Understanding Algorithmic Bias in Job Recommender Systems: An Audit Study Approach

Shuo Zhang [*] Peter Kuhn [†]

December 28, 2022

Abstract

To social science researchers, the recommendation algorithms used by job boards to recommend jobs to workers are a proprietary 'black box'. To derive insights into how these algorithms work, we conduct an algorithmic audit of four Chinese job boards, where we create fictitious applicant profiles and observe which jobs are recommended to profiles that differ only in age and gender. Focusing on the jobs that were recommended to just one of the two genders that applied, we find that only-to-women jobs propose lower wages, request fewer years of working experience, and are more likely to require literacy skills and administrative skills. Only-to-women (men) jobs also disproportionately contain words related to feminine (masculine) personality characteristics, as measured by three distinct approaches for identifying such characteristics. Finally, we assess the patterns in the recommendations generated by our audit study for their consistency with four processes the algorithms could be using: item-based collaborative filtering, content-based matching, matching based on recruiters' profile

[*]Department of Economics and Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115. E-mail: s.zhang@northeastern.edu.

[†]Department of Economics, University of California, Santa Barbara, CA 93106. E-mail: pjkuhn@econ.ucsb.edu .

We thank Kelly Bedard, Clément de Chaisemartin, Heather Royer, Catherine Weinberger, Sarah Bana for their helpful comments and suggestions. This study was approved under UCSB IRB No.17-20-0451 and is in the AEA RCT Registry under AEARCTR-0006101.

views, and rules-based matching based on employers' stated gender preferences. We find evidence suggesting that the algorithms are relying on all but the last of these processes.

Keywords: *Recommender System, Algorithm, Gender, Job Platform*

JEL Codes: C93, J71, J16, O33, M50

1 Introduction

With the rapid development of the Internet, the explosive growth of information makes it increasingly challenging for people to process a huge amount of data and to find desired information, products and workers. The personalized recommender system, first proposed in the 1990s, is a powerful tool to alleviate the information overload problem by prioritizing the delivery of information and showing every user a different list of new items that match her personal interests and preferences (Lee and Brusilovsky, 2007). Recommender systems have been widely and successfully applied in online websites and e-commerce services. For instance, a customer on Amazon possibly sees a page called "Customers Who Bought This Item Also Bought," which displays the products that she is likely to be interested in. After people watched a movie on Netflix, it often suggests people what to watch later, called "People Who Liked This Movie Also Saw"(Jannach et al., 2010).[1]

Similar scenarios can be found on internet-based recruiting platforms, which have now accumulated a vast volume of information on workers and jobs. According to statistics from The Conference Board and Glassdoor.com, in the US, there were 8.85 million jobs posted online by employers in 2021, and more than half of job seekers preferred finding job opportunities on online job sites.[2] In addition, the wide usage of online job searching and recruiting enables internet job boards to characterize behaviors and activities of job seekers and employers, which together foster the development of job recommender systems. Job recommender systems apply the concept of personalized recommendation to the job recruiting domain to suggest better matches between job seekers who search for job positions and recruiters who find candidates on the Internet. Virtually all internet job boards

[1] Recent evidence shows that 35% of purchase on Amazon and 80% of stream time on Netflix are driven by the recommendation systems. See https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48 and https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers.

[2] See The Conference Board®-The Burning Glass® Help Wanted OnLine® (HWOL) Index https://www.conference-board.org/topics/help-wanted-online, and Glassdoor's Job & Hiring Trends for 2020 https://www.glassdoor.com/research/app/uploads/sites/2/2019/11/Job_Hiring_Trends_2020-FINAL-1-1.pdf.

now recommend jobs to the workers who use their platforms. These customized recommendations are generated by algorithms, using criteria that include the worker's characteristics and previous behaviors, and the match between the worker's characteristics and the job's requirements. While job recommendation algorithms have the potential to help workers and firms find better matches faster, they also have sparked deep concerns about fairness: even when there is no discriminatory intent from designers, the recommended jobs may reinforce gender and other stereotypes. For instance, in content-based recommendation algorithms, gender might be associated with certain types of jobs and specific personalities in the workplace, which leads to gender segregation in job recommendations (Chaturvedi et al., 2021; Gaucher et al., 2011). Furthermore, based on job seekers' application behaviors, item-based collaborative filtering algorithms, as well as algorithms that incorporate the past behaviors of hiring agents, can create and perpetuate previous gender differences in recommendations received by workers.

This paper measures whether, to what extent, and how job board algorithms systematically treat male and female job seekers differently by conducting an *algorithm audit*, which is a new research approach proposed in recent years to study the black-box of algorithm features and to ascertain whether algorithms result in harmful discrimination by using fictitious correspondence in online platforms (Sandvig et al., 2014; Hannák et al., 2017). More specifically, we created otherwise identical male and female worker profiles on the four largest Chinese job boards, and observed which jobs were recommended to those profiles. In each job board, we selected 35 types of jobs on each platform based on three criteria: the number of active job openings, the job's gender-type (female-dominated jobs, gender-balanced jobs and male-dominated jobs), and hierarchy level (entry, middle, and high). These types spanned a wide range of skill levels, ranging from unskilled jobs such as sales and warehouse keeper, to high-level jobs such as financial manager and software engineer. Then we created resumes that were qualified for the above jobs; these come in pairs that are identical except for applicant gender. Since Chinese employers' gender preferences appear to interact strongly with the worker's age (Helleseter et al., 2020), we made two versions of each profile pair — a 'young' version and an 'older' version, in which the older applicants have 10 more years of working experience than young applicants. In order to track how algorithms update their recommendations based on workers' application behaviors, my fictitious workers then applied for the top jobs in their recommendation lists. We repeated this application process up to three times (each time responding to a new set of recommendations), then compared the job recommendations received by male and female applicants.

We find that identical male and female applicants do not always receive the same job recommendations: out of 100 job recommendations received by my applicants, 12.3 jobs were uniquely displayed to male or female applicants on average. Young workers, workers who are in the gender-neutral job types, and workers with middle- and high-level jobs received a greater number of gender-specific recommendations. Importantly, gender divisions in recommendations are even higher after fictitious applicants started applying for jobs: The raw difference rate between male and female applicants is 8.9% in the first round, whereas after three rounds of applications, 19.2 percent of recommendations are gender-specific.[3]

Although job boards do not always display the same jobs to our identical male and female workers, a particular concern is that the gender difference in job recommendations is caused by the computation randomness of the system. We address this problem by comparing the quality of only-to-men and only-to-women jobs in two phases. First, we focus on the job characteristics on the gender-specific jobs and find that on average, only-to-male jobs, which are seen by men rather than women, posted wages that were 2,616.1 RMB higher than jobs recommended to women, which is equivalent to 1.8% of the average wage of our fictitious workers. While the gender gap in requested education is close to zero in jobs recommended to male and female applicants, jobs recommended only to men have 0.17 more years of working experience requirement than only-to-female jobs (7.5% of the average requested working experience in recommended jobs).

[3] Because jobs displayed at the top of the recommendation list receive more attention, We further define the *ranking difference* in job recommendations in Appendix B, in which two job recommendations are the same only if both the job and the rank are identical in the recommended lists for pairwise workers (e.g., the third job in the men's list is the same with the third job in the women' list), and find that around three in four recommendations are different across male and female applicants.

In the second, we extract words used in the job descriptions reflecting six aspects of jobs' quality: *Skills*, *Work Timing and Location*, *Benefits*, *Company*, *Other Qualifications* and *Personality, Age, and Appearance* and estimate the association between the appearance of the word in the job description and the gender of applicants that receive the jobs. We find that literacy skills and administrative tasks are more likely to show up in female-only jobs, while influencing skills such as *leadership* and *decision-making* are mentioned more in male-only jobs. On the other hand, female applicants are recommended to apply for more jobs with *flexible* working hours and *regular breaks* in comparison to men with identical characteristics, while male applicants see more jobs that need *night work* and *overtime*. For benefits, only-to-female jobs place more emphasis on family related terms such as *marriage leave*, and *maternity leave*, while only-to-male jobs focus on more performance incentives such as *reward* and company *stocks*. Company-related words do not significantly differ between male-only and female-only jobs, except that orientation *training* is involved in more female-only jobs, while male-only jobs are more likely to be in *publicly-listed* companies.

The other requirements contained in the job descriptions also reflect gender-based differences in job recommendations. Words in jobs recommended to women are often related to feminine personality, such as *patient* and *careful*, and have more descriptions on desired workers' appearance such as *facial features*, *figure*, and *temperament*. Jobs recommended to men prefer workers who are *self-motivated*, *entrepreneurial*, and are able to *work under pressure*. Moreover, these male and female words in recommended jobs are consistent with gendered words summarized in previous literature in language (Fitzpatrick et al., 1995), in political science (Roberts and Utych, 2020), in psychology (Rudman and Kilianski, 2000) and in labor economics (Gaucher et al., 2011; Kuhn et al., 2020; Chaturvedi et al., 2021). To collect the gendered perceptions of words, we conducted two surveys on Amazon MTurk and on Chinese workers, and found that feminine words emerge more within jobs seen by female applicants and jobs recommended to men contain more masculine words. This suggests that words used in gender-specific jobs are associated with widely held gender stereotypes in the workplace, and the inclusion of stereotype-linked words contributes to the gender bias in job recommendation systems.

Finally, we attempt to isolate the precise mechanisms accounting for gender bias in job recommendations. *Content-based recommendations*, which link gender with jobs' features must play a role because words about gender-related personality traits (e.g., patient in female, work under pressure in male) and gender stereotypes in the workplace (e.g., women are good at literacy skills, men have leadership) occur differently in gender-specific recommendations. Moreover, hiring agents' behaviors also appear to contribute to gender-biased job recommendations. When more hiring agents read their profiles, the pairwise male and female applicants will see more different job ads in their recommendations, indicating that human bias may be maintained in and interact with recommender systems. Lastly, by comparing jobs recommended before and after workers apply for jobs, we find that *item-based collaborative filtering* which recommends jobs based on workers' application history may reinforce and amplify the gender bias in the system.

This paper contributes to five existing literatures, the first of which studies gender differences in job search: where, and how, do people look for jobs? Key findings in this literature include the fact that women are less likely to search for jobs far from their homes or in different occupations (Eriksson and Lagerström, 2012; Le Barbanchon et al., 2021), are attracted to jobs with flexible hours (Mas and Pallais, 2017), demonstrate higher levels of risk aversion when accepting job offers (Cortés et al., 2021), avoid competitive work environments (Flory et al., 2015), are attracted to co-operative work environments (Kuhn and Villeval, 2015), are less likely to negotiate starting wages (Card et al., 2016; Leibbrandt and List, 2015; Exley et al., 2020; Roussille, 2020), are more deterred by ambiguous information about job requirements the the number of competing applicants (Gee, 2019; Coffman et al., 2021; Abraham and Stein, 2022; Kline et al., 2022), and respond positively to affirmative action statements (Ibañez and Riener, 2018). A second related literature uses resume audit methods to study gender differences in how employers respond to job *applications* from identical men versus women. Notably, consistent with the idea that employers' gender preferences reflect stereotypes about gender-appropriate work, several papers in this literature have found that discrimination can run in both directions, depending on the type of jobs employers are trying to fill (Booth and Leigh, 2010; Cediey and Foroni, 2008;

Kline et al., 2022).

We contribute to both the preceding literatures by being one of the first papers to focus on a neglected stage of the job-worker matching process; this stage occurs before employers evaluate resumes, and even before workers apply for jobs. Specifically, we ask "Which job vacancies does a worker get to *see* before deciding where to apply?" As we document, automated job recommender systems display different job openings to identical male and female resumes. Since workers cannot apply to vacancies they are not aware of, gender differences that previous studies have attributed to differences in preferences could be caused, at least in part, by automated job recommender systems that inadvertently channel workers toward jobs that match common gender stereotypes. Put another way, these algorithms can create the appearance that men and women are choosing to apply to different types of jobs, when in fact they are unaware of some less-gender-typical vacancies that are available in their labor market.

Third, we contribute to the methodology of resume audit studies by adapting this widely-used tool to study the behavior of algorithms, rather than people. Algorithms are increasingly important actors in the economy, and their 'preferences' can be just as challenging to measure as humans'. While this is especially true for the proprietary algorithms that operate all the major job boards (which are 'black boxes' to outsiders), the complexity of these algorithms ensures that even their creators have limited understanding of their effects. While we study a different outcome than traditional resume audits, it may be worth noting that the algorithm audits we conduct are considerably easier to conduct on a large scale; this is good news given the increasing importance of algorithms importance as economic actors. Algorithm audits are 'easier' because they can be conducted with sparse worker profiles that do not require the investigator to fabricate detailed personal working histories, statements of purpose, and formatting decisions (font, margins, etc.) that consume investigator time and introduce noise.[4] Algorithm audits also have a validity advantage because they are harder for employers to detect (Avivi et al., 2021), and an ethical

[4] Kline et al. (2022) conducted a large scale resume audit in the U.S.; this was a very resource-intensive exercise compared to ours.

advantage because the inconvenience to human recruiters is negligible: very few of our sparse resumes receive call-backs, thus avoiding a waste of human recruiters' time.[5]

Fourth, our work relates to research in computer science and economics on algorithmic fairness in common contexts that include information retrieval, selection decisions, prediction problems, and recommender systems. In the computer science literature, algorithmic fairness is generally operationalized as a second criterion –in addition to a system's main objective (e.g. user engagement, prediction accuracy) to which the designer would like to assign some weight. Fairness-aware algorithms are then designed to incorporate concerns for objectives that include increasing the relevance of information provided to protected groups (Beutel et al., 2019; Bozdag, 2013) or infrequent users (Fu et al., 2020), increasing employers' exposure to minority job candidates (Li et al., 2020), or increasing news readers' exposure to less popular topics (Gao et al., 2021). Economists studying algorithmic fairness have considered the optimal design of algorithms from the point of view of a social planner who cares about equity (Kleinberg et al., 2018; Rambachan et al., 2020), arguing for example that optimal algorithms should not sacrifice *prediction* accuracy by blinding the prediction process to protected characteristics like race.

More closely related to our own research, both economists and computer scientists have empirically estimated the types and amounts of algorithmic bias in a variety of contexts. These authors have studied racial bias against black defendants (Angwin et al., 2016; Cowgill and Tucker, 2019), racial and ethnic discrimination in mortgage lending and credit approval (Bartlett et al., 2021; Fuster et al., 2020), racial discrimination in the health care system(Obermeyer et al., 2019), algorithmic unfairness in opioid use (Kilby, 2021), and gender disparities in image search and face recognition(Kay et al., 2015; Klare et al., 2012).[6]

[5] Algorithmic audits (where investigators supply a series of inputs to 'black box' algorithms and use the outputs to infer properties of the algorithm) have been used by computer scientists in a variety of contexts. For example, Buolamwini and Gebru (2018) compare the accuracy of commercial gender classifier algorithms (which infer gender from facial photographs) across races. Hannak et al. (2014) searched e-commerce sites in the guise of users with different demographics to measure differences in steering and price discrimination across users. For other examples, see Bolukbasi et al. (2016) and Kay et al. (2015). We are not aware of any uses of algorithmic audits in the economics literature.

[6] A subset of these studies (including Arnold et al. (2021) and Mullainathan and Obermeyer (2021)) focus on *prediction problems*, where an algorithm's impact on different groups can be measured against realized outcomes (such as bail violations and healthcare utilization). Unfortunately this is not possible in

Fifth and finally, we contribute to a growing literature on the effectiveness and the fairness of algorithms in the specific context of worker recruitment. One strand of this research focuses on algorithmic decision tools for selecting employees from a pool of candidates that has already been assembled. These *employee selection* tools perform functions that include resume screening, AI interviews, evaluation of interview performance, pre-employment assessments, and productivity prediction. Key contributions here include Hoffman et al. (2018) and Li et al. (2020). Hoffman et al. (2018) compare the performance of pre-employment screening algorithms and human HR agents by studying what happens when human agents overrule a widely used algorithm. They find evidence that the algorithms outperform humans, in part because the algorithms are *less* biased. Li et al. (2020) build a resume screening algorithm that values candidates' statistical upside potential, then assesses the likely effects of that algorithm using data on past hires at a Fortune 500 firm. They predict that their exploration-augmented algorithm would improve both the quality and diversity of candidates selected for an interview, relative to the firm's existing practices.[7]

A second use of algorithms in hiring is in the design of *resume search engines*, which allow employers to search the internet and other large databases for potential hires in response to the employer's search criteria. Chen et al. (2018) is the only paper we know of that systematically assesses these tools for bias, or the lack of it. Taking the role of employers, the investigators search for resumes in 35 job titles on Indeed, Monster, and CareerBuilder and study the ranking of men and women in the search results, while controlling for observable differences of the suggested resumes.[8] Overall, they find that male resumes rank slightly higher than observationally identical female resumes, but (consistent with other studies of employers' gender preferences) this gap is not uniform across job titles.[9]

our recommender-system context. We therefore focus our attention on characterizing the differences between the jobs that are recommended to identical men versus women, without taking a stand on whether the lower-wage, higher-flexibility jobs recommended to women at better or worse in any absolute sense.

[7] Raghavan et al. (2020) summarizes the advertised capabilities of 18 vendors of algorithmic pre-employment assessments.

[8] Gender is inferred from the first names in the resumes.

[9] As a supplement to their main analysis, Chen et al. (2018) conduct an experiment that creates and posts

While resume search engines are used by some employers in tight labor markets, the large majority of hires in most labor markets result from applications workers have made to jobs, not active search by employers. This suggests that the *job recommender algorithms* we study in this paper play an important role, since they help determine where workers apply. To our knowledge, our paper is the only one to experimentally estimate the presence of bias of any form in the recommender systems used on internet job boards, which is where most workers go to search for jobs online.[10]

The rest of this paper is organized as follows: Section 2 presents the potential mechanisms of the gender-biased job recommendations in online job boards. Section 3 details the experiment design and implementation. Section 4 summarizes the experimental results on the differences in job recommendations between male and female applicants. We explore the potential drivers of algorithmic gender bias in job recommendations in Section 5. Section 6 concludes.

2 Job Recommender Systems and their Potential Effects

On internet job platforms, when a job seeker with a complete profile logs into her account, the website displays a list of jobs that the job seeker may be interested in on her homepage. Unlike the board's search function –which requires job seekers to input keywords– job recommender systems present suggested jobs proactively and automati-

identical male and female resumes, then measures how those resumes are ranked by the resume search engines. They did not find any difference, suggesting that the algorithms on these job boards do not directly make use of applicant's genders (as signaled by the names on their resumes).

[10] Two recent papers (Lambrecht and Tucker, 2019; Ali et al., 2019) do, however, present results about *which workers see which job ads*, which is a central question in our paper as well. Both these papers run *advertising experiments* on Facebook, where they purchase ads, then observe which ones are seen by audiences they have created. For the case of job ads, both papers detect large gender differences, but these differences are driven by a factor that plays no role on job boards (which is where most workers look for jobs). Specifically, Lambrecht and Tucker (2019) found that their ads for STEM jobs were much more likely to be seen by men than women, but this was because job ads on Facebook must compete for visibility with ads for goods and services. Since female eyeballs are more valuable in the shopping domain, Facebook's cost-effectiveness algorithm economized on the delivery of job-related ads to women.

cally.[11] To generate these results, most job boards rely on a combination of methods; we summarize the four most common ones in this Section.[12] Although these algorithms are theoretically intended to be gender-neutral, for each of them we identify reasons why they might recommend different types of jobs to identical male and female worker profiles.

The core of most recommender systems is *item-based collaborative filtering* (IBCF). Used in a large variety of contexts (including shopping and information retrieval), IBCF uses the implicit collaboration of users (in our case, job seekers) to predict an individual job seeker's preferences over items (in our case, job ads). For example, suppose user A has applied for job X. IBCF then finds the 'co-applicants' who also applied for job X and recommends the jobs liked by those co-applicants to user A.[13] In simpler terms, IBCF tells workers, "people who applied to this job also applied to..." (Jannach et al., 2016). Put another way, IBCF recommends jobs to workers that are similar to jobs they previously expressed interest in, where 'similarity' is defined by the behavior of other workers. Importantly, because IBCF is based on a worker's previous expressions of interest (such as clicking or applying), it cannot generate recommendations for workers who have just created a profile on the job board. This is known as the 'cold start' problem of IBCF.

On their own, IBCF algorithms are unlikely to cause differences in the types of jobs recommended to identical male and female resumes. This is because the IBCF process for each worker is 'seeded' by that worker's first set of applications. The process then continues to recommend jobs that share co-applicants with the worker's previous applications. Thus, IBCF cannot be responsible for any gender recommendation gaps in Round 0 of our experiment, because these recommendations are made immediately after the profiles have been created– i.e. before our profiles engage in any actions, including applications. That said, IBCF can, for example, reproduce and intensify gender-stereotypical job prefer-

[11] Job recommendations play an increasingly important role in online job boards because Boolean search methods based on keywords entered by the user are frequently ineffective in generating useful matches (Lang et al., 2011).

[12] Our descriptions in this Section are based on a a combination of previous survey research (Al-Otaibi and Ykhlef (2012), Hong et al. (2013) and Siting et al. (2012)) and personal experience with several job boards.

[13] To simplify the discussion of IBCF, we use 'co-applicants' as a shorthand to include the other workers who expressed interest in the same job, whether by applying, viewing, saving it, or otherwise.

ences suggested by a worker's early applications (which can differ in our experiment due to other components of the job recommender system). To illustrate, suppose that the first job a woman applies to is "driver", which is extremely male on Chinese job boards (Kuhn and Shen, 2021). Because mostly men apply to driving jobs, IBCF will then continue to show her other male-intensive jobs. The opposite applies to a woman whose first applications are to receptionist jobs with flexible hours– future recommendations will have the same features. Thus, IBCF simply perpetuates and reinforces the gender mix of the worker's past applications.

One widely-used recommendation method that can alleviate IBCF's cold-start problem is *content-based* recommendations, which are widely used as a supplement to IBCF. Based on text analysis and natural language processing, content-based recommendation algorithms identify similarity between two documents by comparing the words in the documents. In our case, 'content' refers to the text of the job ad and the words in the job seeker's profile. In online job boards, content similarity can be established between jobs, between workers, and between jobs and workers. Two jobs are defined as similar when the same keywords appear in their job descriptions; in this case the algorithm can recommend one of these jobs to the worker if she has applied to the other. Two workers are similar if their resumes have the same keywords; thus the jobs one worker has applied to can be recommended to the other. While both these uses of content rely on previous expressions of user interest, content-based measures of the similarity *between jobs and workers* do not require any history at all. For example, if a job ad and the worker's resume contain the same keyword, such as a skill, then the system will suggest the job seeker to apply for that job.

Content-based recommendations are based on natural language processing, which is well known to encode gender stereotypes. For instance, NLP algorithms are more likely to associate female names with family than career words, compared with male names (Nosek et al., 2002). Nurse, teacher are more likely to be associated with she or her, while engineer, scientist are associated with he or him, suggesting that implicit gender-occupation biases are linked to gender gaps in occupational participation (Caliskan et al., 2017). In addition, the selection of keywords to identify similarity in contents may embed gender bias. If keywords for matching contents between workers and jobs are encoded to be correlated with gender identity, the algorithm eliminates workers whose resumes do not contain the gender-related keywords from jobs listing these keywords in their descriptions(Savage and Bales, 2016). Furthermore, if the keywords associated with strong gender tendency are used to define similarity between workers, workers with the same gender consequently are more likely to be similar and see the same jobs, leading to gender segregation in job recommendations. Finally, when jobs having gendered keywords are defined as similar, a worker that applies for one job with gendered words, will be recommended to other jobs that also contain such gendered words.

This noted, content-based recommendations in Round 0 of our experiment can only create systematic gender differences in recommendations in job recommendations under one condition: if algorithms make explicit use of the the worker's declared gender field in her profile. This is because our sparse, fictitious applicant profiles (which contain very few gender-stereotypical words to begin with) are identical in all respects but gender within each gender pair. In later rounds, gender-typical words that appear in the jobs our profiles apply to when they follow their first top-ten recommendations, could accentuate any such initial differences, but content-based recommendations can only cause systematic Round 0 gender gaps if content-based recommendations routine make use of workers' declared gender in their profiles.

A third method used in job recommender systems applies a *rule-based approach* to generate suggestions based on the match between jobs and workers. Rather than inferring similarity using natural language-based approaches, rule-based approaches use the encoded fields of the job ad and worker profile to measure the 'fit' between a worker and a job. For example, if a worker satisfies the education requirement of a job, the job board is more likely to recommend this job to the worker. Like content-based recommendations (when applied to job-worker matches), rule-based recommendations are useful in cold-start situations because they do not rely on any previous user behavior.

Given that only content-based and rule-based recommendations have the potential to create gender differences in Round 0, we focus our discussion of the likely impact of rule-based systems on this Round. Since Chinese employers are still allowed to state a gender preference in their job profiles (even though Chinese law prevents that preference from being posted in the job ad), one straightforward rule that could generate a gender gap would be "don't recommend jobs that request women to men." We cannot test for such rules directly because we do not observe which job ads contain gender requests, but we note that explicit gender preferences of the above form have become extremely rare on the major Chinese job boards (Kuhn and Shen, 2013, 2021; Kuhn et al., 2022).

We note that naturally-occurring female profiles are likely to have a very different set of observed qualifications than male profiles (for example, more experience in female-dominated occupations and industries). These differences, combined with rules-based matching on occupation, industry and other observables, could easily account for large gender differentials in job recommendations, even in profiles that have no application history. Importantly, however, these factors cannot account for any Round 0 gender differences we detect, because our paired worker profiles are identical in all these respects. For example, while males are considered as 'more suitable' to driving jobs and females are considered as 'more suitable' to receptionist jobs, we are comparing identical male and female applications to driving jobs, and identical male applications to receptionist jobs, and rules cannot create a Round 0 gender gap unless rules take the form of gender.

Finally, some job boards apply more sophisticated systems that incorporate the behavior of *hiring agents* into their recommender systems.[14] Aspects of an agent's behavior that can be used by these systems include browsing a worker's resume, downloading it, or marking it as a 'target' in the job board's application processing interface.[15] Based on these behaviors, recruiter-based recommender algorithms recommend a recruiter's job to workers who are similar to the workers that recruiter recently downloaded, targeted, or

[14] On Facebook, this would be the equivalent of using peoples' responses to friend requests to learn the types of people they would like to have as friends; it is not relevant to shopping sites.

[15] In addition to improving job-worker matching, job boards have incorporated indicators of recruiter engagement into their systems to keep job seekers engaged. Specifically, their concern is that workers who receive no feedback may become frustrated and switch to other sites (Kim, 2017).

called back (Yu et al., 2011).[16] Notably, in most cases job boards can estimate recruiter preferences at the *job ad* level (so, for example, the same recruiter can prefer experienced workers for higher-level positions and new graduate for another).[17]

Consider recruiter-behavior based approach that incorporates hiring agents' rating behaviors (e.g., viewing and downloading profile, sending a message to target worker) into the recommender system. As far as we know, there are three scenarios in which the hiring agents' behaviors could affect job recommendations.[18] Suppose a hiring agent posted a job and received some applications from both genders, but has consistently ignored female applicants (for example, never downloaded female resumes).[19] Two points are learnt from this process: First, this job is not going to hire female workers, then it will not be recommended to other female workers. Second, if a female applicant did not get positive feedback from the job, the algorithm infers that she is unlikely to get callbacks from other jobs that are similar to that job, so those similar jobs will not be recommended to her. That is to say, workers' recommendation results are affected by the processing decisions of the hiring agents who posted jobs that they have already applied to, as well as the spillover effects from other hiring agents. Moreover, in most online job boards, hiring agents can search for and contact suitable workers directly. When a hiring agent searches for workers and clicks into a worker's resume, the jobs posted by this hiring agent will be recommended to that worker, as the hiring agent has shown interests to that worker (Köchling and Wehner, 2020). If a hiring agent persistently views resumes of male workers, those male workers will be suggested to apply while female workers do not have this priority (Burke et al., 2018). We note that recruiter-behavior-based algorithms, like IBCF, have a cold-start problem: they require previous recruiter reactions to a resume to provide

[16] Interestingly, similarity in this context, to our knowledge, is not established in a collaborative way, as in IBCF. (The analogy to IBCF here would be to find other recruiters who also expressed interest in the same resume.) Instead –in part because recruiters' actions are much sparser than jobseekers' actions– similarity is based on the workers' characteristics and on the unstructured text of the worker's resume.

[17] The main exception to this is when recruiters pro-actively search for resumes on the site; in these cases the searches are not necessarily tied to a particular job opening.

[18] Algorithms targeting at click maximization are likely to deliver biased results, due to the feedback loop (Jiang et al., 2019) and learning-to-rank approach (Jiang et al., 2016; de Sá et al., 2016).

[19] The four online job boards allow recruiters to filter workers' profiles by demographics (e.g., gender, age) and characteristics (e.g. education, experience) when they process received applications or search for suitable candidates.

suggestions to that resume. However, since IBCF relies on past *worker* behavior, it may still be possible to distinguish between these systems from algorithmic audits of a job board.

While the above channels seem to be plausible ways a recruiter-based algorithm could send different recommendations to male and female candidates, it is worth distinguishing between how such algorithms are likely to affect naturally occurring applications versus the ones in our experiment. To the best our knowledge, the recruiter-based algorithms that are currently in use do not make adjustments for the composition of the applicant pool a recruiter faces. Thus, in naturally-occurring data, recruiter-based algorithms will reflect not only the gender preferences of individual recruiters but also the application decisions of real workers: Resumes that are viewed by recruiters for driving jobs might be almost exclusively male simply because women don't apply to those jobs. In other words, in naturally occurring data, recruiter-based algorithms could be perpetuating historical patterns of worker application (self-segregation) decisions, at least as much as reflecting recruiter biases.

In our experimental data, however, this application-based mechanism is largely shut down, because our profiles are programmed to only apply to the jobs that are recommended to them. This rules out any choices made by real men and women to segregate themselves by choosing to apply to different types of jobs, leaving only our algorithms as possible channels. Put another way, while 'men' and 'women' could still have different application histories in our experiment –and while those histories will affect recruiter behavior by changing the set of applications they can react to– the difference in those application histories will be caused purely by the algorithms used by the four job boards in our study.

A final, noteworthy feature of recruiter-based algorithms is, of course, that they require recruiter feedback to operate. Thus, like IBCF, they cannot account for gender differences in the 20 job recommendations we collect in Round 0 of our experiment. That said, one feature of our experiment could help us distinguish between IBCF and recruiter-based algorithms. To see this, consider the two-week interval between Round 0 and Round 1.

During this period, our profiles do not make any applications. They could, however, be viewed by recruiters who are using the resume search feature of the job board. Thus, the first 10 recommendations that are harvested in Round 1 (before any applications are sent out) could be affected by rules-based algorithms but not by IBCF. Further, the change in content between the between the first and second 10 recommendations in Round 1 can only be driven by IBCF (because applications are submitted and recommendations collected immediately afterwards). The same reasoning applies within rounds 2 and 3, giving us another potential way to distinguish IBCF and recruiter-based algorithms.

The four potential mechanisms described above can interact with each other to create a complex job recommendation system.[20] More generally, algorithms may replicate the errors stemming from the training data, such as choosing parameters based on data with existing stereotypes, which detracts from gender fairness (Hellström et al., 2020; Rambachan and Roth, 2019; Kim, 2016; Barocas and Selbst, 2016). Overall, recommender systems may reproduce and magnify pre-existing gender bias in the labor market.

3 Experimental Design

3.1 Platform Environments

To represent a broad sample of jobs and workers, our experiment was conducted on the four largest job boards in China, which serve millions of job seekers and job postings.[21] The large size of these markets allows us to create a large number of fictitious workers, while minimizing any distortions we might impose on the existing job search and recruiting processes or the job recommender systems. The four job sites have similar interfaces and functions for users, which are typical of most online job platforms. Job seekers can

[20] Both direct discrimination and indirect discrimination on gender potentially exist in these algorithms, which are distinguished by whether sensitive features (gender) are not explicitly used as inputs in algorithms (Pedreshi et al., 2008).

[21] The four job sites have the most active workers and job advertisements, and cover more than 70% of online labor markets in China.

register and create a profile for free, while employers are charged for posting job advertisements and using recruiter tools. Job seekers submit applications by sending their resumes to the jobs that they are interested in, and firms' hiring agents can check and process the applications online and contact applicants through each website's message system. Furthermore, in line with current industry practice, we expect that all four of these boards use sophisticated forms of machine learning to suggest jobs to workers.

3.2 Job Type Selection

When a job seeker sets up her profile, the job platforms ask her to select her current and desired industry and occupation from drop-down list supplied by each board. For our audit study, we selected 35 industry-occupation cells (a.k.a. *job types*); this selection was based on three criteria: sample size, the cell's incumbent gender mix, and hierarchy level. As a first step, we chose industry-occupation cells that have a large number of job postings to ensure that there were enough new job vacancies to be recommended to workers.[22] For instance, the internet industry has the most job postings, while sales are the most popular occupations in job sites, so internet-sales is one of our job types. Second, because male-and female-dominated jobs might prefer applicants whose gender is typical for their occupation-industry cell, we included examples of three types of jobs: female-dominated (e.g. administrative assistant), (approximately) gender-balanced (e.g. sales), and male-dominated jobs (e.g. software engineer).[23] Finally, because algorithms might reflect employer gender preferences that vary with the position's rank (Bertrand et al., 2010; Pekkarinen and Vartiainen, 2006), we diverse jobs in industry-occupation cells by three levels of a firm's hierarchy: entry-level, middle-level and high-level. Taking sales in the Internet industry as an example, sales clerk is the entry-level job, sales manager is a middle-level job, and sales director is a high-level job. The details of these job types and

[22] Our industry-occupation cells are quite narrow; in fact they refer to what the job boards call sub-industry and sub-occupations. These 'sub' categories are the ones workers generally use to set up their profiles.

[23] Information on the predominant gender in job types was calculated from platforms' annual reports, which include the share of female workers working in each industry and occupation based on the resumes in the platform.

the related characteristics of our fictitious workers are described in Appendix A1.

3.3 Resume Setup

We next created resumes that are qualified for the above jobs. The fictitious resumes come in pairs, and the two workers in each pair are identical except for gender. To increase our profiles' relevance and realism, the resume information was generated from real job ads and resumes. For each job type, we scraped 50 job ads and 50 resumes as the information pool for fictitious profiles. Compared to a typical audit study, our resumes are quite sparse and contain only the mandatory information that is required to set up a worker profile. This is the information that is most likely to be used by job recommender systems; excluding additional details ensures that recommendation results are not driven by idiosyncratic factors.

A fictitious applicant's resume consists of four parts: personal information, education, job history, and job intention. The personal information section contains the worker's name, birth date, years of working experience, current wage, city, employment status, phone number, and email address. In contrast to most audit studies which rely on workers' names to proxy for gender, the jobseeker's gender (male or female) is a compulsory input to create an account on a Chinese job board. The applicant's names were randomly selected from the most popular names in the 2015 Chinese Census 1% Population Sample, and the first name matches the worker's declared gender (See Appendix A2.1 for more details). Since Chinese employers' gender preferences appear to interact strongly with the worker's age (Helleseter et al., 2020), we created two versions of each matched profile pair—a 'young' and an 'older' version, in which 'older' workers refer to ones who have more working experience.

The fictitious workers' age, education and working experience are jointly determined. Young workers graduated in 2017, have three years of working experience, and are 25 years old (born in 1995) if they have a college degree (which takes three years to achieve), or 26 years old (born in 1994) if they have a bachelor's degree, which takes four years to achieve. The older workers are 35 or 36 years old respectively, with 13 years of working experience. The workers' education levels and academic majors satisfy the requirements of the job type that is advertised, and the school's name is randomly drawn from the Chinese High Education Institution List.[24] All the applicants are currently employed, and their wages are selected to match the wages of existing job seekers by job type, education level, and years of working experience using data from real resumes posted on each site. As over half of the job postings on our four job boards are from China's four first-tier cities, we restrict the location of applicants to those cities, specifically Beijing, Shanghai, Shenzhen, and Guangzhou to simplify the resume creation process. Each applicant has a unique and active email address and mobile phone number.

In terms of job history, young workers started their current jobs in August 2017, just after they graduated with the highest degree. For older workers, the start date of their current job is August 2015, implying that they have 5 years of tenure in their recent position. Each worker's current occupation and industry are the same as the job type's occupation and industry, and their job title and job description match the job type's occupation. To minimize the disturbance to real employers and job seekers, all the fictitious workers' current employer names are fictitious. Each company name is a combination of the worker's city, industry and a randomly generated name, as in "Beijing Dongya Internet Technology Company". In the "desired job" section, a worker's desired wage is 120% of their current wage, and the desired city, industry and occupation are aligned with current ones.[25] Appendix A2 summarizes the details of resume generation process.

To sum up, we created groups of four resumes that vary along two dimensions, gender and age, with all the other characteristics and information held constant or randomized, except that the older resumes' experience and current wages are adjusted to be age-appropriate. Given that the four workers in each group are designed to have the same job type and 35 job types are selected in each job board, we created 560 fictitious profiles in

[24] More details on fictitious workers' education information are provided in Appendix A2.2.

[25] According to the salary reports from the job boards, 20% is normal and moderate wage growth for an average worker switching to a new job.

four cities on each platform.[26] After finishing the profile-creation process, the male and female applicants in each pair published their profiles at the same time. This made their resumes accessible to recruiters and head hunters on the site, and immediately generated a first set of job recommendations for each job seeker.

3.4 Implementation

We harvested data from our fictitious profiles in four stages, the first two of which (Rounds 0 and 1) occur immediately after the profiles were created. Rounds 2, 3, and 4 occurred at two-week intervals after that. In more detail:

- **Round 0**. The male and female workers with newly created resumes log into their accounts at the same time, and we collect the first 20 job ads shown to each worker. Then the workers log off.
- **Round 1**. Two weeks later, the male and female workers simultaneously log into their accounts again, and we record the top 10 jobs in their recommendation lists. The two workers then *apply* to these top 10 recommendations by submitting their resumes. Immediately afterwards, the workers refresh their web pages and we record the top 10 recommended jobs that appear. We also record the number of *views* of the worker's resume by hiring agents since the worker's account was created.
- **Round 2**. After two weeks, we repeat the Round 1 procedures.
- **Round 3**. After two more weeks, we again repeat the Round 1 procedures again.
- **Round 4**. After two more weeks, the male and female workers log into their accounts at the same time, and we collect a final count of cumulative views for each resume.[27]

[26] As noted, some of the identifying information in these profiles was customized to reflect the city in which the job was located.

[27] After a worker's resume becomes public, it can be viewed by all recruiters on the job boards. Recruiters of the applied jobs can read applicants' profiles, and other recruiters can find workers by searching resume, or by worker recommendations from job boards. The number of views indicates how many times that the resume is read by hiring agents.

Figure 1 demonstrates the timeline of this experiment. Ideally, each fictitious worker received 160 recommended jobs and applied for 30 jobs in an 8-week job searching spell, and the collected outcomes include the information of 160 jobs as well as the number of hiring agents' profile views during that 8-week spell.

Notably, starting in Round 1 the workers in our experiment apply for jobs in a naive fashion: They mechanically apply to the top 10 jobs that were recommended to them each time they log on. By holding all workers' application strategies constant, this design guarantees that any observed gender differences in the job recommendations are caused solely by our randomized assignment of genders to identical resumes. We decided to follow our workers over several rounds of applications because we suspect that job recommender systems use workers' browsing and application behaviors to deliver increasingly customized recommendations. Our experiment allows us to see the consequences of following the board's recommendations over a period of time.

Compared to our naive workers, real workers' application strategies could either mitigate or accentuate the amount of gender bias we detect in later rounds of our experiment. Workers who are searching for gender-atypical jobs may ignore the stereotypical recommendations they receive; if their board's algorithm learns from their past application behavior the next recommendations they receive may be less gender-typed than the ones we see. On the other hand, workers seeking gender-typical jobs may see an increasingly stereotypical set of job ads that reflect their past choices. In this case, the recommendations received by our workers in the later rounds of our experiment will probably be more homogeneous (and more gender-typed) than those received by real workers. Still, it is of interest to see the recommendations received by a worker who followed the board's recommendations over a period of time. That said, the 100 job recommendations received by each worker in Round 0 are not preceded by any applications. These recommendations give us clean estimates of how the algorithms treat identical workers who also have identical application histories.

4 Results

Our audit study of job recommendation algorithms started in July 2020 and the last collection of hiring agents' profile views was completed in April 2021. In total, 2,240 fictitious profiles were created on four job sites, and those workers received 177,108 job recommendations from 77,802 individual job advertisements.[28]

Table 1 presents descriptive statistics for our sample of fictitious *workers*. As applicants are paired and have fixed characteristics, the information in Table 1 reflects the levels we have assigned, based on averages taken from real resumes on the job boards. The average annual wage of our resume sample is 142,507 RMB, which is around twice the 2020 average wage in urban China.[29] The worker's desired wage is 26.1% higher than the current wages, and the average years of education is 15.56, indicating that about half of the fictitious workers hold a bachelor's degree.[30]

The characteristics of the job ads that were recommended to our fictitious workers are summarized in Table 2. Over 95% of recommended jobs posted a wage (or wage range), and one-third of the recommended positions are from companies that have more than 1,000 employees.[31] The average posted wage in recommended jobs was of 212,611 RMB; mean requested years of education and experience were 14.9 and 2.3 respectively. On average, this posted wage was 18.3 percent higher than the fictitious workers' desired wages;

[28] There are several reasons why the recorded number of job recommendations is smaller than the designed number 2,240*80 = 179,200. One reason is that job boards froze suspicious workers' accounts and a few of them were blocked after the resumes were open to the public. If one account in a gender pair was blocked, we terminated the experiment for the whole gender pair. Another reason is some job links were blank and we were unable to scrape detailed information in job ads. The missing data is less than 0.5% and occurs randomly; importantly, it is independent of the gender of fictitious applicants. Thus it is unlikely to bias our analysis.

[29] According to the statistics from National Bureau of Statistics of China, the average annual wage of workers in the urban non-private sector in 2020 was 97,379 yuan (US$15,188), and workers in the urban private sector had an annual wage of 57,727 yuan (US$9,004).

[30] Some job boards let the worker choose a desired wage *range*; in these cases the desired wage is the midpoint of the range we assigned. Typically, students who earn a Bachelor's (college) degree have 16 (15) years of education.

[31] While some empirical evidence suggests that better jobs (i.e. higher requirements on education and experience) are less likely to explicitly post wages (Marinescu and Wolthoff, 2020), this is not the case in our data.

the requested years of education and experience were lower than the applicants'. Overall, the jobs recommended to our fictitious workers are well-matched with those workers, as shown in Table 3. In around 90% of cases, the workers satisfied the jobs' education and experience requirements, and almost all of the recommended jobs' locations aligned with the worker's current location. 87.9% of recommended jobs posted wages that were higher than the workers' lowest desired wage.[32]

4.1 Differences between the Job Recommended to Men and Women

This section answers the most basic question about gender bias in job recommendations: To what extent are the jobs recommended to identical male and female workers the same ones, or different ones? While the variation of job recommendations to male and female workers does not necessarily indicate bias because it can result from the randomness in websites' recommender systems, it is still worth quantifying the gender-specific jobs in job recommendations before we look into the details of jobs' quality in Section 4.2 and Section 4.3.

We measure the difference as the share of jobs that are only recommended to one gender, without considering the sequence of recommended jobs.[33] Figure 2 demonstrates the difference: Suppose for two workers in a gender pair, male applicant receives jobs that are in set A and C, and the female applicant is recommended to jobs in set B and C, in which set C contains the overlapped jobs of female and male recommendations, while set A represents the only-to-male jobs, and set B includes the only-to-female jobs. Then the difference rate is defined as the share of only-to-one-gender jobs on the whole pool of recommended jobs received by the pairwise male and female applicants:

$$Difference\ Rate = \frac{\#\ jobs\ in\ A + \#\ jobs\ in\ B}{\#\ jobs\ in\ (A+C) + \#\ jobs\ in\ (B+C)} \quad (1)$$

[32] Table A3 presents the descriptive statistics of job recommendations by the applicants' gender.
[33] We provide the discussion of the ranking difference in recommendations in Appendix B.

We present the average difference rate by the worker's age, by the job's gender type, by the job's skill level, and by city in Table 4. These numbers are based on all four rounds of recommendations received in rounds 0 - 3 of the experiment.[34] In total, the difference rate between male and female applicants is 12.30%, meaning that on average, out of 100 jobs recommended to male and female applicants, 87.7 jobs are displayed to all applicants, and 12.3 jobs are unique to one gender while applicants with the opposite gender cannot see those jobs in their recommendation lists.

Table 4 breaks down the 12.30 per 100 overall gender difference rate by applicant age and city, and by two job characteristics: the predominant gender in the job type (Female, Neutral, or Male) and the job's hierarchy level (Entry, Middle, and High). While we find little variation across age and city, the significant differences are detected in jobs' predominant gender and hierarchy level.

In the case of predominant gender, male and female applicants working in gender-neutral jobs observe about 1 additional different job per 100 recommended jobs, compared to workers in male- or female-dominated job types. This pattern is somewhat surprising, however, for the following reason. If job recommender systems reinforce existing patterns of gender segregation by recommending jobs to workers that match the gender of the majority in their occupation/industry, we would expect the difference in job recommendations to be greater in both male-and female-dominated jobs than in non-gendered jobs (Such an algorithm would tend to shift minority-gender applicants into jobs that are more 'popular' among their gender; it would not do such shifting in gender-balanced jobs), but we do observe the opposite pattern in our data. Figure 3 examines the effect of majority gender in more detail by asking whether the majority-gender effect differs between young and older workers, and between entry, middle and high level jobs. The tendency for the difference to be slightly higher in gender-neutral types of work is present within most of these subgroups as well. In terms of the job hierarchy variation, gender-specific jobs appear the least frequently in entry-level jobs, and the most in middle-level jobs.

[34] Disaggregated results are shown in Figure 4.

Figure 4 displays the dynamics of the difference rate. In Round 0 (before workers apply to any jobs), the share of gender-different jobs is 8.91 percent. Two weeks after the release of workers' profiles, the share increases to 13.09% in the first 10 job recommendations in Round 1. After workers apply to those 10 jobs, the difference rate further rises to 14.93% in the top 10 jobs that are displayed to workers. In the following rounds, we observe the same trend such that the share rises within as well as across rounds as workers make additional applications (which follow the top recommendations they previously received). In the last 10 jobs of Round 3, the chance that an applicant will be shown a gender-specific job is more than doubled relative to the share in Round 0, at 21.47 percent. In Section 5 we explore the implications of this pattern for the types of algorithms the job boards are likely to be using.

In addition to the number of only-to-one-gender jobs, the ranking of jobs in the recommendation lists can affect workers' decision on where to apply because jobs displayed at the top receive more attention, and are more likely to be seen and clicked into by workers (Craswell et al., 2008; Richardson et al., 2007). We address the ranking difference in job recommendations between male and female applicants in Appendix B. In general, the ranking difference rate has a quite consistent pattern with the difference rate in Table 4 across the subsamples by age, job's gender type, hierarchy level and city.

4.2 Differences in Job Characteristics: Wages, and Education and Experience Requirements

As shown above, job recommendations to male and female workers are not the same. If the dissimilarity results from computation randomness, we should expect the gender-specific jobs are equally good to male and female workers. However, if systematic gender bias actually exists in job recommendations, jobs recommended to one gender will have different characteristics from the jobs shown to the other gender. To begin to address this question, this section compares the observable characteristics (wages, education requirements, and experience requirements) of the jobs recommended to men versus women.

To make these comparisons, the analysis sample in this section is restricted to job recommendations unique to the male applicant (i.e. Set A in Figure 2), plus job recommendations unique to the pairwise female applicant (Set B). The recommendations that are common within each gender pair (Set C) are excluded.

We examine whether the average characteristic of male-only jobs differs from the average of female-only jobs by estimating the following specification:

$$Y_{pj} = \beta_0 + \beta_1 M_{pj} + \beta_2 X_p + e_{pj} \tag{2}$$

where Y_{pj} is the characteristic (i.e. posted wage, education requirement and experience requirement) of job j received by the applicants in gender pair p. The variable of interest is M_{pj}, which takes the value of 1 if the recommended job j is only seen by the man in gender pair p. We control for the gender pair fixed effect X_p, so β_1 estimates the average gender gap (male-female) in the characteristic between male-only and female-only recommendations within gender pairs.

In Table 5, column 1 to 3 presents the results with the regression outcomes of job's posted wage, requested years of education and requested years of working experience, respectively. Conditional on the wage being advertised publicly, wages in jobs recommended to men are 2,616 RMB higher than in jobs recommended to women on average. This difference amounts to 1.8% of the average current wage of fictitious workers, and is statistically significant at the 5% level.[35] The requested education is statistically indistinguishable between male-only and female-only jobs, but the required working experience in male-only jobs is significantly higher than the requirement in women-only jobs by 0.17 years (or about two months), which corresponds to 2.1% of the average worker's working experience, and 7.5% of the average requested working experience in recommended jobs.

Combining these small differences in posted wages and experience requirements with the fact that about 88 percent of the jobs recommended to men and women are the same

[35] Instead of an exact wage, most jobs posted a wage range. The job's wage in the analysis is the midpoint of posted wage range.

jobs, it appears that the overall difference in the quality of jobs recommended to identical men and women on these job boards is quite modest in size. As we argue below, however, greater differences appear in the job characteristics, skill types, desired personality characteristics, and desired worker attributes in jobs recommended to women versus men.

Figure 5 explores how the gender gaps documented above vary across the experimental rounds, and with other observed characteristics like the worker's age, the job's gender type, and its hierarchy level. According to Figure 5(a), the gender wage gap does appear to increase across successive experimental rounds, except for Round 3, but remains insignificant for all individual rounds. While the recommended wage gap does not differ between young workers and older workers, between female-dominated, gender-neutral and male-dominated jobs, the posted wages in male-only jobs in entry- and middle- level are significantly higher than ones in female only-jobs, and the greatest gender wage gap emerges is entry-level jobs is 3905.2 RMB, which is 4.0% of the current wage of workers whose jobs are in the entry level. Figure 5(b) shows that the differences in education requests of recommended jobs are positive and significant in gender-neutral jobs, entry- and middle- level jobs. In Figure 5(c), the higher requirement on working experience in male-only jobs is significant in almost all subgroups, and more pronounced in older applicants, in gender-neutral jobs, and in middle- and high-level jobs.

4.3 Learning from Words: Which Words are Over-Represented in Jobs Recommended to Men versus Women?

We now begin to explore the unstructured text in the *job description* section of the ads that are recommended to workers. In this Section our goal is a straightforward one– to identify words or phrases that are statistically over-represented in job ads that are only recommended to women, compared to ads that are only recommended to men. We measure the extent to which these over-represented words are associated with widely held gender stereotypes in the following Section.

In a typical job ad, the job description section is one or two paragraphs of text, which is placed after the explicit characteristics of jobs and contains rich information about the position. While the contents of job descriptions are highly diverse, they can be broadly aggregated into *six* categories:

(*1*) *Standardized* (*PIACC*) *Skills.* Skills are the core part of most job descriptions, and recruiters express skills in various ways. While a variety of methods have been developed to categorize the many skill requests that appear in job ads, we adopt the skill classification of the OECD's Programme for the International Assessment of Adult Competencies (PIAAC) (OECD, 2016), which has been used in a wide variety of research contexts, including gender skill differentials (Christl and Köppl-Turyna, 2020; Pető and Reizer, 2021). PIACC skills are divided into seven subsets, which focus on literacy, numeracy, information and communication technology (ICT), problem-solving, influencing, co-operation, and self-organization.

(*2*) *Benefits.* In addition to offered wages, employers frequently advertise jobs' benefits to attract applicants. In Chinese job boards, commonly advertised benefits are often tagged, and their expressions are quite uniform across job types and platforms. Based on the information we have extracted information from job ads, we classify job benefits into four types: payment, break, facility and insurance.

(*3*) *Work Timing and Location.* Job ads frequently contain information about work time arrangements, using words about work schedules, the need to travel for work, breaks, and overtime.

(*4*) *Work Environment.* Job ads provide information on both the position and the company. These include the specific workplace environment, the company type, and job title.

(*5*) *Other Qualifications.* Other desired qualifications that frequently appear in the text of job ads include a desire for a specific college major, overseas work experience, and specific types of work experience.

(*6*) *Personality, Age, and Appearance.* Chinese job ads frequently indicate a desired age range for the workers they are seeking. Requests for a variety of personality attributes (such as "innovative" and "careful"), and for an attractive physical appearance are also quite common.

Based on the above structure, the information in job descriptions was extracted in the following way: For all the jobs collected from four job boards, we first segmented a chunk of text into words, and retained only the high-frequency words. [36] Next, we combined the words that have the same or close meaning (e.g., leadership vs leading) to make the remaining words clearly contrast with each other, and assigned them to one of the six categories above. In total, we extracted 172 words from job ads, listed in Appendix Table D1.

Figure 6 presents the word cloud of job descriptions, with a a larger size representing a higher frequency of words in job ads. Words related to job benefits, such as insurance, vacation and payment scheme, are the most common ones in job descriptions, but employers also frequently ask for communication skills, coordination skills, teamwork skills and leadership.

If the job recommender systems used by our job boards are gender-neutral, the words listed above should appear with roughly equal frequency in the jobs recommended to the identical male and female job profiles in our experiment. In contrast, if gender stereotyping exists, there should be clear differences. We begin our investigation of this question using the same specifications in equation (2), but replacing the outcome variable with the dummy for the appearance of word in the gender-specific job recommendation j in pair p.

We run the regression for all 172 words in the six word categories described above, and display results in Table 6. The coefficients in parentheses represent the estimates of β_1, and a positive coefficient means that jobs recommended to male workers are more likely to contain that word in their job descriptions than female workers' jobs. The left panel lists 35 female words, which have a higher probability of being included in female-only jobs

[36] The vast majority of 'words' in this procedure were single words like "flexible", "leadership" and "data", but we also include some frequently-occurring compound words like "medical insurance", "tier-one school", and "no crime history".

at 5% significance level, and the right panel includes 32 male words that are significantly mentioned more in male-only jobs.

Starting first with the standardized (PIACC) skills, we can see that literacy skills, such as *listen*, *writing*, *speak* and *documentation*, and co-operative skills such as *cooperation*. *communication*, and *negotiation* are more common in only-to-female jobs. Furthermore, female applicants are more likely to see job ads mentioning *data*, *chat tools administrative tasks* and *collect*, while male applicants see more jobs that require *problem-solving skills*, such as *planning*, *decision-making*, *engineering*, and *working independently*, and *influencing skills* such as *leadership*, *charge* and *supervise*. These findings coincide with the results from previous literature on the gender gap in skills that document women tend to carry out more executed tasks, fewer skill-intensive tasks, and use their cognitive skills less than men (Pető and Reizer, 2021; Black and Spitz-Oener, 2010).

Turning to the *Work Timing and Location* panel, jobs with *regular working hours*, *eight-hour working*, *weekly break* or *flexible* schedules are more likely to be recommended to women, and jobs with decreased flexibility, such as *overtime working*, *night work* and *long travel*, are more likely to be recommended to men. This is in line with the finding that women are more willing to pay for flexible work arrangements (Flory et al., 2015; He et al., 2021; Mas and Pallais, 2017; Bustelo et al., 2020).

Turning next to *Benefits*, female-only jobs are more likely to mention *marriage leave*, *maternity leave*, *parental leave*, *social security*, *maternity insurance* and *medical insurance* in their descriptions while only-to-male jobs emphasize *commuting friendly* and providing *shuttle*, *commission*, *allowance*, *free meal*, *reward* and *stock*. Under *work environment*, *training* and workplace *atmosphere* is mentioned more frequently in female-only jobs, while jobs from *publicly-listed* companies are more frequently recommended to men.

With respect to *Other Qualifications*, jobs recommended to women are more likely to request *new graduates*, workers *without working experience* and workers who have *certificate*. Only-to-men jobs request workers who have *science and engineering backgrounds* and *no crime history*.

Finally, under *Personality, Age, and Experience*, jobs recommended to men request workers who are *self-motivated*, *innovative*, *entrepreneurial*, and able to handle work *pressure*. Jobs recommended to women are more likely to mention *punctual*, *patient*, *careful*, *active*, *outgoing*, *temperament*, and *generous*. Words associated with physical appearance, such as *figure*, and *facial* are also more common in only-to-female recommendations.

Figure 7 presents the number of male and female words in subgroups by rounds, worker's age, and job's gender type and level. The overall trend is quite consistent with the difference rate, and in the groups that a have higher share of gender-different jobs, we find more male and female words in those gender-specific jobs (for example, Round 0, Young group). But the association is reversed in the jobs' gender type: While applicants in gender-neutral jobs see more gender-specific jobs than in one-gender-dominated jobs, they see fewer words that show up differently between those male-only and female-only jobs.

4.4 Learning from Words: Relating Over-Represented Words to Gender Stereotypes

In the preceding Section, we established that the jobs recommended to identical male and female job-seeker profiles contained systematically different groups of words. But in what sense, if any, do these words reinforce gender stereotypes? In this Section we exploit three data sources that are external to our job boards to assess which, if any, of the over- and under-represented words identified in the last Section are, in fact, associated with widely-held gender stereotypes.[37]

The first external data source we use is the previous literature on gendered words, which identifies masculine and feminine words that are widely associated with gender

[37] Previous research has shown that the wording in job advertisements frequently reflects commonly held gender stereotypes and affects how readers react to ads. For instance, women found jobs less appealing when the job advertisements included more masculine wording (Gaucher et al., 2011), while feminine wording increases the share of female applicants (Kuhn et al., 2020; Chaturvedi et al., 2021).

stereotypes. While linguists studying stereotypes focus on commonly used words in daily life and the effect of gendered words on people's behaviors (Fitzpatrick et al., 1995; Gastil, 1990; Lindqvist et al., 2019), researchers in political science (Roberts and Utych, 2020) and in psychology (Bem, 1981; Hoffman and Hurst, 1990; Rudman and Kilianski, 2000) identify gendered words in different contexts and argue that their use can shape people's support for policies. Most of the gendered words identified by these literatures are adjectives, which describe men's and women's personalities (for example, confident, aggressive, and strong are masculine words, while sensitive, kind, and beautiful are feminine words).

More relevant to our context, three papers have encoded the gendered words most frequently used in job advertisements. Gaucher et al. (2011) collected masculine and feminine words from published lists of agentic and communal words, and masculine and feminine trait words, and showed that these words affected readers' perceptions of gender representation in jobs. Kuhn et al. (2020) and Chaturvedi et al. (2021), on the other hand, took advantage of the fact that jobs with *explicit* gender requests are still common in many developing countries. This allows the authors to train text analysis and machine learning techniques to predict the effect of observing a particular word in an ad on the probability the ad explicitly requests only male or female applicants. The words with the largest predictive power often refer to worker's personalities and required skills. In more detail, Kuhn et al. (2020) apply the naïve Bayesian classifier to identify the likelihood of an explicit gender request based on the words in job titles in a Chinese job board, and Chaturvedi et al. (2021) make use of the text contained in detailed job descriptions in India and construct measures on whether the job ad text is predictive of an employer's explicit male or female preference using a multinomial logistic regression classifier. Our first external list of male and female words simply combines all the male and female words identified in these three studies of job ads.

Our second and third approaches are based on two surveys we conducted to collect people's perceptions about stereotypically male and female words in job ads. In the English survey, we recruited participants from Amazon Mechanical Turk (MTurk), and let them rate words on maleness and femaleness. A corresponding Chinese version was conducted on Chinese workers in a survey platform, wenjuanxing. com, which provides professional online questionnaire survey service. In both surveys, we asked the following question about every word extracted from job descriptions: "Suppose you are a recruiter and you craft a job advertisement containing the following word, would you tend to hire (a) no gender requirement, (b) men, (c) women?". In this setting, respondents would perceive the most likely gender of the candidate for jobs that are worded in a masculine or feminine way. Details on the surveys are provided in Appendices D2 and D3.

The heat map in Table 7 displays the stereotypical maleness and femaleness of words, as determined by these three approaches: previous literature, our Mturk survey, and our Chinese survey. The words listed are the male and female words that were identified by our quantitative analysis of job recommendations (from Table 6), while the color intensity represents those words' femaleness or maleness as defined by our three external data sources. If a word is highlighted with bright red, it is defined as a female word in all three approaches. Words in light red are defined as female words in two approaches, and pink indicates the word was female in just one approach. Male words are marked with blue colors, in which bright blue, light blue and pale blue represent maleness from three, two and one approach, respectively.

Overall, the dominance of red colors in the left panel and blue colors in the right panel clearly demonstrates that the words that are over-represented in identical only-to-male and only-to-female job recommendations are indeed correlated with commonly held gender stereotypes, which associate men with words like *engineering*, *leadership*, and *overtime*, and women with words like *assist*, *administrative*, *patient* and *facial* features. In other words, the algorithmic job recommender systems used by these job boards recommend different jobs to identical male and female job seekers in a way that reinforces commonly held gender stereotypes.[38]

We use four different approaches to quantify the statistical significance of the associations described above. The first and basic method is OLS regression applied to the

[38] This is consistent with results from Chaturvedi et al. (2021), who found that words related to hard-skills and flexibility are critical in explaining gender disparities in labor market outcomes.

combined sample of only-to-male jobs plus only-to-female job ads. In these regressions, the outcome variable equals 1 if the job was only recommended to male applicants, and 0 if was only recommended to female profiles. The regressors are dummy variables for the presence of 172 words in the recommended job ad. Column 1 in Table 8 lists the top 10 words that are significant at the 5% level, in order of coefficient magnitude. The overall F-test result is $F(172, 16267) = 7.72$ ($p<0.0001$), indicating that the 172 words are jointly significant. As the data matrix for these regressions is large, sparse, and some of the words are correlated with each other, one may want to select variables that have a larger impact on the outcome rather than including all of them. Our second and third methods therefor apply lasso and ridge approaches, respectively, to the preceding regressions. These approaches impose a penalty parameter for adding an extra variable to determine which words contribute most to the diverging recommendations to men and women. We applied 20-fold cross-validation to find the optimal penalty parameters, and the selected top 10 words by lasso and ridge regressions are shown in columns 2 and 3.

Our final method of identifying words that contribute to the classification of jobs recommended to men and women is a random forest approach. Given our binary outcome and independent variables, our data structure is well suited to decision tree methods like random forests that search for the best way to split the sample into two groups– in our case, male-only and female-only jobs. Column 4 in Table 8 presents the top 10 words with high feature importance based on 100 decision trees and Gini impurity.

Overall, Table 8 shows that the three regression-based methods produce similar results, with words about work schedules (regular hours, long travel, overtime, flexible) and benefits (commission and medical insurance) occurring frequently among the top ten. The words identified by the random forest approach do not overlap as much with the other three methods, but emphasize similar considerations, including vacation, allowance, commuting, reward and shuttle.[39]

[39] In Appendix C, we measure the word dissimilarity in job recommendations by different groups based on the vector distance.

5 Explanations for the Gender Gap in Job Recommendations

Section 2 described four mechanisms that could create a gender gap in jobs recommended to identical male and female workers. In this Section, we search for patterns in our data that would suggest the presence, or absence, of each of these mechanisms in our job boards' algorithms.

One pattern in our data that is consistent with a role for *item-based collaborative filtering* in accounting for the gender gap in job recommendations is the fact that the jobs recommended to our male and female profiles become more different across rounds of our experiment, as shown in Figure 4 and Figure 5. This is consistent with IBCF because workers' application histories –which follow the board's previous recommendations– also diverge across the rounds. In other words, IBCF could explain why the small initial recommendation gap in Round 0 -which cannot be explained by application histories– becomes magnified over time.

In more detail, Figure 4 has already shown that the difference rate rises after applicants send out profiles in Round 0, and within Round 1 to Round 3, the second 10 applications' difference rate is on average 9.63 percentage higher than the one in the first 10 applications, implying that the pairwise men and women would see nearly one more gender-specific job after they apply for the first 10 recommended jobs.

While these patterns are consistent with a role for IBCF, they could also be caused by two other factors: randomness and diversification processes. Specifically, since the data used by algorithms is updated continuously, split-second differences in the times at which the paired male and female profiles are posted, or the recommendations are collected, could introduce random differences in recorded recommendations. Diversification of recommender systems –which increases workers' exposures to a larger spectrum of jobs than the application histories – could have a similar effect.[40]

[40] Diversification in recommendations are developed to solve the overfitting problem of recommenda-

We further check the divergence in jobs that are recommended to men and women by splitting our sample into the first 10 jobs and the second 10 jobs in Round 1 to 3, and compare the mean *characteristics* of the gender-specific jobs between the 10 jobs within each round. Figure 8 shows that the gender gaps in recommended job characteristics, especially working experience, grow after workers apply for the first 10 jobs. This also applies to the words used in job ads: In Figure 8(d), the total number of male and female words (words that are significantly correlated to male-only and female-only jobs), increases from 55 in the first 10 jobs to 57 in the second 10 jobs in Round 1 to Round 3.

Moreover, because we record the 10 job recommendations (applications 11-20) just after applying the first 10 jobs (applications 1-10), hiring agents' behaviors on workers' profiles can hardly vary during this short period of time. That is to say, the above comparisons are conducted conditional on the recruiters' reactions, and the gender differences between the first 10 and the second 10 jobs within each round should be purely driven by IBCF. Later in this Section, we consider direct evidence concerning presence of recruiter-behavior effects, and we conduct one test that should distinguish between IBCF and recruiter-behavior-based drivers of the increasing gender gap in job recommendations across experimental rounds.

Turning next to *content-based* recommendations, if these play a role in the gender recommendation gap two things must be true: First, the unstructured text of job descriptions should affect job recommendations, and second, this effect should occur even in the absence of any application history, or recruiter reactions. For the pooled sample of recommendations across all rounds, Table 7 has already shown that the boards are disproportionately recommending jobs containing female words to women, and jobs containing male words to men. For instance, *facial*, *patient*, and *assist* are overrepresented in jobs recommended only to women, while *engineering* and *leadership* are overrepresented in jobs recommended only to men.

tions, and becomes one of the most important topics in recent research in recommendation algorithms. It addresses that accuracy-related metrics are insufficient to measure recommendation quality, and recommendation systems should not only make relevant but also diversified recommendations to improve overall user satisfaction (Szpektor et al., 2013; Wu et al., 2016; Kunaver and Požrl, 2017).

Notably, this is happening even before identical men and women start applying to jobs, and before any recruiters have had a chance to react to these workers' resumes. Table D4 in Appendix D presents the heatmap in Table 7 but only based on the gender-specific jobs in Round 0. 6 out of 9 female words and 8 out 13 male words that have are significantly associated with only-to-female only-to-female and only-to-male jobs in Round 0 reflects workplace gender stereotypes.

Third, a *rule-based* approach that complies with employers' stated gender requests probably has a very limited effect on gender-biased recommendations. While we cannot observe the preferred gender from public job ad postings, recent studies show that number of jobs advertised specifically for men or women in China have declined dramatically due to the recent policy interventions (Kuhn and Shen, 2021). In Kuhn and Shen (2013), jobs that specified desired gender accounted for about 10.5% in Zhaopin.com in 2008, and Kuhn et al. (2022) suggests that the share was lower than 1% in Liepin.com in 2018. Moreover, if the gender requests still exist, they are more likely to appear in the fields that are dominated by one gender, thus we expect to find greater gender bias in male- and female-dominated jobs. However, when we disaggregate the difference rate differences in measured job characteristics, there is no strong evidence that applicants in male- or female-dominated jobs received more gender-specific job recommendations, or that the jobs recommended to men and women were more different in terms of wages, experience requirements, or education requirements.

Finally, we now consider the likely role of *recruiter-behavior* based algorithms in creating a gender gap in recommendations, by using data on the number of times each of our profiles was *viewed* by hiring agents. Although our sparse profiles rarely receive callbacks from employers, they are viewed quite frequently by hiring agents, and each profile receives 23 views from hiring agents on average.[41] Thus, if a resume has been viewed by a lot of agents, an algorithm could use that information to target its job recommendations more precisely.

[41] These recorded profile views include views of resumes that have applied to a job, as well as views by hiring agents who found the worker's resume using a board's resume search function. Thus it is possible for resumes to be viewed even before they have applied to any jobs.

As a first approach to this question, we use data on the total number of views two profiles in a gender pair received during the entire 8 weeks of our experiment. We use this as a proxy for the *amount of information* that was available to a potential algorithm from recruiter reactions to the two profiles, and check how this indicator of information affects the amount of gender-specific jobs recommended to workers by estimating the following regression on the gender pair level:

$$Y_i = \beta_0 + \beta_1 ViewT_i + AX + e_i \quad (3)$$

The outcome variable Y_i is the number of gender-different jobs per 100 recommendations in gender pair i (100*difference rate), and the variable of interest $ViewT_i$ is the total number of views on the identical male and female applicants in gender pair i.

Table 9 reports the regression results. Column 1 only includes the hiring agents' views on gender pairs. Column 2 and 3 add controls for the worker's age and the job's gender type. In column 4, we further control for the job board fixed effect to absorb various behaviors of hiring agents in different job boards. The estimations show that the views from hiring agents are a significant contributor to the quantity of different jobs seen by identical men and women, although modest in size: One more view on the male's or the female's profile will increase 0.03 gender-specific jobs in 100 recommendations.

In Table E1 in Appendix E, we further investigate the effect of hiring agents' reactions on the quality of jobs recommended to only men and women. We adopt the specification in column 4 in Table 9 with controls on profile's age, job's gender type and job board, but replace the outcome variable with the gender gap (male - female) in posted wage, requested education, and requested experience between only-to-male jobs and only-to-female jobs in each gender pair. Moreover, based on the female words and male words derived in Table 6, we count the frequency of these words in gender-specific jobs, and regress the frequency of words (male words + female words) on the total views of hiring agents in column 5.[42] The coefficients of total views on paired applicants' profiles stay

[42] For instance, if a job advertisement mentions assist, careful, leadership in the description, then the total gendered words in this ad is 3 (2 female words, assist, careful + 1 male word leadership. Then the outcome

positive, indicating that the hiring agents' views magnify the gender gap in recommended jobs' characteristics, but they are not statistically significant.

As mentioned in Section 3, the number of views on workers' profiles is recorded before they send out applications in every round. We take advantage of this structure and refine the estimation of the hiring agents' effect on job recommendations by disaggregating the previous tests by rounds. Specifically, the regressor is the total hiring agents' views on the two profiles in a gender pair until Round t ($t = 1, 2, 3$), and the outcome variable is the difference rate in Round t, the gender gap in job characteristics (wage, education and experience) and frequency of male and female words in the only-to-one-gender jobs in Round t. We provide the estimation results in Table E2 in Appendix E. Under the full set of controls of experimental rounds, profile's age, job gender type and job board, the total number of views in Round t increases the amount of gender-specific jobs that the identical men and women see in Round t, but the gender gaps on those jobs' characteristics and wording are not significant, however.[43]

Finally, as already noted, both IBCF and recruiter-behavior-based algorithms share the prediction that job recommendations to male and female profiles should diverge across rounds of the experiment. To attempt to distinguish between these two contributing factors, we take advantage of the fact that there are no applications between Rounds 0 and 1, but there are some reads (because recruiters search profiles on the sites to find applicants). We therefore regress the *change* of difference rate, as well as the *change* of gender gap in recommended job characteristics and male and female word frequency between Round 0 and Round 1, on the number of total views on male and female profiles in Round 1. Estimation results in Table E3 in Appendix E demonstrate that more hiring agents' views on workers' profiles before they send any applications in Round 1 lead to a higher share of only-to-one-gender jobs compared to Round 0, but they does not widen or narrow the gender gap in these gender-specific jobs' characteristics and wording.

variable is the total number of male and female words in gender-specific jobs in each gender pair.

[43] In addition to the gender pair level results, we discuss how the hiring agents' views affect the evolution of jobs received by the individual applicants in Appendix E.

In sum, several patterns in data suggest that three common processes –item-based collaborative filtering, content-based recommendations, and recruiter-behavior based algorithms– all play a role in accounting for the gender gaps in job recommendations we document in our audit study. It is also possible that these three mechanisms interact, leading to additional consequences, both intended and unintended. The simple nature of our experiment, however, does not allow us to quantify these interactions, or even to quantify the relative importance of these three factors in accounting for gender recommendation gaps. We hope that a next generation of algorithm audit studies can be designed to address these issues.

6 Discussion

Personalized recommender systems have become indispensable tools for managing information overload by helping people find items, matches (including friends and romantic partners), and information that suit their individual interests and preferences. Depending on the algorithms on which they are based, however, recommender systems can also have unintended consequences, including information silos, echo chambers, unequal information quality for protected versus unprotected groups, and the perpetuation of stereotypes. Assessing these unintended consequences is challenging for outsiders, because the algorithms used by the most influential web platforms are proprietary black boxes.

In this paper we have adapted a widely used tool in the study of discrimination – the resume audit study– to take a first peek inside these black boxes, by assessing the causal effect of a job seeker's gender on the jobs that are recommended to them on four large job boards. We find that these recommender systems do recommend different jobs to identical male and female job seekers, though most of the recommended jobs overlap between the two genders and the gender wage gap in the jobs that differ is quite small: jobs recommended to women post wages that are 1.9 percent lower.

In contrast, we find that the recommender systems have a much stronger effect in steering men and women to different *types* of work, and to different work environments, which are stereotypically male and female respectively. These stereotypes are present even in the jobs that are recommended before an applicant has applied for any jobs. Further, if job seekers 'follow the board's advice' by applying to the jobs that are most highly recommended, these stereotypes are magnified across rounds of job applications. Together, these processes could make it hard for workers interested in counter-stereotypical types of work to find suitable matches. Finally, we scrutinize the patterns of recommendations our fictitious workers encounter for evidence of the different types of processes that might be used by these boards' recommendation algorithms. We argue that at least three widely used processes –item-based collaborative filtering, context-based recommendations–, and recruiter-behavior-based recommendations– are being used by these recommender algorithms.

As the first resume algorithm audit that we are aware of, our paper has some inevitable, but important limitations. The first of these is that our results are confined to a very early stage of the job search process: Which jobs are suggested to male versus female workers? While it seems likely that displaying different job ads to men and women will also affect the types of jobs they are offered, and where they are hired, we do not observe these longer-term outcomes. Still, such consequences seem likely, due in part to the simple fact that one needs to see an ad in order to apply to it. Extensions of our audit methods that merge it with internal job board data about later stages of the recruiting process might be able to address these issues.[44]

Second, while we have provided evidence suggesting the presence of three common processes –item-based collaborative filtering, context-based recommendations–, we have only scratched the surface of inferring the types of processes used by job board algorithms

[44] Two measures of applicant success could be collected without access to internal data: application reads and call-backs. While we have counts of reads per application, we do not observe the firms or jobs in which they occurred. Future studies could purchase this information –which is available to workers on all four job boards we study– for each fictitious profile that is created. To count call-backs, future studies could design more detailed user profiles and resumes for which call-backs occur more frequently. Our sparse profiles did not generate enough call-backs to make this a valid indicator of worker success here.

from audit studies conducted by outsiders. This is a difficult problem, given the high and increasing complexity of job recommender systems (Hannák et al., 2017), but it is important because the large platforms play important roles in allocating workers to jobs, and because their internal algorithms are proprietary.

Third, after we collect and study the first 100 recommendations received by every worker profile, in this paper we have only studied what happens if job seekers mechanically follow the recommendations they receive by applying to their top ten recommendations. This provides an important benchmark, because it guarantees that the only cause of the subsequent recommendation gaps is the recommendation algorithm (by holding workers' application strategies fixed.) But in reality, there may be important heterogeneity within genders in preferred types of work, and in the extent to which each person's preferences match the stereotypes for their gender. In an expanded study, it would be fascinating to explore how quickly (if at all) the job board algorithms learn this type of heterogeneity.

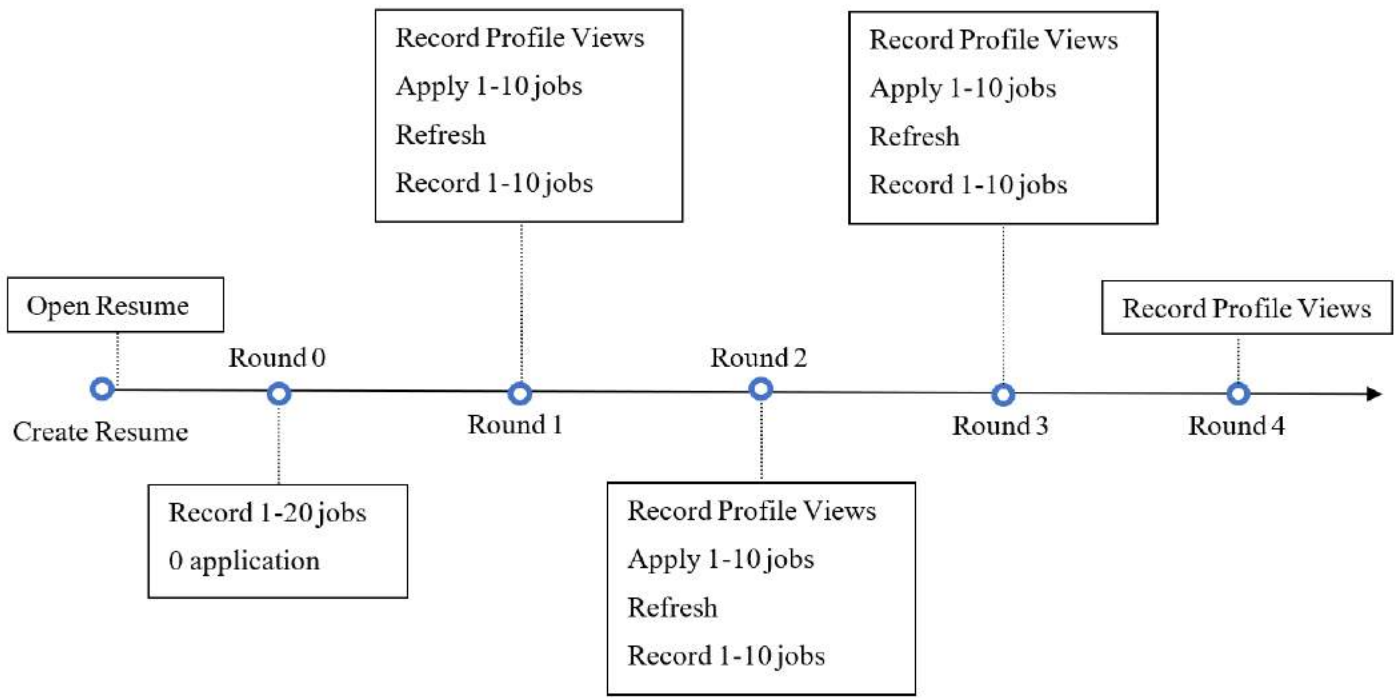
To illustrate, suppose there are two types of women, with high versus low values of hours flexibility, and the 'high-value' type is in the majority; workers' types are private information. Then, before a woman has applied for any jobs, it will be hard for a job board algorithm to know her type. Thus we expect the algorithm to cater to the majority type and assume a high value. (It can only use rules-based and content-based methods since the profile has no application history, and no history of recruiter reactions). To learn how quickly a job board infers an jobseeker's type, one could do an experiment like ours, but with two profiles per gender, which differ in their unobserved types (e.g. one is a woman that wants hours flexibility, the other is a woman that cares more about wages). These two profiles are programmed to make different application choices from the jobs that are recommended to them, with the latter placing more emphasis on non-gender-stereotypical jobs. Observationally, the two profiles should have identical (or equivalent) recommendations before they submit any applications, but these profiles should begin to diverge in a specific way it the algorithm uses application behavior to infer the applicant's type (i.e. if the algorithm uses IBCF).

Fourth, we have not discussed the discussed the types of algorithmic changes that might reduce the stereotyping we detect on these job boards, nor have we discussed whether such changes would themselves have undesirable side effects. For example, if improving a measure of algorithmic fairness involves showing workers many jobs they have almost no chance of getting, is that desirable? In the longer run, how can algorithmic fairness be maintained in a dynamic context where the menu of choice objects (jobs), choosers (workers) and their preferences are constantly changing (Ge et al., 2021)?

Finally and related, we have not discussed the impacts of possible de-biasing algorithms on firms and hiring agents, who are the job board's paying customers. Would de-biasing be costly to employers by sending them workers who are truly less appropriate or less interested in their jobs? On the other hand, to what extent should we allow recommender systems to reflect employers' preferences when those preferences themselves are biased? We hope that future researchers in this area will address all these important issues.

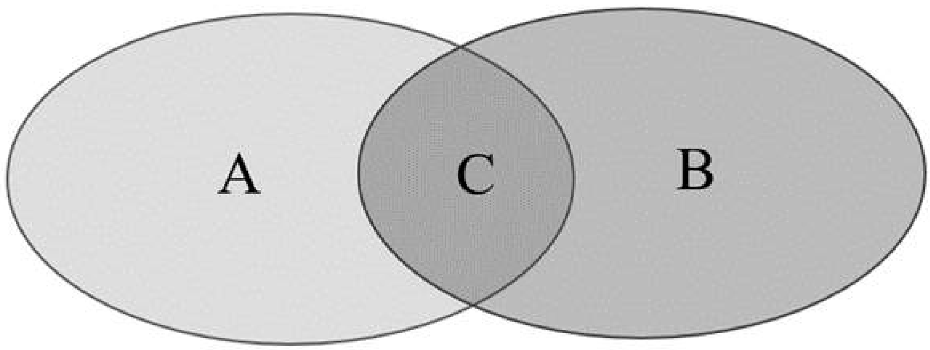
Tables & Figures

Figure 1: Timeline of the Experimental Steps



Note: Two profiles in each gender pair follow the same timeline. From Round 1 to Round 3, fictitious workers apply for the first job to the 10th job that are displayed in their customized job recommendation interfaces, and the time interval for each round is two weeks.

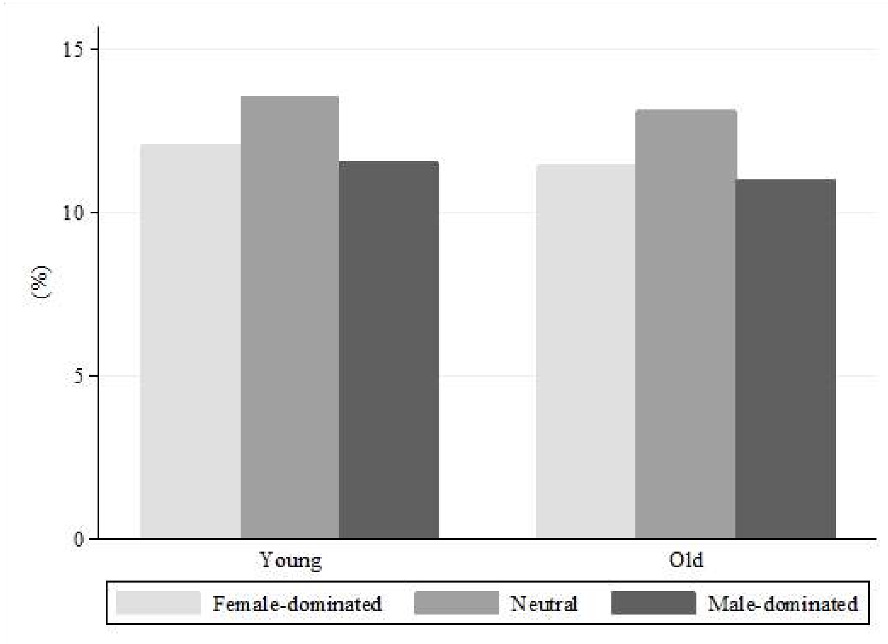
Figure 2: Difference Measure in Job Recommendations



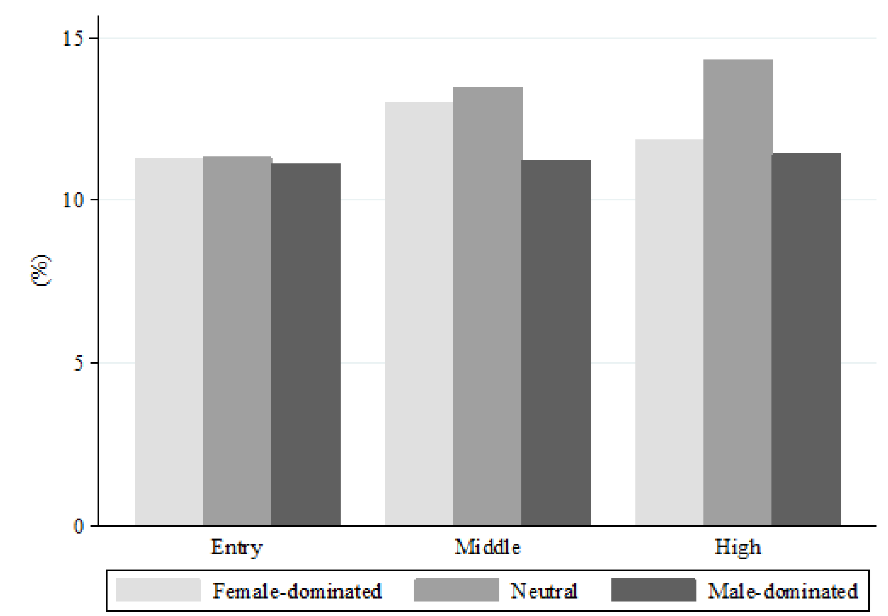
Note: For job recommendations received by two workers in a gender pair, set A represents job recommendations that are only displayed to the male applicant, set B represents job recommendations that are only for the female applicant, and set C represents the jobs that are recommended to both the male and the female. The difference rate is defined as the share of gender-specific recommendations on the total recommendations for the two workers, (A+B)/(A+C+B+C).

Figure 3: Difference Rate by Age, Job Gender Type, and Job Hierarchy

(a) Difference Rate by Job Gender Type and Age

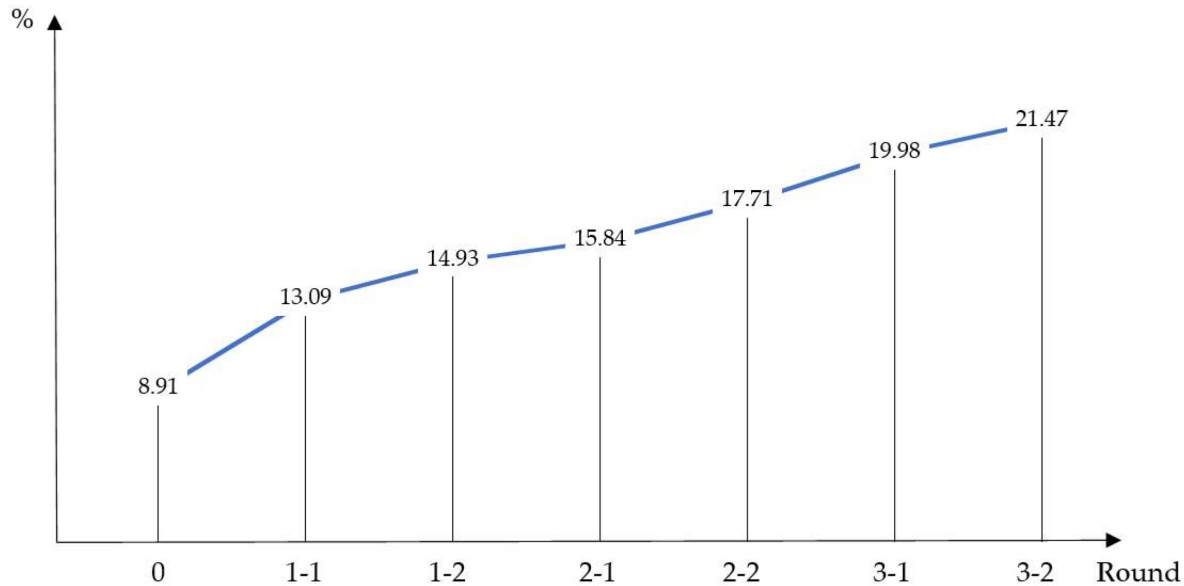


(b) Difference Rate by Job Gender Type and Hierarchy



Note: Difference rate is defined on group level. For instance, the first bar in (a) is the average share of gender-specific recommendations on the total recommendations to young pairs in female-dominated fields.

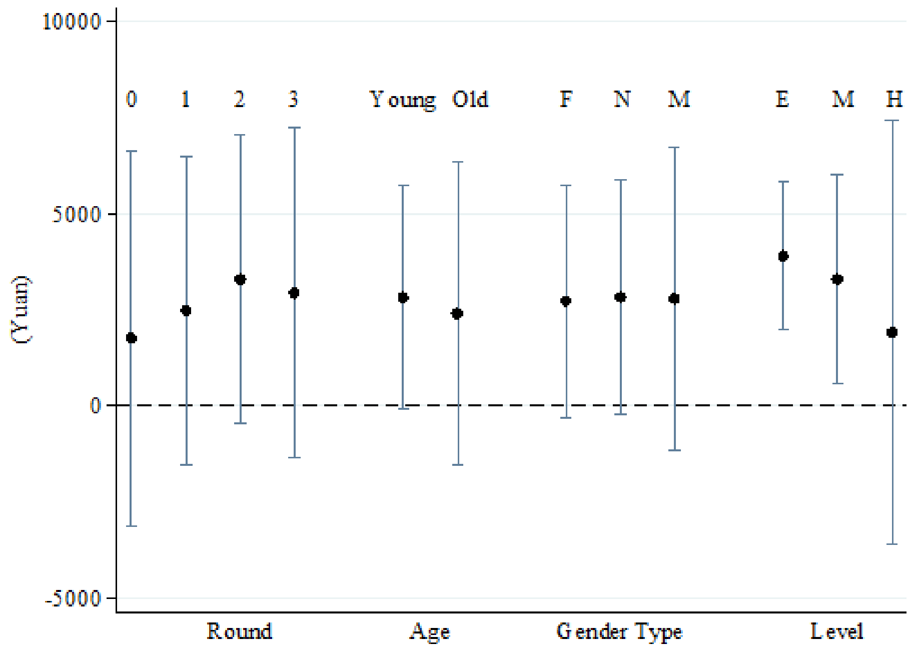
Figure 4: Difference Rate by Experimental Rounds



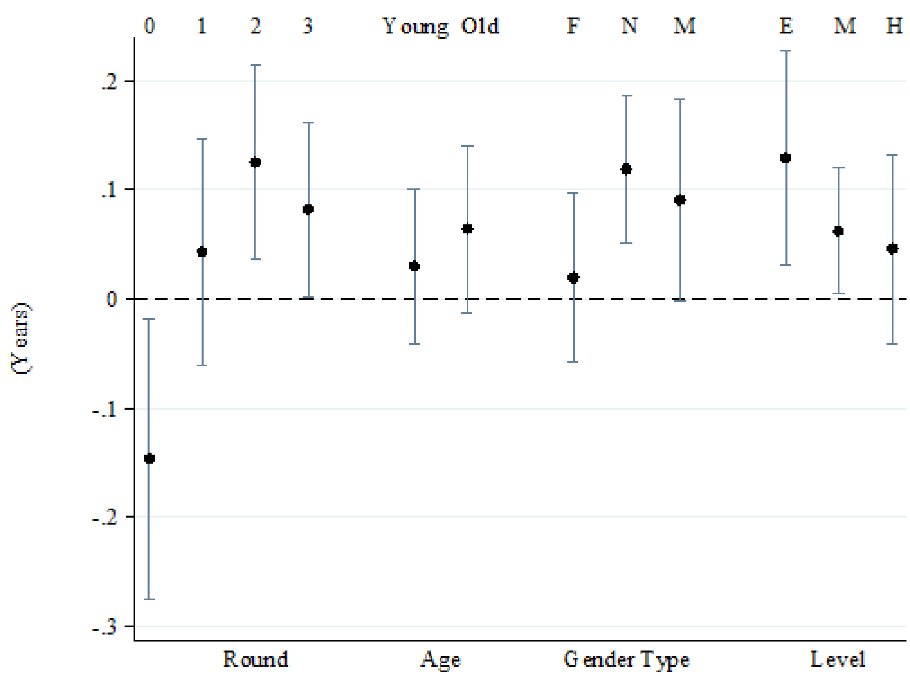
Note: The number of job recommendations in Round 0 to Round 3 is 20. 1-1 represents the first 10 jobs in Round 1, 1-2 is for the second 10 jobs in Round 1.

Figure 5: Gender Differences in Job Characteristics by Groups

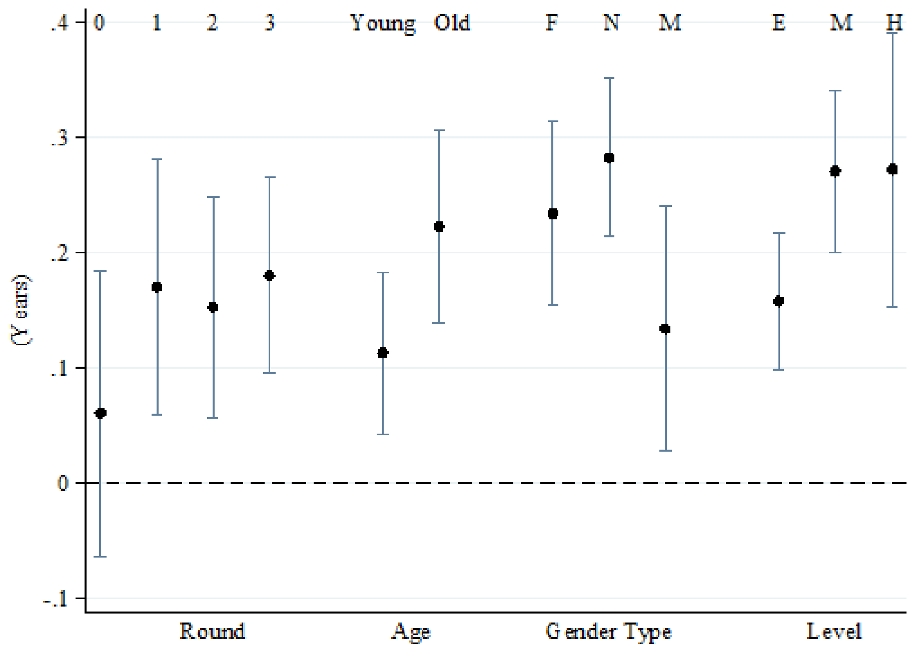
(a) Gender Differences in Posted Wage by Groups



(b) Gender Differences in Requested Education by Groups



(c) Gender Differences in Requested Experience by Groups

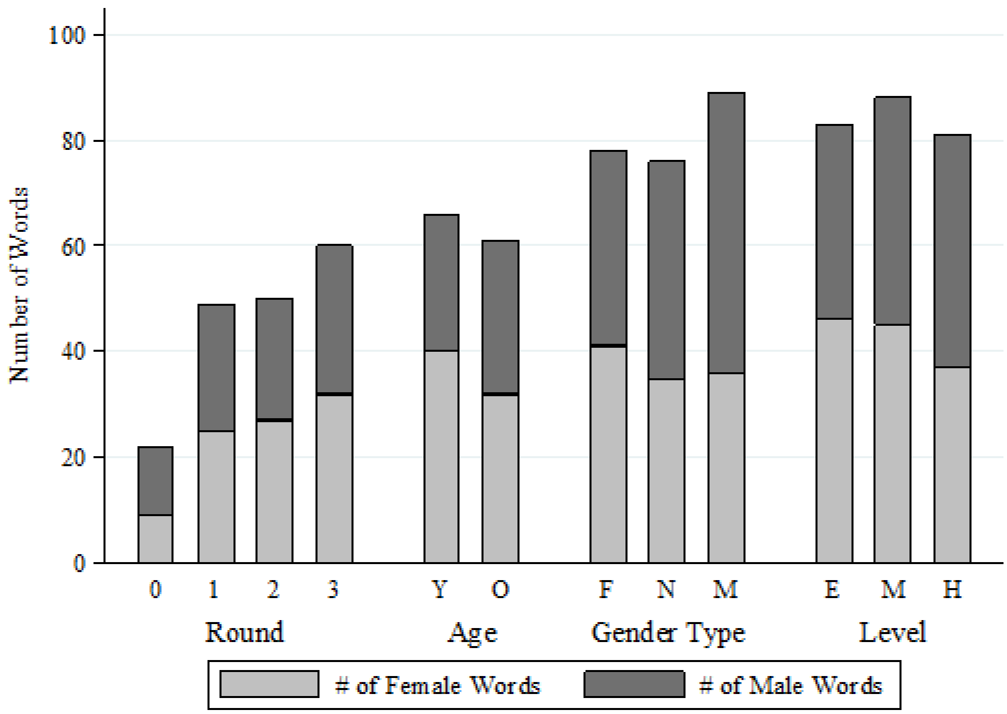


Note: In the job gender type, F denotes female-dominated jobs, N denotes gender-neutral jobs, and M denotes male-dominated jobs. In the job hierarchy level, E denotes entry-level jobs, M denotes middle-level jobs, and H denotes high-level jobs.

Figure 6: Word Cloud from Job Ads

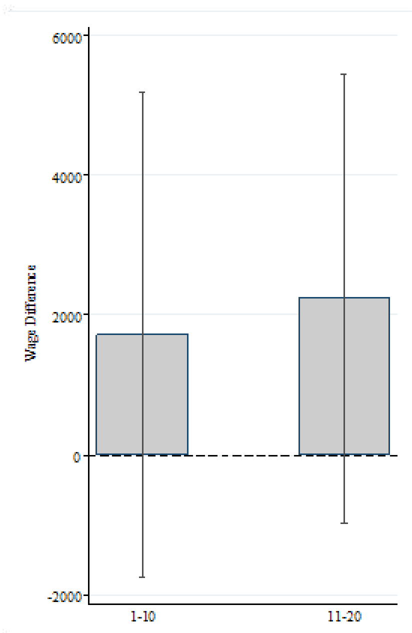
Note: The word cloud is based on the extracted words in the job descriptions from 77,802 recommended job advertisements, and the size corresponds the word frequency. The Chinese version is shown in Appendix D Figure D1.

Figure 7: The Number of Male and Female Words by Groups

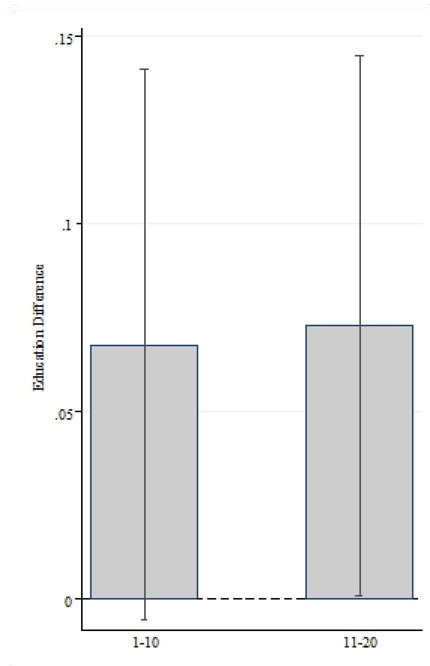


Note: The (fe)male words are defined as words that are correlated with only-to-(fe)male job recommendations (as in Section 4.2) in each group.

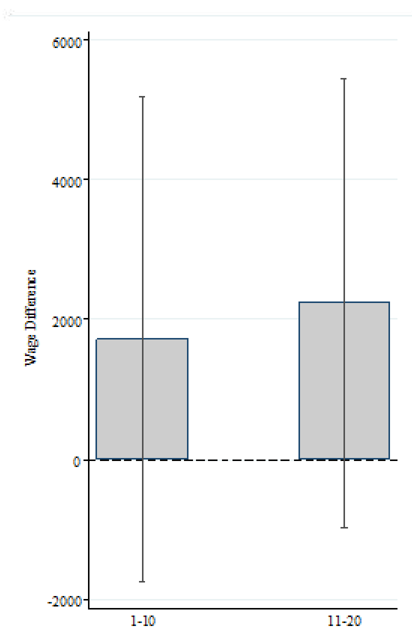
Figure 8: Comparison of Gender-Specific Jobs Before and After Applications



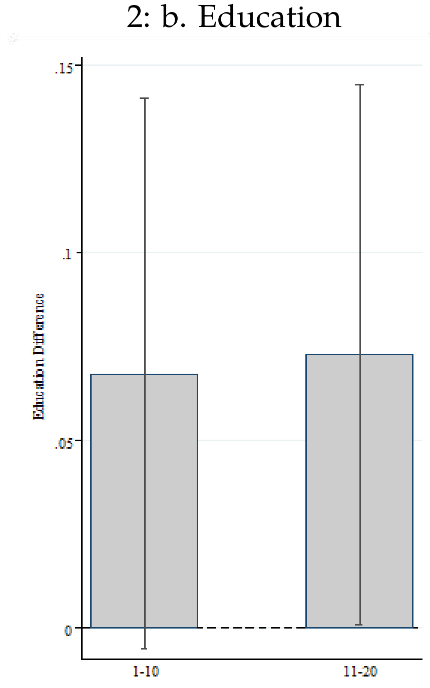
1: a. Wage



2: b. Education



3: c. Experience



4: d. Number of Words

Note: Figure 8(a) to (c) show the gender gap in the characteristics of jobs that are recommended before (1-10 jobs in each round) versus after applications (11-20 jobs in each round) are made. Figure 8(d) compares the number of male and female words in gender-specific jobs before and after applications.

Table 1: Descriptive Statistics: Applicant Sample

	Mean
Current Wage	142507.1
	(65141.8)
Desired Wage	179732.1
	(81818.9)
Education	15.56
	(0.4960)
Sample Size	2,240

Note:
1. Current wage and desired wage are annual wage in RMB.
2. Education levels in resumes are transformed to the years of education. A college degree is equivalent to 15 years of education, and a bachelor's degree is equivalent to 16 years of education.
3. Standard errors are in parentheses.

Table 2: Descriptive Statistics: Recommended job Sample

	Mean
Posted Wage?	0.9503
	(0.2173)
Wage, if posted	212611
	(821107)
Required Education	14.852
	(2.2369)
Required Experience	2.3064
	(2.1563)
Large Company	0.3562
	(0.4727)
Sample Size	77,802

Note:
1. Wage is the midpoint of the posted range of wages.
2. Education levels in job ads are transformed to the years of education. Middle school takes 9 years of education, tech school and high school are equivalent with 12 years of education, college is 15 years of education, and bachelor's degree is equivalent with 16 years of education, master/MBA is 18 years of education, and doctoral degree is 23 years of education.
3. Large company refers to companies that have more than 1,000 employees. The company size is self-reported by hiring agents.
4. Standard errors are in parentheses.

Table 3: Descriptive Statistics: Job Recommendation Sample

	Mean
Desired Wage Match	0.8789
Education Match	0.8925
Experience Match	0.9160
Location Match	0.9916
Sample Size	177,108

Note:
1. Desired wage match equals 1 if the recommended job's upper bound of posted wage range is higher than the worker's lowest desired wage.
2. Education (experience) match is 1 if the worker's years of education (experience) are above the request from the recommended job.
3. Location match is 1 if the worker's city is consistent with the job's city.

Table 4: Difference Rate in Job Recommendations

	Difference Rate (S.D.)	Group Difference
All	0.1230 (0.0588)	
Age		
Young	0.1255 (0.0572)	0
Old	0.1204 (0.0603)	-0.0052
Gender		
Female	0.1174 (0.0577)	-0.0107***
Neutral	0.1281 (0.0600)	0
Male	0.1125 (0.0552)	-0.0156***
Hierarchy		
Entry	0.1186 (0.0465)	0
Middle	0.1281 (0.0653)	0.0095**
High	0.1274 (0.0594)	0.0088**
City		
Beijing	0.1284 (0.0537)	0
Shanghai	0.1201 (0.0555)	0.0083
Shenzhen	0.1235 (0.0599)	0.0049
Guangzhou	0.1209 (0.0616)	0.0075

Note:
1. Difference rate is computed by the number of gender-specific recommendations over the number of total recommendations received by both male and female applicants in the gender pair.
2. For the comparisons within each group, the rate differences and significance levels are derived from t-test.*** $p<0.01$, ** $p<0.05$, * $p<0.1$.
3. Duplicate job recommendations of different rounds are counted once.

Table 5: Gender Differences in Characteristics of Job Recommendations

	(1) Posted Wage	(2) Education	(3) Experience
Male	2,616.1232**	0.0466	0.1666***
	(1,240.305)	(0.031)	(0.028)
N	20,321	22,245	22,245
R^2	0.620	0.374	0.382

Note:
1. The regression sample is the jobs that are only recommended to one worker in gender pairs, and the outcome variables are the job's posted wage, requested years of education and requested years of working experience. Male is the indicator for only-to-male jobs.
2. Pair fixed effect is controlled in all columns. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Gender Difference in Words in Job Recommendations

	Female Words	Male Words
Skills	listen (-0.0078), speak (-0.0078), write (-0.186), documentation (-0.0183), data (-0.0390), chat tools (-0.0142), cooperation (-0.0522), communication (-0.0183), assist (-0.0186), negotiation (-0.0105), administrative (-0.0186), collect (-0.0266)	decision-making (0.0172), planning (0.0170), engineering (0.0286), independent (0.0234), leadership (0.0306), charge (0.0157), supervise (0.0524), design (0.0156)
Work Timing and Location	eight-hour working (-0.0070), flexible (-0.0597), weekly break (-0.0429), regular hour (-0.0149)	nightwork (0.0020), work overtime (0.0124), long travel (0.0153)
Benefits	marriage leave (-0.0121), maternity leave (-0.0057), parental leave (-0.0057), social security (-0.0098), maternity insurance (-0.0031), medical insurance (-0.0208)	commission (0.0992), stock (0.0190), allowance (0.0415), reward (0.0696), meal (0.0195), shuttle (0.0207), commute friendly (0.0470), injury insurance (0.0190)
Company	training (-0.0239), atmosphere (-0.0156)	public company (0.0488)
Other Qualifications	certificate (-0.0104), new grad (-0.0092), non experience (-0.0057)	science&engineering (0.0133), no crime history (0.0225)
Personality, Age, and Appearance	careful (-0.0368), patient (-0.0097), active (-0.0115), outgoing (-0.0149), generous (-0.0060), punctual (-0.0076), figure (-0.0474), temperament (-0.0296), facial (-0.0042)	self-motivated (0.0065), pressure (0.0515), innovative (0.0146), entrepreneurial (0.0125)

Note: Table 6 displays words that are significantly correlated with male-only or female-only jobs. Female and male words are derived from regressions in equation (2), in which the outcome variable is the dummy for the appearance of the word in the gender-specific jobs. The estimated coefficient of female-only jobs are in parentheses. Words that have higher probabilities of be included in female-only (male-only) jobs than male-only (female-only) jobs at 5% significance are female (male) words.

Table 7: Gender Differences on Words and Gender Stereotypes

	Female Words	Male Words
Skills	listen, speak, write, documentation, data, chat tools, cooperation, communication, assist, negotiation, administrative, collect	decision-making, planning, engineering, independent, leadership, charge, supervise, design
Work Timing and Location	eight-hour working, flexible, weekly break, regular hour	nightwork, work overtime, long travel
Benefits	marriage leave, maternity leave, parental leave, social security, maternity insurance, medical insurance	commission, stock, allowance, reward, meal, shuttle, commute friendly, injury insurance
Company	training, atmosphere	public company
Other Qualifications	certificate, new grad, non experience	science&engineering, no crime history
Personality, Age, and Appearance	careful, patient, active, outgoing, generous, punctual, figure, temperament, facial	self-motivated, pressure, innovative, entrepreneurial

Note: Table 7 shows the relation between gendered words in job ads and gender stereotypes. The color intensity indicates the maleness and femaleness consistency with gender stereotypes from literature and two survey results. Female words are highlighted with red colors, male words highlighted with blue colors, and strong color indicates high consistency.

Table 8: Top 10 Words in Prediction of Gender-Specific Recommended Jobs

OLS	Lasso	Ridge	Random Forest
supervise	engineering	independent	commission
flexible	regular hour	public	vacation
long travel	commission	engineering	commute
regular hour	flexible	cooperation	logic
commission	independent	Eight-hour	public
medical insurance	cooperation	certificate	shuttle
careful	public	regular hour	allowance
pressure	overtime	careful	collect
engineering	flexible	night work	oversea
cooperation	supervise	supervise	documentation

Note:
1. Table 8 presents the top words in predicting whether a job is only recommended to male applicants. The outcome variable is binary and equals 1 for male-only jobs, and independent variables are 172 dummy variables for the existence of words in job ads.
2. Column 1 lists words from the OLS regression, which are significant at 5% level and sorted in descending order of the magnitude of coefficients.
3. Column 2 and 3 present words that are selected by the Lasso and Ridge regression. The penalty parameter for Lasso regression is 0.25 and is 0.25 in Ridge regression. Those are determined by using 20-fold cross-validation for the highest R squared. Words are sorted in descending order of the magnitude of estimation effects.
4. In column 4, random forest is applied to find words that have high impacts on the classification of male-only and female-only jobs based on 100 bootstraps and Gini impurity. Words are sorted in descending order of the importance factor.

Table 9: Effects of Views from Hiring Agents on Difference Rate

	(1)	(2)	(3)	(4)
ViewT	0.0310***	0.0310***	0.0308***	0.0313***
	(0.003)	(0.003)	(0.003)	(0.008)
Age		Yes	Yes	Yes
Job Gender Type			Yes	Yes
Job Board				Yes
N	1,031	1,031	1,031	1,031
R^2	0.336	0.336	0.345	0.366

Note:
1. The dependent variable is the number of gender-specific jobs in 100 recommendations.
2. In column 1, the regressor is the total number of views on the female's profile and the male's profile in each gender pair. Column 2 controls for young or older pairs. Column 3 further controls the worker's job gender type. Column 4 adds the job board fixed effect.
3. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

References

Abraham, L. and Stein, A. (2022), 'Words matter: Experimental evidence from job applications', *Unpublished manuscript* .

Al-Otaibi, S. T. and Ykhlef, M. (2012), 'A survey of job recommender systems', *International Journal of Physical Sciences* **7**(29), 5127–5142.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A. and Rieke, A. (2019), 'Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes', *Proceedings of the ACM on Human-Computer Interaction* **3**(CSCW), 1–30.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016), 'Machine bias', *ProPublica, May* **23**(2016), 139–159.

Arnold, D., Dobbie, W. and Hull, P. (2021), Measuring racial discrimination in algorithms, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 49–54.

Avivi, H., Kline, P., Rose, E. and Walters, C. (2021), Adaptive correspondence experiments, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 43–48.

Barocas, S. and Selbst, A. D. (2016), 'Big data's disparate impact', *Calif. L. Rev.* **104**, 671.

Bartlett, R., Morse, A., Stanton, R. and Wallace, N. (2021), 'Consumer-lending discrimination in the fintech era', *Journal of Financial Economics* .

Bem, S. L. (1981), 'Gender schema theory: A cognitive account of sex typing.', *Psychological review* **88**(4), 354.

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.

Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E. H. et al. (2019), Fairness in recommendation ranking through pairwise comparisons, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', pp. 2212–2220.

Black, S. E. and Spitz-Oener, A. (2010), 'Explaining women's success: technological change and the skill content of women's work', *The Review of Economics and Statistics* **92**(1), 187–194.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V. and Kalai, A. T. (2016), 'Man is to computer programmer as woman is to homemaker? debiasing word embeddings', *Advances in neural information processing systems* **29**.

Booth, A. and Leigh, A. (2010), 'Do employers discriminate by gender? a field experiment in female-dominated occupations', *Economics Letters* **107**(2), 236–238.

Bozdag, E. (2013), 'Bias in algorithmic filtering and personalization', *Ethics and information technology* **15**(3), 209–227.

Buolamwini, J. and Gebru, T. (2018), Gender shades: Intersectional accuracy disparities in commercial gender classification, *in* 'Conference on fairness, accountability and transparency', PMLR, pp. 77–91.

Burke, R., Sonboli, N. and Ordonez-Gauger, A. (2018), Balanced neighborhoods for multi-sided fairness in recommendation, *in* 'Conference on Fairness, Accountability and Transparency', PMLR, pp. 202–214.

Bustelo, M., Díaz Escobar, A. M., Lafortune, J., Piras, C., Salas, L. M. and Tessada, J. A. (2020), What is the price of freedom? estimating women's willingness to pay for job schedule flexibility, Technical report, IDB Working Paper Series.

Caliskan, A., Bryson, J. J. and Narayanan, A. (2017), 'Semantics derived automatically from language corpora contain human-like biases', *Science* **356**(6334), 183–186.

Card, D., Cardoso, A. R. and Kline, P. (2016), '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.

Cediey, E. and Foroni, F. (2008), *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office*, ILO.

Chaturvedi, S., Mahajan, K., Siddique, Z. et al. (2021), Words matter: Gender, jobs and applicant behavior, Technical report, IZA DP No. 14497.

Chen, L., Ma, R., Hannák, A. and Wilson, C. (2018), Investigating the impact of gender on rank in resume search engines, *in* 'Proceedings of the 2018 chi conference on human factors in computing systems', pp. 1–14.

Christl, M. and Köppl-Turyna, M. (2020), 'Gender wage gap and the role of skills and tasks: evidence from the austrian piaac data set', *Applied Economics* **52**(2), 113–134.

Coffman, K., Collis, M. and Kulkarni, L. (2021), Stereotypes and belief updating, Technical report, Harvard Business School, Working Paper 19-068.

Cortés, P., Pan, J., Pilossoph, L. and Zafar, B. (2021), Gender differences in job search and the earnings gap: Evidence from business majors, Technical report, National Bureau of Economic Research.

Cowgill, B. and Tucker, C. E. (2019), 'Economics, fairness and algorithmic bias', *preparation for: Journal of Economic Perspectives* .

Craswell, N., Zoeter, O., Taylor, M. and Ramsey, B. (2008), An experimental comparison of click position-bias models, *in* 'Proceedings of the 2008 international conference on web search and data mining', pp. 87–94.

de Sá, C. C., Gonçalves, M. A., Sousa, D. X. and Salles, T. (2016), Generalized broof-l2r: a general framework for learning to rank based on boosting and random forests, *in* 'Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval', pp. 95–104.

Eriksson, S. and Lagerström, J. (2012), 'The labor market consequences of gender differences in job search', *Journal of Labor Research* **33**(3), 303–327.

Exley, C. L., Niederle, M. and Vesterlund, L. (2020), 'Knowing when to ask: The cost of leaning in', *Journal of Political Economy* **128**(3), 816–854.

Fitzpatrick, M. A., Mulac, A. and Dindia, K. (1995), 'Gender-preferential language use in spouse and stranger interaction', *Journal of Language and Social Psychology* **14**(1-2), 18–39.

Flory, J. A., Leibbrandt, A. and List, J. A. (2015), '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.

Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., Xu, S., Geng, S., Shah, C., Zhang, Y. et al. (2020), Fairness-aware explainable recommendation over knowledge graphs, *in* 'Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval', pp. 69–78.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2020), 'Predictably unequal? the effects of machine learning on credit markets', *The Effects of Machine Learning on Credit Markets (October 1, 2020)* .

Gao, R., Ge, Y. and Shah, C. (2021), 'Fair: Fairness-aware information retrieval evaluation', *arXiv preprint arXiv:2106.08527* .

Gastil, J. (1990), 'Generic pronouns and sexist language: The oxymoronic character of masculine generics', *Sex roles* **23**(11), 629–643.

Gaucher, D., Friesen, J. and Kay, A. C. (2011), 'Evidence that gendered wording in job advertisements exists and sustains gender inequality.', *Journal of personality and social psychology* **101**(1), 109.

Ge, Y., Liu, S., Gao, R., Xian, Y., Li, Y., Zhao, X., Pei, C., Sun, F., Ge, J., Ou, W. et al. (2021), Towards long-term fairness in recommendation, *in* 'Proceedings of the 14th ACM International Conference on Web Search and Data Mining', pp. 445–453.

Gee, L. K. (2019), 'The more you know: Information effects on job application rates in a large field experiment', *Management Science* **65**(5), 2077–2094.

Hannak, A., Soeller, G., Lazer, D., Mislove, A. and Wilson, C. (2014), Measuring price discrimination and steering on e-commerce web sites, *in* 'Proceedings of the 2014 conference on internet measurement conference', pp. 305–318.

Hannák, A., Wagner, C., Garcia, D., Mislove, A., Strohmaier, M. and Wilson, C. (2017), Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr, *in* 'Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing', pp. 1914–1933.

He, H., Neumark, D. and Weng, Q. (2021), 'Do workers value flexible jobs? a field experiment', *Journal of Labor Economics* **39**(3), 709–738.

Helleseter, M. D., Kuhn, P. and Shen, K. (2020), 'The age twist in employers' gender requests evidence from four job boards', *Journal of Human Resources* **55**(2), 428–469.

Hellström, T., Dignum, V. and Bensch, S. (2020), 'Bias in machine learning–what is it good for?', *arXiv preprint arXiv:2004.00686* .

Hoffman, C. and Hurst, N. (1990), 'Gender stereotypes: Perception or rationalization?', *Journal of personality and social psychology* **58**(2), 197.

Hoffman, M., Kahn, L. B. and Li, D. (2018), 'Discretion in hiring', *The Quarterly Journal of Economics* **133**(2), 765–800.

Hong, W., Zheng, S., Wang, H. and Shi, J. (2013), 'A job recommender system based on user clustering.', *J. Comput.* **8**(8), 1960–1967.

Ibañez, M. and Riener, G. (2018), 'Sorting through affirmative action: Three field experiments in colombia', *Journal of Labor Economics* **36**(2), 437–478.

Jannach, D., Resnick, P., Tuzhilin, A. and Zanker, M. (2016), 'Recommender systems—beyond matrix completion', *Communications of the ACM* **59**(11), 94–102.

Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010), *Recommender systems: an introduction*, Cambridge University Press.

Jiang, R., Chiappa, S., Lattimore, T., György, A. and Kohli, P. (2019), Degenerate feedback loops in recommender systems, *in* 'Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society', pp. 383–390.

Jiang, S., Hu, Y., Kang, C., Daly Jr, T., Yin, D., Chang, Y. and Zhai, C. (2016), Learning query and document relevance from a web-scale click graph, *in* 'Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval', pp. 185–194.

Kay, M., Matuszek, C. and Munson, S. A. (2015), Unequal representation and gender stereotypes in image search results for occupations, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', pp. 3819–3828.

Kilby, A. E. (2021), Algorithmic fairness in predicting opioid use disorder using machine learning, *in* 'Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency', pp. 272–272.

Kim, P. T. (2016), 'Data-driven discrimination at work', *Wm. & Mary L. Rev.* **58**, 857.

Kim, P. T. (2017), 'Auditing algorithms for discrimination', *U. Pa. L. Rev. Online* **166**, 189.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V. and Jain, A. K. (2012), 'Face recognition performance: Role of demographic information', *IEEE Transactions on Information Forensics and Security* **7**(6), 1789–1801.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Rambachan, A. (2018), Algorithmic fairness, *in* 'Aea papers and proceedings', Vol. 108, pp. 22–27.

Kline, P., Rose, E. K. and Walters, C. R. (2022), 'Systemic discrimination among large us employers', *The Quarterly Journal of Economics* **137**(4), 1963–2036.

Köchling, A. and Wehner, M. C. (2020), 'Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of hr recruitment and hr development', *Business Research* pp. 1–54.

Kuhn, P. J. and Shen, K. (2021), What happens when employers can no longer discriminate in job ads?, Technical report, National Bureau of Economic Research.

Kuhn, P. J., Zhang, S., Liu, T. and Dai, K. (2022), Should i show or should i hide: When do jobseekers reveal their wages?, Technical report, Working Paper.

Kuhn, P. and Shen, K. (2013), 'Gender discrimination in job ads: Evidence from china', *The Quarterly Journal of Economics* **128**(1), 287–336.

Kuhn, P., Shen, K. and Zhang, S. (2020), 'Gender-targeted job ads in the recruitment process: Facts from a chinese job board', *Journal of Development Economics* **147**, 102531.

Kuhn, P. and Villeval, M. C. (2015), 'Are women more attracted to co-operation than men?', *The Economic Journal* **125**(582), 115–140.

Kunaver, M. and Požrl, T. (2017), 'Diversity in recommender systems–a survey', *Knowledge-based systems* **123**, 154–162.

Lambrecht, A. and Tucker, C. (2019), 'Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads', *Management science* **65**(7), 2966–2981.

Lang, S., Laumer, S., Maier, C. and Eckhardt, A. (2011), Drivers, challenges and consequences of e-recruiting: a literature review, *in* 'Proceedings of the 49th SIGMIS annual conference on Computer personnel research', pp. 26–35.

Le Barbanchon, T., Rathelot, R. and Roulet, A. (2021), 'Gender differences in job search: Trading off commute against wage', *The Quarterly Journal of Economics* **136**(1), 381–426.

Lee, D. H. and Brusilovsky, P. (2007), Fighting information overflow with personalized comprehensive information access: A proactive job recommender, *in* 'Third International Conference on Autonomic and Autonomous Systems (ICAS'07)', IEEE, pp. 21–21.

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.

Li, D., Raymond, L. R. and Bergman, P. (2020), Hiring as exploration, Technical report, National Bureau of Economic Research.

Lindqvist, A., Renström, E. A. and Sendén, M. G. (2019), 'Reducing a male bias in language? establishing the efficiency of three different gender-fair language strategies', *Sex Roles* **81**(1), 109–117.

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.

Mullainathan, S. and Obermeyer, Z. (2021), On the inequity of predicting a while hoping for b, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 37–42.

Nosek, B. A., Banaji, M. R. and Greenwald, A. G. (2002), 'Harvesting implicit group attitudes and beliefs from a demonstration web site.', *Group Dynamics: Theory, Research, and Practice* **6**(1), 101.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science* **366**(6464), 447–453.

OECD (2016), *The survey of adult skills: Reader's companion*, Organisation for Economic Co-operation and Development.
URL: *https://doi.org/10.1787/9789264204027-en*

Pedreshi, D., Ruggieri, S. and Turini, F. (2008), Discrimination-aware data mining, *in* 'Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 560–568.

Pekkarinen, T. and Vartiainen, J. (2006), 'Gender differences in promotion on a job ladder: Evidence from finnish metalworkers', *ILR Review* **59**(2), 285–301.

Pető, R. and Reizer, B. (2021), 'Gender differences in the skill content of jobs', *Journal of Population Economics* **34**(3), 825–864.

Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2020), Mitigating bias in algorithmic hiring: Evaluating claims and practices, *in* 'Proceedings of the 2020 conference on fairness, accountability, and transparency', pp. 469–481.

Rambachan, A., Kleinberg, J., Ludwig, J. and Mullainathan, S. (2020), An economic perspective on algorithmic fairness, *in* 'AEA Papers and Proceedings', Vol. 110, pp. 91–95.

Rambachan, A. and Roth, J. (2019), 'Bias in, bias out? evaluating the folk wisdom', *arXiv preprint arXiv:1909.08518* .

Richardson, M., Dominowska, E. and Ragno, R. (2007), Predicting clicks: estimating the click-through rate for new ads, *in* 'Proceedings of the 16th international conference on World Wide Web', pp. 521–530.

Roberts, D. C. and Utych, S. M. (2020), 'Linking gender, language, and partisanship: Developing a database of masculine and feminine words', *Political Research Quarterly* **73**(1), 40–50.

Roussille, N. (2020), 'The central role of the ask gap in gender pay inequality', *URL: https://ninaroussille. github. io/files/Roussille_ askgap. pdf* **34**, 35.

Rudman, L. A. and Kilianski, S. E. (2000), 'Implicit and explicit attitudes toward female authority', *Personality and social psychology bulletin* **26**(11), 1315–1328.

Sandvig, C., Hamilton, K., Karahalios, K. and Langbort, C. (2014), 'An algorithm audit', *Data and Discrimination: Collected Essays. Washington, DC: New America Foundation* pp. 6–10.

Savage, D. D. and Bales, R. (2016), 'Video games in job interviews: using algorithms to minimize discrimination and unconscious bias', *ABAJ Lab. & Emp. L.* **32**, 211.

Siting, Z., Wenxing, H., Ning, Z. and Fan, Y. (2012), Job recommender systems: a survey, *in* '2012 7th International Conference on Computer Science & Education (ICCSE)', IEEE, pp. 920–924.

Szpektor, I., Maarek, Y. and Pelleg, D. (2013), When relevance is not enough: Promoting diversity and freshness in personalized question recommendation, *in* 'Proceedings of the 22nd international conference on World Wide Web', pp. 1249–1260.

Wu, L., Liu, Q., Chen, E., Yuan, N. J., Guo, G. and Xie, X. (2016), 'Relevance meets coverage: A unified framework to generate diversified recommendations', *ACM Transactions on Intelligent Systems and Technology (TIST)* **7**(3), 1–30.

Yu, H., Liu, C. and Zhang, F. (2011), 'Reciprocal recommendation algorithm for the field of recruitment', *Journal of Information & Computational Science* **8**(16), 4061–4068.

Appendices

Appendix A

Resume Audit Study Experimental Design

A1: Job Type Selection

In each job board, 35 types of jobs were selected based on three criteria: the number of active job openings, the job's gender type (female-dominated jobs, gender-balanced jobs, and male-dominated jobs), and hierarchy level (entry, middle, and high). For each job type, we scraped 50 job ads to determine the education level and academic major that are required by most employers. In addition, 50 resumes in the job type were employed to derive the current wages (adjusted to be age-appropriate).

Table A1 lists the selected job type (industry-occupation cell) in each job board, the corresponding hierarchy level (low, middle, high), the required education level, and the major. Current wage (,) represents current wages for (young, older) workers in 10k RMB, respectively.

Table A1.1 Selected Job Types in Job Board 1

Gender	Industry	Occupation	Hierarchy Level	Education Level	Major	Current Wages
M	Computer Software	Software Engineer	Low	Bachelor	Computer Science	(14, 17)
	Computer Software	Senior Software Engineer	High	Bachelor	Computer Science	(17, 23)
	Internet/ E-Business	Operations Specialist	Low	College	Computer Science	(7, 9)
	Internet/ E-Business	Operations Manager/Supervisor	High	Bachelor	Computer Science	(11, 14)
	Machine Manufacturing	General Worker /Operator		College	Machinery	(8, 13)
	Automobiles/Motorcycles	General Worker /Operator		College	Machinery	(9, 13)
	Transportation/Shipping	Courier		College	Econ&Management	(5, 6)
	Internet/ E-Business	Courier		College	Econ&Management	(6, 7)
	Wholesale/Retail	Warehouse Keeper		College	Econ&Management	(4, 5)
N	Internet/ E-Business	Data Analyst		Bachelor	Statistics	(11, 14)
	Computer Software	Data Analyst		Bachelor	Statistics	(11, 14)
	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(13, 17)
	Internet/ E-Business	Product Manager/Supervisor		Bachelor	Econ&Management	(13, 17)
	Internet/ E-Business	Sales Representative	Low	College	Marketing	(5, 7)
	Education/Training	Sales Representative	Low	College	Marketing	(5, 7)
	Real Estate Services	Sales Representative	Low	College	Marketing	(6, 8)
	Internet/ E-Business	Sales Manager	Middle	College	Marketing	(12, 17)
	Computer Software	Sales Manager	Middle	College	Marketing	(12, 17)
	Wholesale/Retail	Sales Director	High	Bachelor	Marketing	(16, 21)
	Internet/ E-Business	Sales Director	High	Bachelor	Marketing	(16, 21)
	Internet/ E-Business	Front Desk	Low	College	Econ&Management	(6, 8)

F	Professional Services	Front Desk	Low	College	Econ&Management	(6, 8)
	Professional Services	Executive Assistant	Low	College	Econ&Management	(7, 9)
	Computer Software	Executive Assistant	Low	College	Econ&Management	(7, 9)
	Internet/ E-Business	Executive Manager	High	College	Econ&Management	(11, 13)
	Wholesale/Retail	Store Clerk	Low	College	Marketing	(5, 7)
	Wholesale/Retail	Store Manager	High	College	Marketing	(9, 11)
	Internet/ E-Business	Customer Service	Low	College	Marketing	(5, 6)
	Finance/Securities	Customer Service	Low	College	Marketing	(5, 6)
	Internet/ E-Business	Customer Service Manager	High	College	Marketing	(8, 12)
	Trade/Import-Export	Accountant		Bachelor	Accounting	(8, 12)
	Wholesale/Retail	Accountant		Bachelor	Accounting	(8, 12)
	Internet/ E-Business	HR Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Professional Services	HR Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Internet/ E-Business	Human Resources Manager	High	College	Econ&Management	(9, 12)

Table A1.2 Selected Job Types in Job Board 2

Gender	Industry	Occupation	Skill Level	Education Level	Major	Current Wages
M	Computer Software	Software Engineer		Bachelor	Computer Science	(15, 23)
	Internet	Mobile Development Engineer		Bachelor	Computer Science	(16, 23)
	Internet	Algorithm Engineer		Bachelor	Computer Science	(17, 24)
	Internet	Operations Specialist	Low	College	Computer Science	(7, 9)
	Internet	Operations Manager/Supervisor	High	Bachelor	Computer Science	(11, 14)
	Real Estate Development	Real Estate Project Management		Bachelor	Architecture	(14, 22)
N	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(14, 20)
	Internet	Product Manager/Supervisor		Bachelor	Econ&Management	(14, 20)
	Computer Software	Project Manager/Supervisor		Bachelor	Econ&Management	(13, 19)
	Internet	Project Manager/Supervisor		Bachelor	Econ&Management	(13, 19)
	Internet	Data Analyst		Bachelor	Statistics	(12, 18)
	Big Data	Data Analyst		Bachelor	Statistics	(12, 18)
	Securities/Investment	Data Analyst		Bachelor	Statistics	(12, 18)
	Advertising/Public Relations	Public Relations Specialist/Assistant		College	Marketing	(11, 14)
	Advertising/Public Relations	Public Relations Manager/Supervisor		Bachelor	Marketing	(15, 20)
	E-Business	Sales Representative	Low	College	Marketing	(7, 12)
	Internet	Sales Representative	Low	College	Marketing	(7, 12)
	Education/Training	Sales Representative	Low	College	Marketing	(7, 12)
	Real Estate Services	Sales Representative	Low	College	Marketing	(8, 13)
	Wholesale/Retail	Sales Manager	Middle	College	Marketing	(12, 17)
	Real Estate Services	Sales Manager	Middle	College	Marketing	(12, 17)

	Internet	Sales Director	High	Bachelor	Marketing	(14, 19)
	Wholesale/Retail	Sales Director	High	Bachelor	Marketing	(14, 19)
F	E-Business	Web Customer Service	Low	College	Marketing	(6, 8)
	Banking	Telephone Customer Service	Low	College	Marketing	(6, 8)
	E-Business	Customer Service Manager	High	College	Marketing	(12, 15)
	Banking	Customer Service Manager	High	College	Marketing	(12, 15)
	E-Business	Accountant		Bachelor	Accounting	(9, 14)
	Internet	HR Specialist/Assistant	Low	College	Econ&Management	(6, 9)
	Professional Services	HR Specialist/Assistant	Low	College	Econ&Management	(6, 9)
	Internet	Human Resources Manager/Supervisor	High	Bachelor	Econ&Management	(11, 14)
	Computer Software	Human Resources Manager/Supervisor	High	Bachelor	Econ&Management	(11, 14)
	Internet	Executive Assistant/Secretary	Low	College	Econ&Management	(7, 9)
	Internet	Administration Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Internet	Administration Manager/Supervisor	High	College	Econ&Management	(9, 14)

Table A1.3 Selected Job Types in Job Board 3

Gender	Industry	Occupation	Skill Level	Education Level	Major	Current Wages
M	Internet/E-Business	WEB Front-end Developer		Bachelor	Computer Science	(17, 24)
	Machine Manufacturing	Mechanical Engineer		Bachelor	Machinery	(16, 21)
	Computer Software	Software Engineer	Low	Bachelor	Computer Science	(18, 25)
	Computer Software	Senior Software Engineer	High	Bachelor	Computer Science	(22, 27)
	Internet/E-Business	Operations Specialist	Low	College	Computer Science	(10, 13)
	Internet/E-Business	Operations Manager/Supervisor	High	Bachelor	Computer Science	(14, 20)
	Real Estate Development	Architect		Bachelor	Architecture	(15, 22)
N	Pharmaceuticals/Biotechnology	Sales Representative	Low	College	Marketing	(10, 15)
	Securities/Investment Funds	Sales Representative	Low	College	Marketing	(11, 15)
	Pharmaceuticals/Biotechnology	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(14, 18)
	Internet/E-Business	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 18)
	Securities/Investment Funds	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 18)
	Pharmaceuticals/Biotechnology	Sales Director	High	Bachelor	Marketing	(17, 24)
	Internet/E-Business	Sales Director	High	Bachelor	Marketing	(16, 25)
	Commodity	Sales Director	High	Bachelor	Marketing	(16, 24)
	Internet/E-Business	Product Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Internet/E-Business	Project Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Computer Software	Project Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Commodity	Marketing Manager/Supervisor		Bachelor	Marketing	(14, 22)
	Wholesale/Retail	Marketing Manager/Supervisor		Bachelor	Marketing	(14, 22)

	Real Estate Development	Legal manager/Supervisor		Bachelor	Law	(15, 25)
	Internet/E-Business	Legal manager/Supervisor		Bachelor	Law	(15, 24)
F	Internet/E-Business	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Real Estate Development	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Internet/E-Business	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Real Estate Development	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Internet/E-Business	Human Resources Director	High	Bachelor	Econ&Management	(16, 26)
	Real Estate Development	Human Resources Director	High	Bachelor	Econ&Management	(16, 26)
	Internet/E-Business	Accountant	Low	Bachelor	Accounting	(12, 18)
	Securities/Investment Funds	Financial Manager	High	Bachelor	Finance	(15, 20)
	Internet/E-Business	Administration Specialist/Assistant	Low	College	Econ&Management	(9, 13)
	Real Estate Development	Executive Assistant/Secretary	Low	College	Econ&Management	(10, 14)
	Internet/E-Business	Administration Manager/Supervisor	Low	Bachelor	Econ&Management	(15, 20)
	Internet/E-Business	Administration Vice President	High	Bachelor	Econ&Management	(51, 88)

Table A1.4 Selected Job Types in Job Board 4

Gender	Occupation	Skill Level	Education Level	Major	Current Wages
M	WEB Front-end Developer		Bachelor	Computer Science	(19, 25)
	Operation and Maintenance Engineer	Low	Bachelor	Computer Science	(18, 24)
	Operation and Maintenance Director	High	Bachelor	Computer Science	(19, 26)
	Pattern Recognition		Bachelor	Computer Science	(19, 25)
	Machine Learning		Bachelor	Computer Science	(19, 25)
	Operations Assistant	Low	College	Computer Science	(7, 9)
	Operations Specialist	Middle	College	Computer Science	(10, 12)
	Operations Manager/Supervisor	High	Bachelor	Computer Science	(14, 19)
	Test Engineer	Low	Bachelor	Computer Science	(15, 22)
	Test Manager	High	Bachelor	Computer Science	(19, 25)
	Data Architect		Bachelor	Computer Science	(17, 25)
N	Sales Representative	Low	College	Marketing	(8, 12)
	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 17)
	Sales Director	High	Bachelor	Marketing	(18, 25)
	Product Assistant	Low	College	Econ&Management	(9, 10)
	Product Manager	High	Bachelor	Econ&Management	(15, 23)
	Project Assistant	Low	College	Econ&Management	(9, 10)
	Project Manager	High	Bachelor	Econ&Management	(15, 23)
	Data Analyst		Bachelor	Statistics	(13, 19)
	Design Assistant	Low	College	Arts	(8, 10)
	Designer	Middle	College	Arts	(13, 19)

	Design Manager	High	Bachelor	Arts	(15, 23)
	Strategy Consultant		Bachelor	Econ&Management	(13, 19)
F	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 10)
	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Human Resources Director	High	Bachelor	Econ&Management	(17, 26)
	Accountant	Low	Bachelor	Accounting	(13, 17)
	Training Specialist		College	Econ&Management	(10, 12)
	Customer Service	Low	College	Marketing	(7, 8)
	Customer Service Manager	High	College	Marketing	(13, 17)
	Media Specialist	Low	College	Marketing	(7, 8)
	Media Manager	High	Bachelor	Marketing	(10, 15)
	Administration Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Administration Manager/Supervisor	Middle	Bachelor	Econ&Management	(13, 18)
	Administration Director	High	Bachelor	Econ&Management	(16, 25)

Note: The industry in job board 4 is set as “all industries”.

A2. Fictitious Resume

The resumes only contain the basic information required by each job board to register as a valid job seeker. The first section of a fictitious resume is personal information, including worker's name, birth date, years of working experience, current wage, city, employment status, phone number, and email address. The second part is about worker's education: the highest education level, time period, university name and major. The third part describes worker's working experience of the most recent job including the time period, company name, occupation, industry, job title, and job description. The last part is worker's intention for future jobs, including desired wage, desired location, desired industry, and occupation. Two workers in each gender pair have identical backgrounds, and four workers in each group (young male, young female, older male, older female) are placed in each job type.

A2.1 Personal Information

Name: We picked up the most popular first and last names to make up the names of fictitious applicants. Based on the statistics from 2015 Chinese Census 1% Population Sample, we chose the top 20 last names, top 15 male first names, and top 15 female first names as the applicants' name pool (listed in Appendix A2.1). For each applicant, the last name and first name corresponding to the applicant's gender will be randomly drawn from the name pool. Although gender is explicitly stated in the resume and we do not need applicant's name to denote gender, we still adopted first names that are consistent with a worker's gender to make the fictitious profile as common and real as possible.

Names of Fictitious Applicants

Last name: 李(Li), 王(Wang), 张(Zhang), 刘(Liu), 陈(Chen), 杨(Yang), 赵(Zhao), 黄(Hunag), 周(zhou), 吴 (Wu), 徐(Xu), 孙(Sun), 胡(Hu), 朱(Zhu), 高(Gao), 林(Lin), 何(He), 郭(Guo),马(Ma), 罗(Luo).

Male First Name: 伟(Wei), 强(Qiang), 磊(Lei), 军(Jun), 洋(Yang), 勇(Yong), 杰(Jie), 涛(Tao), 超(Chao), 平(Ping), 刚(Gang), 浩(Hao), 鹏(Peng), 宇(Yu), 明(Ming).

Female First Name: 芳(Fang), 娜(Na), 敏(Min), 静(Jing), 丽(Li), 艳(Yan), 娟(Juan), 霞(Xia), 婷(Ting), 雪(Xue), 丹(Dan), 英(Ying), 洁(Jie), 玲(Ling), 燕(Yan).

Birth Date: Employers infer worker's age from the birth date. Instead of varying workers' age directly, we used their graduation year to classify the age level, and "older" workers refer to ones who graduated earlier and have more working experience. Applicants have two potential age levels: Young workers graduated in 2017, and old workers graduated in 2007. After a worker's graduation year is fixed, his age is jointly determined by the graduation year and his education level. The advantage of this design is that workers' years of working experience are equalized within each age level. More specifically, young workers are 25 (with a college degree, born in 1995) or 26 (with a bachelor's degree, born in 1994) with three years of working experience, 35 or 36 years old are for the senior workers with more than 5 years of working experience. Workers in the gender pair have the same randomly drawn birth month and day.

Years of Working Experience: To simplify the profiles, we assumed workers started to work just after they graduated from the university/college of their highest degree. As discussed above, years of working experience is the difference between the current year (2020) and the graduation year. For instance, if a worker graduated in 2017, then he has 2020 – 2017, three years of working experience.

Current Wage: Fictitious workers' wages are drafted based on wages of active workers in job boards by matching their current job position as well as working experience. we used the hiring agent account in each platform and searched for workers that were currently in the job positions and specified the working experience as "1 to 3 years" and "5 to 10 years" in March 2020. For each experience level in every job position, we recorded the first 50 workers' current wages shown in the search result and took the average as the fictitious worker's wage.

City: All of the four job boards are nationally recognized and cover most of the regions in China, and over half of job postings are from first-tier cities. To achieve enough amount of job recommendations, fictitious workers are currently living in the first-tier cities, including Beijing, Shanghai, Shenzhen, and Guangzhou.

Employment Status: All of the workers are currently employed.

Phone number and email: Each applicant has a unique and active email address and mobile phone number.

A2.2 Education

Workers' education level is designed to match jobs' education requirements. For each job type, we checked 50 job advertisements in February 2020 and listed the most common education request. 85% of job ads required workers had a bachelor's or junior college degree. Bachelor's degree often takes 4 years to achieve, while junior college takes 3 years. The end time of school is the graduation year, and the start time of school depends on worker's education degree, which is three years (college degree) or four years (bachelor's degree) earlier than the graduation year. For instance, a young worker, graduated with a bachelor's degree in June 2017, is 26 years old (born in 1994) and started his university program in August 2013.

Two workers in the same gender pair have the same educational background, and the school's name is randomly drawn from the Chinese High Education Institution List, released by the Ministry of Education in 2019, and the school locations match the worker' current location.[1] Majors will also match job positions: Computer Science/Software is for IT jobs, Mathematics/Statistics is for data position, and economics/management/marketing majors are for other jobs.

A2.3 Recent Job History

As we assume all the workers are currently employed, their recent jobs are their current jobs. For young workers, their current jobs started in August in the year when they graduated with the highest degree (2017); for old workers, their current jobs started five years ago, in March 2015, implying that they have 5 years tenure in their recent positions.

We made up company names to minimize the disturbance to both job seekers and employers on job platforms. The company name consists of three parts: (1) company's location. It will be the same with worker's current city. (2) company's name. We used an online business name generator to collect 100 company names listed below. The company name will be randomly assigned to each gender pair. (3) company's industry. It will be consistent with the job's industry. An example of the company name is, Beijing Dongya Internet Technology Company.

[1] Schools are randomly drawn from the surrounding provinces that workers currently locate in, and we excluded the provinces that have ethnic minority groups, such as Xinjiang, Yunnan, Qinghai, Tibet and Guangxi.

Worker's current occupation and industry will be the same as the job's occupation and industry. Job title and job description are filled in by words, and we set them as the job's occupation.

Names of Company

东艾, 森利, 先卓, 利晟, 同通, 富长盛, 芯达, 精典, 尼佳, 益复捷, 生德, 晶长, 森益, 金伙伴, 德光, 茂全, 鲜派, 信顺康, 龙丝, 新耀协, 佳丽, 昇晖, 佳洲, 森道尔, 皇祥千, 润飞昌, 福中荣, 基玉, 如和, 茂乾, 翔鹏, 南湘, 圣泰, 吉春, 本寿, 亚义金, 耀浩, 邦洁, 宝复, 洪进贵, 永泰满, 显郦, 华行, 韵仪, 格派, 晶佩, 迪和, 领速, 贝耀, 信华诚, 世力, 舜杰, 久福, 曼新, 仁大兴, 金祥元, 泰伟飞, 亚和金, 吉振, 和伟中, 盛金缘, 立韦, 宏久, 吉至, 曼展, 天联, 金涛, 网诚, 系广, 圣金龙, 易露发, 嘉利华, 聚顿, 公同宏, 威邦, 力涛, 恒蓝, 铭航, 中美公, 永逸, 同捷, 发和, 易龙, 汉金, 干亚, 翔洋, 新都, 茂进永, 达通, 娇罗, 浩中和, 东升, 龙姿, 隆新弘, 仟顺, 越福, 川实, 中协吉, 霸辉, 洪谦, 裕飞

A2.4. Job Intention

A worker searches for full-time jobs, in which the desired wage is 120% of his current wage (or the wage range), and the desired city, industry, and occupation will be the same as the current ones.

Table A2.1 summarizes the information included in worker's resume.

Table A2.1: Resume Information Generation

	Method	Note
Personal Information		
Name	Randomly assigned to each worker	Appendix A2.1
Birth Date	Young worker graduated in 2017, and older worker graduated in 2007. Birth year is decided by graduation year and education level.	Young, bachelor's =1994, Young, college=1995. Older, bachelor's =1984, Older, college=1985.
Years of Working Experience	2020 - graduation year	3 or 13 years
Current Wage	Average wage of the collected workers in the platforms.	Adjust with job type and experience.
City	Beijing, Shanghai, Shenzhen, Guangzhou	
Employment Status	Currently employed.	
Phone Number & Email	Uniquely assigned for each worker.	
Education		
Highest degree	Assigned on group level, based on job type's education requirement.	Bachelor's degree or junior college.
Time Period	Graduation year – years to achieve the highest degree.	4 years to achieve bachelor's degree, 3 years to achieve college degree.
School Name	Randomly drawn for each gender pair.	Chinese High Education Institution List (2019)
Major	Same on group level.	Depends on job type.
Recent Job		
Time Period	Young worker: after graduation (2017) until now,	

	Older worker: 2015 until now.	
Company Name	Location +name + industry, name will be randomly assigned to each worker.	Appendix A2.2
Occupation	Same with job type	
Industry	Same with job type	
Job Title	Same with occupation	
Job Description	Same with occupation	
Intention		
Desired Wage	Current wage*1.2	
Desired City	Same with city	
Desired Industry	Same with job type	
Desired Occupation	Same with job type	

Table A3: Descriptive Statistics of Job Recommendations by Gender

	Male	Female	Difference
Posted Wage?	0.9497 (0.0008)	0.9491 (0.0008)	0.0006
Wage, if posted	196815 (3337.3)	196449 (3337.4)	366.02
Required Education	14.6337 (0.0086)	14.6264 (0.0087)	0.0072
Required Experience	2.2742 (0.0083)	2.2589 (0.0082)	0.0153
Desired Wage Match	0.8889 (0.0012)	0.8871 (0.0012)	0.0018
Education Match	0.9185 (0.0011)	0.9194 (0.0011)	-0.0008
Experience Match	0.9179 (0.0011)	0.9195 (0.0011)	-0.0016
Location Match	0.9915 (0.0003)	0.9916 (0.0003)	0.0000
Number of Views	23.1005 (0.6281)	23.0060 (0.6226)	0.0945
N	88,544	88,544	

Note: The sample is job recommendations received by male and female applicants. All the differences in column 3 are insignificant at 10% level.

Appendix B

Ranking Difference of Job Recommendations

The ranking difference measure takes the rank of recommended jobs into account in its measure of gender differences. According to this measure, two job recommendation lists are the same only if the two jobs in the same rank are identical. Then the ranking difference rate is defined as:

$$Ranking\ Difference\ Rate = \frac{\sum_{i=1}^{n} ith\ job\ ad\ is\ difference\ in\ gender\ pair}{Length\ of\ recommendation\ list\ (n)}$$

Figure B1: Ranking Difference Measure in Job Recommendations

	Male	Female	
1st	Job 1	Job 1	Same
2nd	Job 2	Job 2	Same
3rd	Job 3	Job 4	
:			
ith	Job i	Job i+1	
:			
nth	Job n	Job 3	

In the above example, only the first two jobs in recommendation lists are the same, then ranking difference rate is (n-2)/n.

Table B1: Ranking Difference Rate in Job Recommendations

	Share
All	0.6838
Round	
0	0.6087
1	0.6580
2	0.7148
3	0.7541
Age	
Young	0.6815
Old	0.6861
Gender	
Female	0.6680
Neutral	0.7038
Male	0.6722
Hierarchy	
Entry	0.6585
Middle	0.6985
High	0.6914
City	
Beijing	0.6868
Shanghai	0.6782
Shenzhen	0.6882
Guangzhou	0.6721

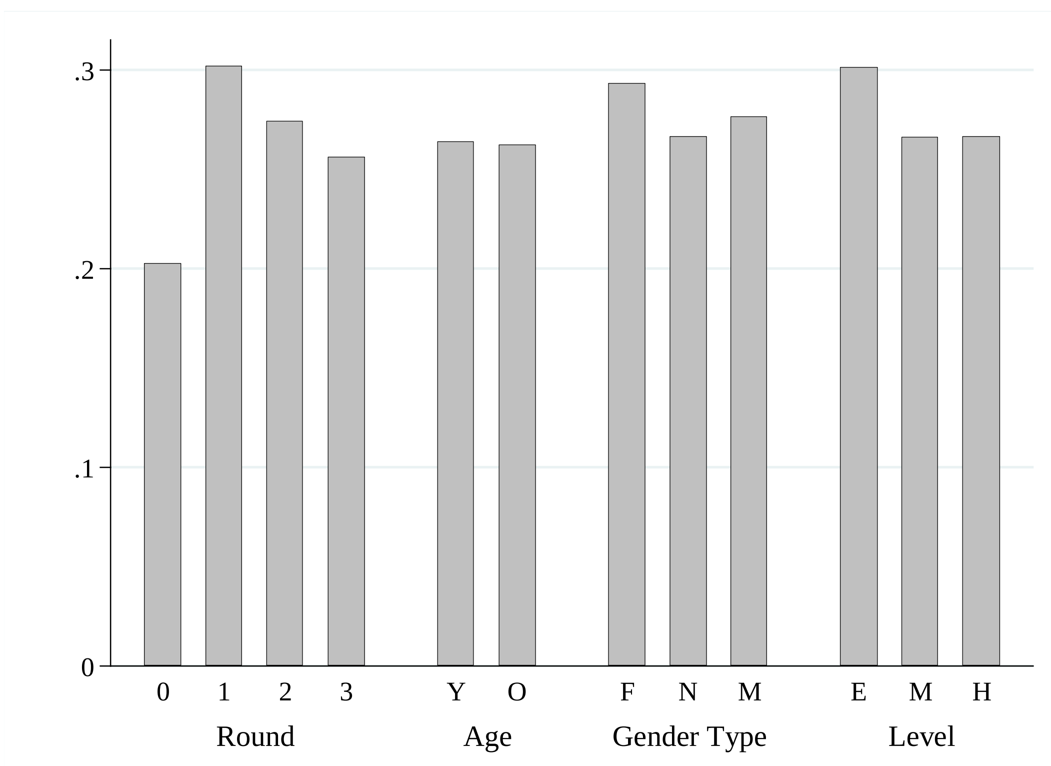
Table B1 summarizes the average ranking difference rates by experimental rounds, the worker's age, and the job's gender type, hierarchy level and city. The overall ranking difference rate is 68.4%, indicating that in a list of 100 recommended jobs, only around 32 jobs are displayed identically to male and female applicants. Similar to the results in Figure 4, the ranking difference rate increases substantially after applicants send out job applications, from 60.8% in Round 0 to 75.4% in Round 3. The ranking difference rate has a quite consistent pattern with the difference rate in Section 4.1 across the subsamples by age, job's gender type, hierarchy level and city.

Appendix C

Word Dissimilarity in Job Recommendations

To achieve an overall evaluation of gender difference in the words contained in only-to-male versus only-to-female jobs, we compute the vector dissimilarity between the average jobs recommended to men and women. Based on the extracted words in the job descriptions, job *i* can be described by a vector S_i with 172 elements, in which the *jth* word s_{ij}, ($j = 1, \ldots, 172$) equals 1 if job *i* contains word *j*. The dissimilarity between the average male-only job, $\bar{S}_M$ and the average female-only job $\bar{S}_F$ is computed as Euclidean distance between two vectors, and is plotted in Figure C1. It suggests that on the aggregate level, the wording in jobs that are recommended to men and women has a dissimilarity about 0.26. This number is slightly higher among young workers applying to entry-level, female-dominated jobs compared to other jobs.

Figure C1: Measure of Words' Dissimilarity in Job Recommendations



Appendix D

Words in Job Recommendations

Figure D1: Word Cloud in Chinese

Table D1: Word List in Job Ads

Skills	Literacy skills: listen, speak, read, write, language, documentation Numeracy skills: data, accounting, analysis ICT skills: programing, microsoft office, chat tools Problem solving skills: learning, comprehension, thinking, logic, decision-making, planning, problem-solving, engineering, independent, insight Influencing skills: leadership, team management, charge, supervise Co-operative skills: cooperation, communication, teamwork, assist, coordination, organize, negotiation, public relation, marketing, sale, client, compliance Self-organizing skills: administrative, design, collect, reception, driving, execution, test, task management
Work Timing and Location	Schedule: work shift, night work, morning work, evening work, big and small week*, eight-hour, flexible, attendance, overtime, no overtime Business travel: regular travel, short travel, long travel Work break: weekly break, monthly break, noon break, regular working hour
Benefits	Payment: base pay, commission, stock, allowance, promotion, reward Break: vacation, marriage leave, parental leave, maternity leave, sick leave, funeral leave, holiday Facilities: office supplements, vehicle, meal, housing, shuttle, subway, commute friendly, snacks Insurance: fiveone*, medical insurance, commercial insurance, social security, housing funds, maternity insurance, unemployment insurance, endowment insurance, injury insurance, disease insurance Other benefits: training, staffing, activities, mentor
Company	Environment: atmosphere, employee care, career, dream, culture, screening Type: direct recruiting, public company, top500, startup, flat management, financing, big company* Title: senior, medium, core

Other Qualifications	Education: non education, certificate, new grad, tongzhao*, tier-one school, fulltime school, top school, nonmajor, major, science&engineering Experience: non experience, experienced, oversea Other: no crime history, law abiding, solitary
Personality, Age, and Appearance	Personality: effective, rigorous, careful, patient, energetic, active, outgoing, optimistic, virtuous, trustworthy, honest, practical, self-motivated, hardworking, passion, tenacious, sharp mind, generous, curious, courageous, innovative, punctual, entrepreneurial, devotion, enthusiasm, kind, responsible, work under pressure, responsive Age: non gender, non age, age below35, age below40 Appearance: figure, temperament, healthy, facial, clothing, shape, voice

Note:

1. Table D1 shows the extracted words from job ads in four job boards, and the restrictions are described in Section 4.2.
2. Every listed word includes its variations, such as leadership vs leading, and confidence vs confident.
3. fiveone represents "five social insurance and one housing fund" (五险一金), including endowment insurance, medical insurance, unemployment insurance, employment injury insurance, maternity insurance, and housing fund. Big and small week describes the working schedule in which workers have one-day rest in one week and two-day rest in the next week. Big company indicates companies that have more than 1000 employees. Tongzhao means university or college admission is through Gaokao in high school.

D2: Survey from Amazon MTurk

To determine the gendered perceptions of words, I recruited participants from Amazon's Mechanical Turk (MTurk) in September 2021 to choose whether the existence of a certain word in the job ad indicates gender stereotypes and implicit gender preferences of employers.

The survey question is: "Suppose you are the hiring agent of a company, and plan to post a job advertisement that contains the word X in the job description. This indicates that you prefer to hire (1) no gender request for worker; (2) male worker; (3) female worker".

In total, 86 valid surveys were collected from people between the ages of 25 to 55, and 56% of them were men. The gender score of a word is computed as:

Score = -1*number of participants choose (3) + 1* number of participants choose (2)

, in which -1 indicates the extreme female word and 1 implies the extreme male word. The average gender score of words in the survey is 0.0905 and the standard deviation is 0.1111. Male words are defined as words whose scores are above one standard deviation from the mean, 0.2016, and female words' scores are below one standard deviation from the mean, -0.0206.[2]

Table D2: Gendered Words from Amazon MTurk Survey

Female Words	Male Words
read, write, documentation, learning, assist, compliance, administrative, design, reception, holiday, marriage leave, maternity insurance, maternity leave, parental leave, sick leave, enthusiasm, kind, patient, careful, figure, temperament, shape, voice	data, analysis, logic, engineering, independent, leadership, supervise, negotiation, driving, work shift, night work, evening work, big and small week, overtime, long travel, commission, promotion, stock, vehicle, mentor, startup, science&engineering, experienced, no crime history, effective, practical, responsible, pressure

[2] tierone university and tongzhao are excluded from the surveyed words because they are only identified in the Chinese high-level education system.

D3: Survey from Chinese Workers

The Chinese version of the survey on people's perceptions about gendered words in job ads was conducted in Wenjuanxing (问卷星) in September 2021. The surveyed question is the same as the one from AMturk, but in Chinese: 假设您是公司 HR，发布的招聘广告中包含以下词汇，代表您倾向于招聘 （1 ）性别不限；（2）男员工；（3）女员工。

79 valid respondents participated in the survey, 81% of them were between 25 to 55 years old and 73% of them were men. The average gender score of words in the survey is 0.0962 and the standard deviation is 0.0721. Male words are defined as words whose scores are above one standard deviation from the mean, 0.1683, and female words' scores are below one standard deviation from the mean, 0.0241.

Table D3: Gendered Words from Chinese Survey

Female Words	Male Words
listen, speak, read, communication, assist, compliance, administrative, design, collect, reception, eight-hour, flexible, office supplements, marriage leave, parental leave, sick leave, maternity insurance, atmosphere, employee care, patient, active, outgoing, passion, kind, figure, temperament, healthy, facial, shape, voice	data, problem-solving, engineering, independent, leadership, charge, teamwork, negotiation, driving, nightwork, overtime, long travel, commission, stock, promotion, meal, commute, unemployment insurance, injury insurance, disease insurance, training, staffing, culture, screening, core, oversea, no crime history, optimistic, practical, self-motivated, tenacious, courageous, punctual, entrepreneurial, responsible, pressure, responsive, age below40

Table D4: Gender Differences on Words and Gender Stereotypes in Round 0

	Female Words	Male Words
Skills	listen, data, cooperation, assist,	decision-making, supervise,
Work Timing and Location	flexible, weekly break	work overtime, big week
Benefits	marriage leave	reward, meal, shuttle, commute friendly, housing, allowance
Company		public company
Personality, Age, and Appearance	figure, temperament	pressure

Appendix E

Hiring Agents' Views and Job Recommendations

Table E1: Effects of HR Views on Gender Differences in Job Characteristics

	(1) Wage	(2) Education	(3) Experience	(4) # Words
ViewT	29.2746	-0.0013	0.0028*	0.0058
	(54.550)	(0.001)	(0.002)	(0.004)
N	1,031	1,031	1,031	1,031
R^2	0.069	0.041	0.047	0.048

Note:

1. The dependent variables are the gender gap (male - female) in the gender-specific jobs' posted wage, requested education, requested working experience and the total number of male and female words, from column 1 to column 4.
2. All regressions control for fixed effects of worker's age, job's gender type and job boards.
3. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table E2: Effects of HR Views on Gender Differences in Job Characteristics By Rounds

	(1) Diff Rate	(2) Wage	(3) Education	(4) Experience	(5) # Words
ViewT	0.0403**	127.6739	-0.0045	0.0043	-0.0022
	(0.018)	(166.293)	(0.005)	(0.005)	(0.009)
N	3,125	3,125	3,125	3,125	3,125
R^2	0.178	0.069	0.041	0.047	0.048

Note:

1. Table E2 decomposes the estimations of Table 9 and Table E1 by experimental rounds. The independent variable is the total number of views received by two workers in the gender pair in a certain round, and the outcome variables are the difference rate, the gender gap (male - female) in the gender-specific jobs' posted wage, requested education, requested working experience and the total umber of male and female words in that round.
2. All regressions control for fixed effects of experimental rounds, worker's age, job's gender type and job boards.
3. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table E3: Effects of HR Views on the Change of Gender Differences in Job Characteristics Between Round 0 and Round 1

	(1)	(2)	(3)	(4)	(5)
	ΔDiff Rate	ΔWage	ΔEducation	ΔExperience	Δ# Words
ViewT	0.0742***	-639.7461	-0.0065	0.0186	0.0029
	(0.028)	(723.665)	(0.021)	(0.024)	(0.003)
N	994	986	994	994	994
R^2	0.212	0.175	0.010	0.012	0.020

Note:

1. Table E3 estimates the effect of HR views on the changes of gender gap between Round 0 and Round 1. The regressor is the total views on male and female profiles in each gender pair in Round 1. The outcomes are the changes of difference rate between Round 0 and Round 1, and the changes of gender gaps on recommended jobs' posted wage, education, experience and the total number of male and female words from Round 0 to Round 1.
2. All regressions control for fixed effects of worker's age, job's gender type and job boards.
3. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

E4: Hiring Agent Views' and Applicants' Job Recommendations

Section 5 explains the effect of views from hiring agents on gender difference in jobs that the pairwise men and women receive. Here we discuss how the views of hiring agents affect the jobs recommended to individual applicants by regressing the changes of jobs' characteristics between two rounds on the views of hiring agents:

$$Z_i = \alpha_0 + \alpha_1 * ViewD_i + AX + e_i$$

The outcome variable Z_i is the change in the characteristics (posted wage, requested education, requested working experience and the total number of male and female words) of jobs recommended to applicant *i* (between rounds *t-1* and *t*, *t = 1, 2, 3*), and $ViewD_i$ represents the extra views from hiring agents that received by applicant *i* from rounds *t-1* to *t*. X is the fixed effects for worker's age, job's gender type, job boards and experimental rounds. We run the regressions separately for male and female applicants, and results are shown in Table E4a and E4b, respectively. Consistent with results on the gender-pair level, the views of hiring agents on applicants' profile increases the share of gender-different jobs in their recommendation lists, but they change little on the quality of jobs that applicants receive.

Table E4a: Effects of HR Views on the Change of Job Characteristics (Male Applicant Sample)

	(1) ΔDiff Rate	(2) ΔWage	(3) ΔEducation	(4) ΔExperience	(5) Δ#Words
ViewD	0.0213**	369.7237	-0.0007	0.0047	0.0017
	(0.010)	(2,700.288)	(0.004)	(0.004)	(0.003)
N	3,295	3,206	3,295	3,295	3,295
R^2	0.029	0.012	0.046	0.017	0.038

Table E4b: Effects of HR Views on the Change of Job Characteristics (Female Applicant Sample)

	(1) ΔDiff Rate	(2) ΔWage	(3) ΔEducation	(4) ΔExperience	(5) Δ#Words
ViewD	0.0209**	144.1467	0.0050	0.0071*	0.0022
	(0.011)	(2,737.239)	(0.004)	(0.004)	(0.003)
N	3,295	3,206	3,295	3,295	3,295
R^2	0.029	0.012	0.042	0.016	0.037