Technical skills: analytic*, code, data, tech*

Work ethic: accountab*, commit*, ethic*, integrity, ownership, responsib*, trust*

Proactive attitude: creativ*, deliver, execut*, fearless, initiat*, innovat*, passion, proactiv*

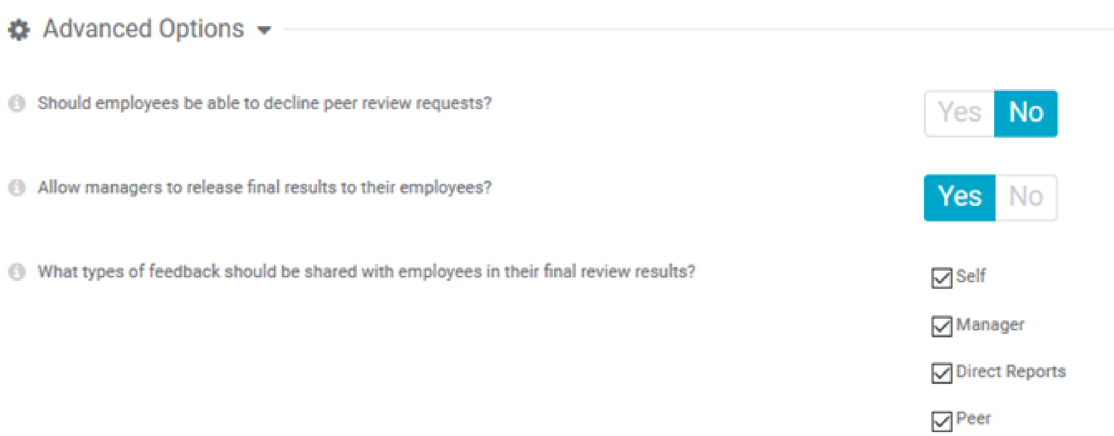
B.3 Supplementary Tables and Figures

Figure B.1: *Designing a Performance Review*



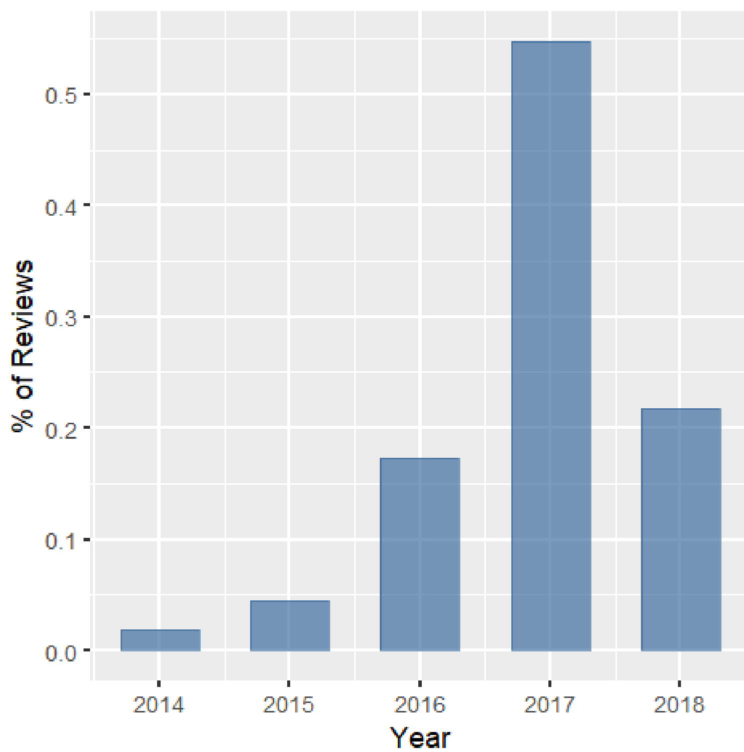
Notes: This figure displays the employee performance management software company's platform. HR representatives at the customer companies first choose the type of review they want to create and then customize the review along several different dimensions, such as who reviews whom and what information is shared with the worker.

Figure B.1: *Designing a Performance Review (Continued)*



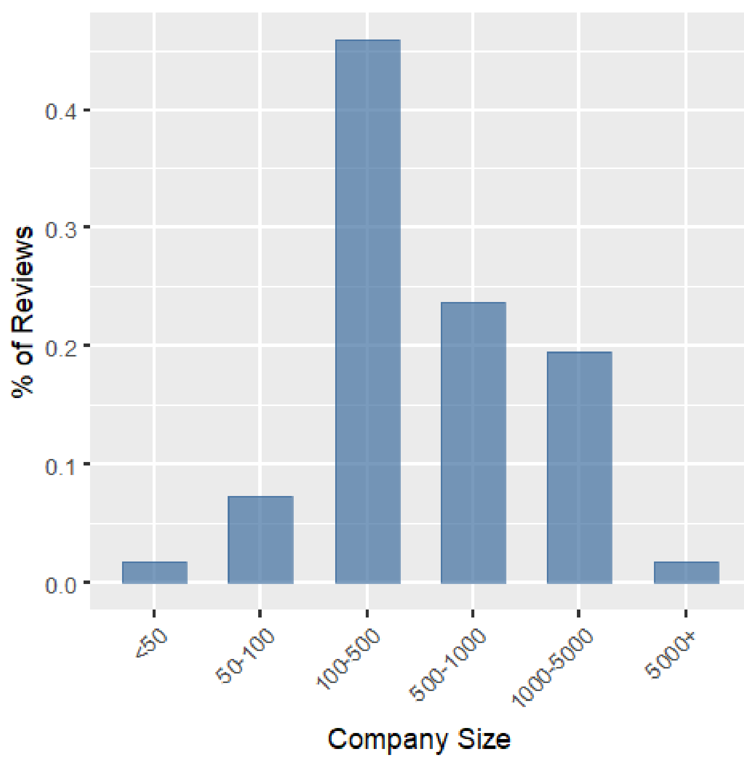
Notes: This figure displays the employee performance management software company's platform. HR representatives at the customer companies first choose the type of review they want to create and then customize the review along several different dimensions, such as who reviews whom and what information is shared with the worker.

Figure B.2: *Year Representation*



Notes: This figure shows the years that are represented in the final dataset as a share of reviews.

Figure B.3: *Company Size Representation*



Notes: This figure shows the number of employees for the companies in the final dataset, with most falling in the 100 to 500 employee-range as a share of reviews. These buckets represent total company size, even if not all employees at the given company are part of the employee performance management software company's dataset.

Table B.1: *Industry Representation*

Industry	Percent
Computer Software	17.3
Advertising & Marketing	15.4
Finance	6.8
IT Services	6.7
Services	5.8
Retail	4.2
Architecture & Engineering	3.9
Real Estate	3.8
Education	2.6
Manufacturing	1.9
Healthcare	1.5
Media	1.2
Entertainment	1.2
Telecom & Communication Services	0.3
Consumer Goods	0.1

Notes: This table displays the industries represented in the final sample as a share of reviews. Industry information is provided for approximately 73 percent of reviews.

Table B.2: *Industry Heterogeneity: Effect of Worker Gender on Self Minus Manager Score*

	Self-Manager Score												
	Advert.	ArchitEng	Computer	Educ	Entertain	Finance	HC	IT	Mfg	Media	Real Estate	Retail	Services
Female Worker	-0.117^{***}	-0.297	-0.036	-0.306^{***}	-0.210	-0.172^{***}	-0.043	-0.204^{**}	0.014	0.014	-0.155^{***}	-0.242^{***}	0.003
	(0.042)	(0.222)	(0.044)	(0.024)	(0.240)	(0.060)	(0.050)	(0.103)	(0.175)	(0.034)	(0.013)	(0.091)	(0.062)
Manager FE	X	X	X	X	X	X	X	X	X	X	X	X	X
Within Cycle	X	X	X	X	X	X	X	X	X	X	X	X	X
Observations	3,217	849	3,588	526	273	1,161	337	1,390	415	283	920	883	1,159
Adjusted R^2	0.122	0.140	0.184	0.107	0.078	0.144	0.209	0.155	0.406	0.283	0.228	0.158	0.231

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notes: This table displays regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each industry group: Advertising & Marketing, Architecture & Engineering, Computer Software, Education, Entertainment, Finance, Healthcare, IT Services, Manufacturing, Media, Real Estate, Retail, and Services. I omit two industries from this analysis (Consumer Goods and Telecom / Communication Services) because of too few observations in the industry group. Robust standard errors are clustered at the company level.

Table B.3: *Manager Tenure: Effect of Gender on Self Minus Manager Score*

	Self-Manager Score				
	<6 mo	6 mo to <1 yr	1 to <2 yr	2 to <3 yr	3+ yr
Female Worker	−0.114	−0.129**	−0.096**	−0.124***	−0.105***
	(0.078)	(0.063)	(0.043)	(0.036)	(0.034)
Manager FE	X	X	X	X	X
Within Cycle	X	X	X	X	X
Observations	1,591	2,799	4,483	3,671	8,385
Adjusted R^2	0.194	0.174	0.167	0.167	0.181
Note:					*p<0.1; **p<0.05; ***p<0.01

Notes: This table displays regressions of the Self-Manager z-score on a female dummy variable and a continuous measure of worker tenure for each manager tenure group: <6 months, 6 months to <1 year, 1 to <2 years, 2 to <3 years, and 3 or more years. Robust standard errors are clustered at the company level.

Table B.4: *Manager Gender: Effect of Worker Gender on Self Minus Manager Score*

	Self-Manager Score	
	Female Mng	Male Mng
Female Worker	-0.172^{***}	-0.087^{***}
	(0.039)	(0.024)
Manager FE	X	X
Within Cycle	X	X
Observations	5,958	14,971
Adjusted R^2	0.158	0.183
Note:	*p<0.1; **p<0.05; ***p<0.01	

Notes: This table displays regressions of the Self-Manager z-score listed in the column on a female dummy variable and a continuous measure of worker tenure. Column 1 is restricted to female managers and column 2 is restricted to male managers. Robust standard errors are clustered at the company level.

Table B.5: *Presence of Another Review: Effect of Gender on Self Minus Manager Score*

	Self-Manager Score	
	Peer/DR Review	No Other Review
Female Worker	-0.097^{***}	-0.123^{***}
	(0.028)	(0.033)
Manager FE	X	X
Observations	10,689	9,833
Adjusted R^2	0.153	0.196
Note:		*p<0.1; **p<0.05; ***p<0.01

Notes: This table displays regressions of the Self-Manager z-score listed in the column on a female dummy variable and a continuous measure of worker tenure. Column 1 reflects reviews with a peer or direct report evaluation in the cycle. Column 2 reflects reviews with no peer or direct report evaluations in the cycle. Both columns restrict the sample to reviews where managers are able to concurrently view peer or direct report evaluations. Robust standard errors are clustered at the company level.

Appendix C

Appendix to Chapter 3

C.1 Glassdoor Data Cleaning Procedures

1. Glassdoor Salary Reports

We made several edits to the original salary report data in our Glassdoor sample in preparation for our analysis.

We delete irrelevant / miscoded data: observations whose residence included American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, or the Virgin Islands; salary reports from 2006 and 2007 (given the much smaller sample size for these years) and two observations with miscoded years; and observations for which the years of reported experience exceeded the individual's age minus 16 (as we assume that an individual cannot begin working before age 16). We also drop 55 Glassdoor occupations which are vague and do not have a readily identifiable counterpart in the publicly available data (e.g., "fellow"), or which do not represent full-time occupations (e.g., "student").

We made several standard edits to the earnings variable. Given that users can report their earnings on an annual, monthly, or hourly basis, we converted all wages to an hourly basis (assuming a 2,000 hour work-year for earnings reported annually and a 12-month year for earnings reported monthly). Eighty percent of observations in our final sample report wages on an annual basis, but we choose to convert to hourly wages for comparison to publicly available data and ease of interpretation for our results. We also drop observations which are below the federal minimum wage in a given year; since we do not have monthly identifiers in the Glassdoor data, we use the federal minimum wage at the start of the year as the threshold. We also eliminate observations that report earnings above $500 per hour. All earnings information are converted to 2018 dollars using the annual Personal Consumption Expenditure Chain-type Price index.

Our final dataset is restricted to salary reports with non-missing information for our variables of interest over the 2008 to 2019 time period: earnings, gender, age (restricted to workers aged 25 to 64), education, years of experience, state, year, industry of employment, employer identifiers, occupation, and job title. We eliminate duplicate observations (the second or more instance of observations which report the same user, employer, specific job title, year, and wage). We further restrict there to be 10 males and 10 females in each Glassdoor occupation and drop single-sex employers for the gender pay analysis. Our final sample represents full-time, non-self-employed workers in the United States.

2. Glassdoor Reviews Data

We eliminate company reviews associated with individuals who submitted two or more reviews for a given employer while they held the same job title; this eliminates individuals who have strong views and could skew the sample (indicated by their submission of more than one review).

C.2 Publicly Available Data: Data Cleaning Procedures & Glassdoor Occupation Crosswalk

1. Current Population Survey MORG Data

We use the Current Population Survey Merged Outgoing Rotation Group files in our analysis. We restrict our sample to 2013 and onward because of an occupational recode in May 2012 which makes prior occupational data incompatible with the later years.

We make several additional edits to our CPS sample in preparation for our analysis, following Autor et al. (2016) and Hoynes et al. (2012) cleaning procedures. We restrict our sample to full-time, non-self-employed workers aged 25 to 64. We eliminated individuals who worked less than 35 hours per week in the week prior to the survey week. We also constructed a potential years of experience variable, calculated as the minimum of *(Age - years of education - 6)* or *(Age - 16)*.

We generated the hourly wage as earnings per week divided by hours last week except for individuals who reported being hourly workers (for whom we use the hourly wage reported). We then made several edits: we winsorize the top-coded value for weekly earnings; drop observations where the earnings data is greater than the 1/35th top-coded wages; and drop hourly earnings below the federal minimum wage.

All summary statistics and regressions for the CPS sample are probability weighted by the earnings weight variable.

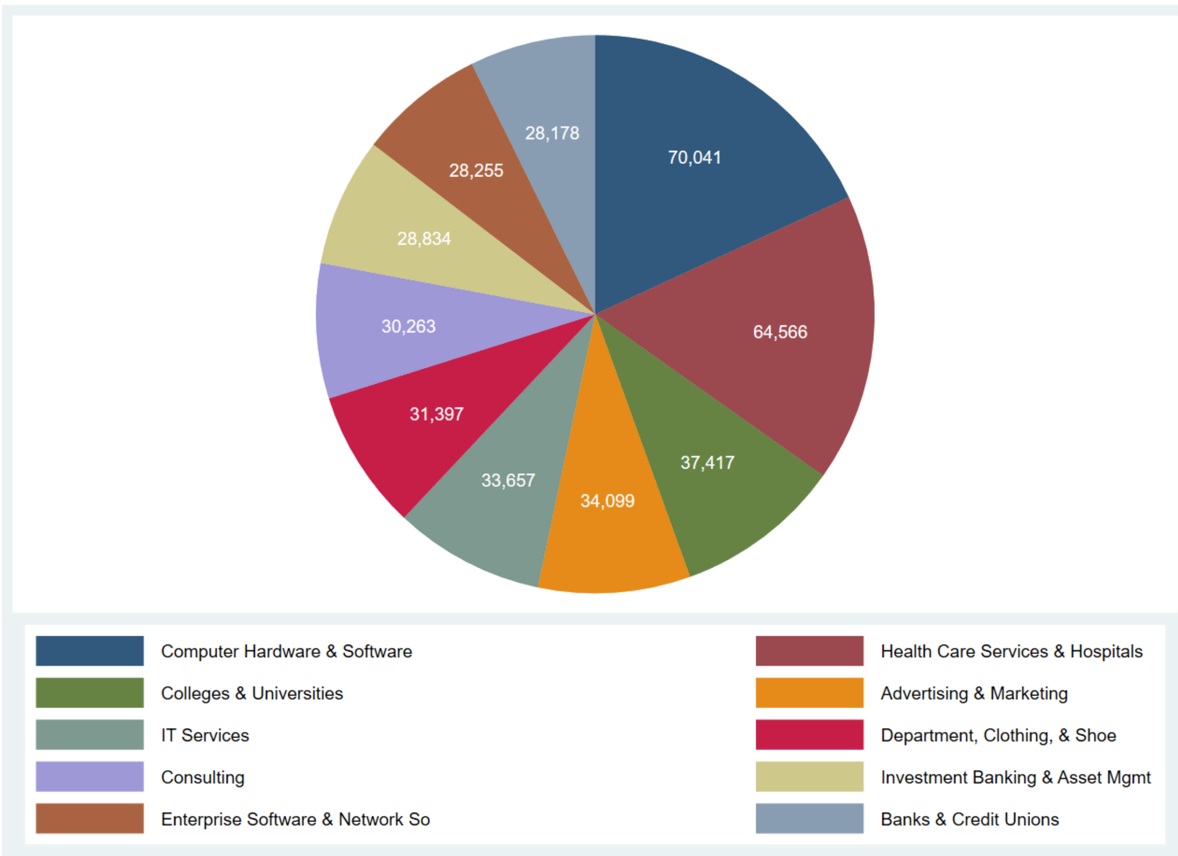
2. Occupational Employment Statistics Data

We use the 2018 Occupational Employment Statistics data in our manual construction between Glassdoor and OES occupations. Glassdoor occupational data is more detailed than the OES occupations, enabling us to map several Glassdoor occupations to a given OES occupation.

The OES-Glassdoor crosswalk enables us to extract the relevant portion of the CPS data using publicly available correspondences between the CPS occupational codes and OES occupations.

C.3 Supplementary Tables and Figures

Figure C.1: *Glassdoor's Top 10 Industries*



Notes: This figure represents the top 10 industries and associated number of salary reports in our data. These industry groups account for approximately 43 percent of the full sample.

Table C.1: *Years in Glassdoor Sample*

	Year
2008	16,064
2009	40,752
2010	58,095
2011	28,383
2012	59,732
2013	78,934
2014	100,223
2015	125,047
2016	128,145
2017	119,306
2018	83,403
2019	68,887

Notes: This table displays the number of observations for each calendar year represented in the Glassdoor sample.

Table C.2: *Reviews Subsample: Glassdoor Gender Wage Gap*

	ln(Wage)				
Female	-0.222***	-0.192***	-0.0852***	-0.0677***	-0.0614***
	(0.00607)	(0.00519)	(0.00446)	(0.00423)	(0.00524)
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	31,529	31,529	31,529	31,529	31,529
Adjusted R^2	0.180	0.423	0.677	0.714	0.785

Note: *p<0.1; **p<0.05; ***p<0.01.

Notes: This table displays regressions of the natural log of the hourly wage in 2018 dollars on a female dummy variable and controls, which include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to individuals who provided both a Glassdoor review for their employer and a corresponding salary review while working for the same employer and holding the same job title. There are 135 industry categories, 1,583 Glassdoor occupation categories, and 10,371 employers in the sample.

Table C.3: *Reviews Data: Do Females Complain about Pay?: Alternative Version*

	Complain about Pay				
Female	-0.0265***	-0.0247***	-0.0230***	-0.0245***	-0.0242***
	(0.00477)	(0.00478)	(0.00519)	(0.00524)	(0.00749)
ln(Wage)	X	X	X	X	X
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	31,529	31,529	31,529	31,529	31,529
Adjusted R^2	0.031	0.034	0.048	0.052	0.080
Note:			*p<0.1; **p<0.05; ***p<0.01.		

Notes: This table displays regressions of a dummy variable indicating whether the individual listed salary, wage, pay, or compensation in the "con" section of the company review on a female dummy variable; controls include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include the natural log of the hourly wage, and state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to individuals who provided both a Glassdoor review for their employer and a corresponding salary review while working for the same employer and holding the same job title. There are 135 industry categories, 1,583 Glassdoor occupation categories, and 10,371 employers in the sample.

Table C.4: *Reviews Data: Do Females Complain about Pay?: Not Controlling for Wage*

	Complain about Pay				
Female	0.00587	0.00456	-0.00335	-0.00795	-0.0134*
	(0.00493)	(0.00494)	(0.00543)	(0.00548)	(0.00777)
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	31,529	31,529	31,529	31,529	31,529
Adjusted R^2	0.002	0.011	0.023	0.029	0.072
Note:			*p<0.1; **p<0.05; ***p<0.01.		

Notes: This table displays regressions of a dummy variable indicating whether the individual listed salary, wage, pay, benefits, or compensation in the "con" section of the company review on a female dummy variable; controls include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to individuals who provided both a Glassdoor review for their employer and a corresponding salary review while working for the same employer and holding the same job title. There are 135 industry categories, 1,583 Glassdoor occupation categories, and 10,371 employers in the sample.

Table C.5: *Reviews Data: Do Females Complain Less about Discrimination?*

	Complain about Discrimination				
Female	0.00541***	0.00525***	0.00402***	0.00395***	0.00460**
	(0.00118)	(0.00118)	(0.00125)	(0.00126)	(0.00186)
ln(Wage)	X	X	X	X	X
Age, experience, education		X	X	X	X
GD Industry FE				X	
GD Occupation FE			X	X	X
Employer FE					X
Observations	31,529	31,529	31,529	31,529	31,529
Adjusted R^2	0.001	0.001	0.008	0.008	0.043
Note:					*p<0.1; **p<0.05; ***p<0.01.

Notes: This table displays regressions of a dummy variable indicating whether the individual listed words with the fragment discriminat, unfair, unequal, inequality, sexism, sexist or bias in the "con" section of the company review on a female dummy variable; controls include a continuous variable for age and its square, experience and its square, and fixed effects for an individual's highest education level. All regressions include the natural log of the hourly wage, and state and year fixed effects. Robust standard errors are in parentheses. The sample is restricted to individuals who provided both a Glassdoor review for their employer and a corresponding salary review while working for the same employer and holding the same job title. There are 135 industry categories, 1,583 Glassdoor occupation categories, and 10,371 employers in the sample.