NBER WORKING PAPER SERIES

MEASURING BIAS IN JOB RECOMMENDER SYSTEMS: AUDITING THE ALGORITHMS

Shuo Zhang
Peter J. Kuhn

Working Paper 32889
http://www.nber.org/papers/w32889

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2024

We thank Sarah Bana, Kelly Bedard, Clément de Chaisemartin, Joanna Lahey, Heather Royer, Benjamin Villena Roldan, Kim Weeden, Catherine Weinberger, seminar participants at the University of Oregon and UC San Diego, and participants at the Chinese Labor Economists Society Conference (Shandong University) and LM2C2 Workshop on Gender and Labor Market Mismatches, Santiago, Chile for many helpful comments. This study was approved under UCSB IRB No.17-20-0451 and is in the AEA RCT Registry under AEARCTR-0006101. The authors declare that they have no relevant financial relationships or other conflicts of interest affecting this manuscript. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring Bias in Job Recommender Systems: Auditing the Algorithms
Shuo Zhang and Peter J. Kuhn
NBER Working Paper No. 32889
August 2024
JEL No. C99, J16, J71, M50, O33

ABSTRACT

We audit the job recommender algorithms used by four Chinese job boards by creating fictitious applicant profiles that differ only in their gender. Jobs recommended uniquely to the male and female profiles in a pair differ modestly in their observed characteristics, with female jobs advertising lower wages, requesting less experience, and coming from smaller firms. Much larger differences are observed in these ads' language, however, with women's jobs containing 0.58 standard deviations more stereotypically female content than men's. Using our experimental design, we can conclude that these gender gaps are generated primarily by content-based matching algorithms that use the worker's declared gender as a direct input. Action-based processes like item-based collaborative filtering and recruiters' reactions to workers' resumes contribute little to these gaps.

Shuo Zhang
Department of Economics
Northeastern University
43 Leon Street, 312A Lake Hall
Boston, MA 02115
s.zhang@northeastern.edu

Peter J. Kuhn
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and IZA
and also NBER
pjkuhn@econ.ucsb.edu

A randomized controlled trials registry entry is available at https://doi.org/10.1257/rct.6101-1.2

1 Introduction

Paradoxically, the explosive growth of information about products, workers, jobs and services created by the Internet can make it harder for people to find what they want in a vast ocean of choices. Personalized recommender systems, first proposed in the 1990s, have provided powerful solutions to this information overload problem by showing every user a personalized list of items (Lee and Brusilovsky, 2007; Jannach et al., 2010). Recommender systems have been successfully applied in a variety of contexts, including e-commerce sites, streaming platforms and internet job boards.[1]

To the best of our knowledge, essentially all major internet job boards now use algorithms to recommend jobs to workers, based on criteria like the match between the worker's characteristics and the job's requirements, and the previous behavior of workers and recruiters on the board. While job recommendation algorithms have the potential to help workers and firms find better matches faster, they have also sparked concerns about fairness: Even when there is no discriminatory intent from designers, the recommended jobs may reinforce gender and other stereotypes. For instance, even algorithms that do not rely directly on the worker's gender can learn to associate genders with certain types of jobs and skills, leading to gender segregation in job recommendations. In addition, algorithms that incorporate the past behavior of hiring agents could learn to accommodate those agents' discriminatory preferences.

This paper uses an *algorithm audit* to measure whether, to what extent, and why job board algorithms treat identical male and female job seekers differently. Our algorithm audit adapts the well known resume audit method to study the behavior of job recommendation algorithms rather than the behavior of people (i.e. recruiters). In more detail, we created identical paired worker *profiles* on the four largest Chinese job boards, which

[1] Recent evidence shows that 35% of purchases on Amazon and 80% of stream time on Netflix are driven by their respective recommendation systems (Chong, 2020; MacKenzie et al., 2013). According to 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 (Board, 2022; Andrew, 2020).

differ only in their declared gender (male or female) and a gender-matched name (henceforth 'gender').[2] Next, we sequentially published pairs of profiles and observed which jobs were recommended to them before they took any actions (such as viewing a job or applying to one). Then, to track how algorithms update their recommendations based on recruiters' reactions to our resumes and our resumes' application histories, our fictitious workers applied to the top ten jobs in their recommendation lists six times over the following six weeks. During this process, we collected data on the new recommendations received and the number of times each of our profiles was viewed by recruiters on the board.

We find that identical male and female applicants do not always receive the same job recommendations: out of every 100 job recommendations received by a typical applicant, 12.4 jobs were not seen by their identical, opposite-gender pair. This *difference rate* grows from 7.34% before our profiles have applied to any jobs to 17.60% after six sets of applications that follow the board's recommendations. Furthermore, the jobs recommended to men and women have different characteristics: Combining all rounds of the experiment, we find that only-to-male jobs –which are seen by men but not women– posted wages that were 3,118 RMB (or 1.54%) higher than only-to-women jobs. While the gender gap in requested education is close to zero, jobs recommended only to men require 0.17 (or 7.19%) more years of working experience, and were 2.76 percentage points or 7.92% more likely to come from firms with 1000 or more employees. Thus the gender gap in recommended wages is associated with a tendency to direct men to larger firms, and to higher-ranked jobs within those firms (as proxied by experience requirements).

Next, we measured the extent to which the words in recommended jobs reinforce commonly-held gender stereotypes using a four-stage approach. First, we extracted all the (stemmed and non-'stop') words that were used at least 100 times in the population of recommendations our profiles received. Next, we identified which of those 172 words

[2] Our causal estimates therefore refer to the combined effects of the worker's declared gender and having a female name. While algorithms could be deriving some of their information about gender from names, this seems unlikely when precise data on gender is easily available and its use by matching algorithms is not explicitly prohibited.

were statistically over-represented in jobs directed at men (22 words) and in jobs directed at women (36 words). Many of these over-represented words appear to promote gender stereotypes; for example, *leadership*, *entrepreneurial*, and *work under pressure* appear more often in male-only jobs. Jobs recommended to women are more likely to include the words *patient* and *careful*, and are more likely to list appearance-related criteria, including *facial features* and *figure*.

Third, to determine whether these over-represented words reflect commonly-held gender stereotypes, we turned to four sources external to our study –published lists, our own surveys of U.S. and Chinese workers, and ChatGPT– and assigned male and female stereotype scores to each of our over-represented words based only on these sources. Finally, to quantify the amount of stereotypical content in jobs recommended to men versus women, we constructed *ad-level* stereotype scores by summing all the words' scores in each job ad and estimated gender differences in stereotypical content. We found that only-to-women jobs have 0.58 standard deviations more stereotypically female content than only-to-men jobs; only-to-men jobs have 0.13 standard deviations more stereotypically male content than only-to-women jobs. Thus, while both male and female stereotypes are reflected in job recommendations, the amount of stereotypically-female content is an especially strong predictor of which gender will see a particular ad.[3]

In the second half of our paper we try to isolate which pieces of information and which commonly-used components of recommender algorithms are used on our job boards, and which ones account for the gender differences in job recommendations documented above. Since our profiles differ only in their gender, and since our profiles have no application history in Round 0 of the experiment, our Round 0 results imply that the job recommender systems on these job boards must be using the worker's gender as an input.[4] The same evidence implies that the boards' algorithms must be using *content-based similarity scores* (either between jobs and workers, or between workers and workers) to

[3] As we discuss in Section 3.5, this result is consistent with the "male as default" concept in psychology and linguistics (e.g. Smith and Zarate (1992)).

[4] All the main results reported in this paper combine data from all four of the job boards we study. It is of course possible that these four boards use different types of algorithms, but Appendix D shows that all our main results apply, with less statistical power, to each board separately.

make recommendations. Next, evidence from later rounds strongly suggests that our job boards use *recruiters' reactions* to our fictitious resumes to help craft their recommendations. While this may not seem surprising, it implies that the job recommender algorithms we study have not simply adapted the main processes used in retail sales and information retrieval –which focus exclusively on finding information or items a user wants– to the labor market. Instead, the worker-facing recommendation algorithms on the job boards we study reflect the *two-sided* nature of matching in labor markets, where the users (workers) and the 'items' (in our case the jobs) must want *each other*. Finally, we find essentially no evidence that applying to the jobs previously recommended by the board increases the gender gap in the types of jobs recommended to men versus women.

This paper contributes to three main literatures, the first of which uses *resume audit* methods to study employers' responses to job applications from workers of different races, genders, or other characteristics. For the case of gender, our findings are consistent with a common finding that hiring discrimination can favor either men or women, largely in concordance with industry and occupation-based gender stereotypes (Booth and Leigh, 2010; Cediey and Foroni, 2008; Kline et al., 2022). Our main contribution to this literature is to adapt the resume audit method to study the behavior of algorithms rather than people, showing that algorithms direct workers to jobs that are stereotypical for their gender, even when identical workers are seeking jobs within the same, narrowly defined industry-occupation category.[5] Understanding the effects of algorithms is important because of their increased prevalence as economic actors, and difficult because many influential algorithms are proprietary. Algorithm audits can help fill this gap.

While we study a different outcome (job recommendations) than traditional resume audits, we note that our algorithm audits are considerably easier to conduct on a large scale than traditional ones. In part, this is because algorithm audits can be conducted

[5] Algorithm audits (where investigators supply a series of inputs to 'black box' algorithms) 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 algorithm audits in the economics literature.

with sparse worker profiles that do not require the investigator to fabricate detailed personal working histories and statements of purpose, or to make formatting decisions (font, margins, etc.) that consume investigator resources and introduce noise.[6] Algorithm audits also have a validity advantage because they are harder for employers to detect (Avivi et al., 2021), and an ethical advantage because the inconvenience to human recruiters is negligible: the recommendations we study are made by machines, not people.[7] Finally, a distinct and useful feature of algorithm audits is that –unlike resume audits, which are a 'one-shot' intervention– algorithm audits can take a series of actions after creating a profile (such as viewing and applying to different jobs). As we demonstrate in the paper, this allows us to gather relatively detailed evidence about the precise mechanisms –i.e. data inputs and algorithmic processes– being used by the algorithms.

A second related literature studies gender and other differences in *application behavior*– a phase of the job search process that precedes the candidate selection phase studied by resume audits. Key findings of 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., 2023), avoid competitive work environments (Flory et al., 2015), are attracted to cooperative 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, 2021), are more deterred by ambiguous information about job requirements and the number of competing applicants (Gee, 2019; Coffman et al., 2023; Abraham et al., 2024; Kline et al., 2022), and respond positively to affirmative action statements (Ibañez and Riener, 2018).[8]

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

[7] Our fictitious resumes applied to jobs in rounds 1-3 of the experiment, and occasionally encountered human recruiters at those times. An audit study based on only Round 0 of our experiment, however, would never apply to jobs, essentially eliminating human contact.

[8] In contrast, Castilla and Rho (2023) find negligible effects of the gendering of job postings on worker search behavior. In other applications, Burn et al. (2022) study the effect of job ad content on older workers' application rates, and Flory et al. (2021) study the effects of affirmative action statements on applications from racial minorities.

We contribute to *both* the resume audit and application behavior literatures by focusing on an even earlier stage of the job search process: Which job vacancies does a worker get to see before deciding where to apply?[9] Since workers cannot apply to vacancies they are not aware of, gender differences in application behavior that appear to be driven by differences in preferences (e.g. for greater hours flexibility) could be caused, in part, by automated job recommender systems that inadvertently channel workers toward jobs that match common gender stereotypes. Put another way, the algorithms we study can create the appearance that men and women are choosing to apply to different types of jobs, when in fact they are never informed of some less-gender-typical vacancies that are available in their labor market.

Third, we contribute to a growing literature in both economics and computer science on the fairness of algorithms in the context of worker recruitment.[10] 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 include Hoffman et al. (2018) and Li et al. (2020).[11] 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. Chen et al. (2018) is the only paper we know of that systematically assesses these tools for bias.[12] A final way that algorithms enter the recruitment

[9] An excellent recent group of group of papers has studied the *informal* aspect of how workers learn about the existence of vacant jobs– referrals (Burks et al., 2015; Friebel et al., 2023; Pallais and Sands, 2016; Gee et al., 2017). To our knowledge we are the first researchers to study how workers engaged in the formal process of applying for publicly posted job ads might still have differential access to available job openings.

[10] Other contexts in which algorithmic bias has studied include racial bias against black defendants (Angwin et al., 2016; Cowgill and Tucker, 2020), racial and ethnic discrimination in mortgage lending and credit approval (Bartlett et al., 2021; Fuster et al., 2022), 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).

[11] Hoffman et al. (2018) compare the performance of pre-employment screening algorithms and human HR agents, and Li et al. (2020) build a resume screening algorithm that values candidates' statistical upside potential, then simulates the algorithm's effects using data on past hires. Raghavan et al. (2020) summarizes the advertised capabilities of 18 vendors of algorithmic pre-employment assessments.

[12] 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. Overall, they find that male resumes rank slightly higher

process is in the *job recommender algorithms* we study in this paper. Since worker-initiated search of job ads is much more prevalent than employer-initiated resume search on all the job platforms we are aware of, job recommender systems are probably a much more important determinant of who ultimately works where.[13] To our knowledge, our paper is the only one to experimentally estimate the amount of bias in the job recommender algorithms used on internet job boards.[14] Finally, while both economics and computer science journals have published many papers on algorithmic fairness, to our knowledge ours is the only paper that uses the audit method to attempt to infer the mechanisms (i.e. input data and processes) being used by proprietary, black-box recommender algorithms. We hope that continued collaboration between economists and computer scientists will soon fill this gap.

2 Experimental Design

2.1 Platform Environments

Our experiment was conducted on the four largest job boards in China, which together cover more than 70% of China's online job postings and active workers. The large size of these markets, comprising millions of postings, ensures that our 2,240 fictitious worker profiles have minimal effects on the existing job search and recruiting processes or the job recommender systems. The four job sites have similar interfaces and functions: Job seekers can register and create a profile for free, while employers are charged for posting

than observationally identical female resumes, but (consistent with other studies of employers' gender preferences) this gap is not uniform across job titles.

[13] For example, our data from one of these job boards in 2018 indicates that 82.4% of resumes that were downloaded by hiring agents came from applications, not from employer-initiated resume search.

[14] Two recent papers (Lambrecht and Tucker, 2019; Ali et al., 2019) use field experiments to study which job ads are displayed to men and women on Facebook, where –in contrast to job boards– job ads must compete with consumer advertising in a marketplace. Lambrecht and Tucker (2019) find that women are less likely to see the job ads purchased by the authors than men, but attribute this to the fact that retail ads command a higher price when they are directed at women than men. As a result, Facebook's cost-effectiveness algorithm responded to this price gap by directing retail ads to women and job ads to men. This confounding factor is not present on the job boards we study.

job ads and using recruiter tools, including resume search. Job seekers can view recommended jobs, search for other jobs using keyword searches, and apply to jobs by clicking the jobs' 'apply' buttons. Firms' hiring agents can view recommended workers, search for workers, process applications and contact applicants through each board's recruiter-facing portal. 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.

2.2 Job Type Selection

When a job seeker sets up their profile, the job platforms ask them to select their current and desired industry and occupation from a drop-down list supplied by each board. To represent a broad sample of jobs and workers we targeted 35 industry-occupation cells (a.k.a. *job types*) on each platform using three criteria: sample size, the cell's incumbent gender mix, and the job's skill 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.[15] Second, because male-and female-dominated jobs might prefer applicants whose gender is typical for their industry-occupation cell, we included female-dominated job types (e.g. computer software industry, administrative assistant), gender-balanced job types (e.g. computer software, data analyst), and male-dominated job types (e.g. computer software, software engineer).[16] Finally, because employers' gender preferences could vary with the position's rank (Bertrand et al., 2010; Pekkarinen and Vartiainen, 2006), we include job types at different ranks. For example, sales representative, sales manager, and sales director are low, middle, and high ranked positions in the 'internet / e-business' industry.[17]

[15] 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.

[16] 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.

[17] All these examples of job types are from Job Board 1. The list of job types varies somewhat across the four job boards, depending on the markets they serve. Complete listings of the job types are provided in Appendix A1.

2.3 Resume Setup

We next created job seeker profiles that are qualified for the above jobs by entering data into fields on the 'create an account' menu. The platforms then convert this information into two formats that are visible to recruiters: *summary cards*, which contain a very short list of characteristics, and *resumes* (which are displayed when the recruiter clicks on a card). Because they are platform-generated, all the resumes on the same platform are formatted in exactly the same way and contain only the machine-readable information that was entered into each worker's profile. Since resumes contain no information beyond what is contained in the profiles, we use the terms 'profiles' and 'resumes' synonymously.

On all four of our platforms, gender is a mandatory field for setting up a profile, and only two choices (male or female) are allowed. As noted, our fictitious resumes come in pairs, and the two workers in each pair are identical except for name and gender. Since other research has documented strong interactions between Chinese employers' and and gender preferences (Helleseter et al., 2020), we created two versions of each profile pair: the young workers graduated in 2017 and have three years of experience; the older workers graduated in 2007 and have 13 years. Depending on the job types they apply to, our candidates have either a college or university degree.[18]

To increase our profiles' relevance and realism, the resume information was generated from an information pool of 50 scraped job ads and 50 resumes for each job type on each job boards. Key features of this process (described in more detail in Appendix A2) include the following: The workers' education levels and academic majors satisfy the most common advertised requirements of the job type the worker is seeking. All our applicants are currently employed, and their wages match the wages of existing job seekers by job type, education level, and years of working experience. Since over half of the job postings on our four job boards are from China's four first-tier cities (Beijing, Shanghai, Shenzhen, and Guangzhou), we restrict our applicants' locations to those cities. Each worker's cur-

[18] Chinese college and university degrees take three versus four years to achieve respectively. It follows that our 'young' workers are either 25 or 26 years old, and the older workers are 35 or 36 years old

rent occupation and industry are the same as the job type's occupation and industry, and all workers are seeking jobs in their current city and occupation-industry cell.

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 within job types (except that the older resumes' experience and current wages are adjusted to be age-appropriate). With four resumes per group applying to 35 job types in four cities, this gave us created 560 fictitious profiles on each of the four platforms we studied, or 2,240 profiles in total. These profiles remained unpublished (i.e. invisible to employers and not able to receive job algorithmic job recommendations) until we initiated the experiment for a particular gender pair in its "Round 0".

2.4 Implementation

As illustrated in Figure 1, we harvested data from our fictitious profiles in five Rounds, separated by four two-week Intervals, as follows:

- **Round 0**. The two completed profiles in a gender pair log into their accounts simultaneously and publish their profiles (i.e. make them public). We then immediately collect the first 20 job ads shown to each worker, and the workers log off.
- **Round 1**. Two weeks later, the male and female workers simultaneously log into their accounts again. We then record the number of times their profile was viewed by HR agents since the worker's account was published.[19] We also collect the top 10 jobs in their recommendation lists. The two workers then *apply* to these top 10 recommendations. Next, the workers refresh their web pages and we record the top 10 recommended jobs that appear at this stage as well.
- **Rounds 2 and 3**. At two-week intervals, we repeat the Round 1 procedures.

[19] All of the job boards in our study give workers cumulative counts of the number of times their profile has been viewed each time they log in. The goal is to keep workers engaged, because workers who receive no feedback may become frustrated and switch to other sites (Kim, 2017).

- **Round 4**. Two weeks after Round 3, the profiles log on one final time and we count the cumulative number of profile views at that point.

In all, each of our resumes applied to 30 jobs in an 8-week job search spell, during which we collected the contents of up to 80 jobs that were recommended to them, plus the number of hiring agents' profile views at two-week intervals.[20] Importantly, throughout the experiment, our workers apply for jobs in a naive fashion, applying only to the top 10 jobs that were recommended to them in Rounds 1-3. Because this procedure holds the workers' application strategies constant, it guarantees that any observed gender differences in job recommendations are caused solely by the job boards' recommendation algorithms. Compared to our 'naive' workers, real workers' application strategies could either mitigate or accentuate any gender gaps we measure in rounds 1-3 of our experiment. Workers who are searching for gender-*atypical* jobs may ignore the stereotypical recommendations they receive; if the board's algorithm learns from these choices, the next recommendations these workers receive should be less gender-typed than the ones we collect in our experiment. On the other hand, workers seeking gender-*typical* jobs may elicit an increasingly stereotypical set of job ads that reflect their own past choices. For these workers, our experiment will understate the amount of gender-typing in later rounds of the experiment. That said, the 20 job recommendations we collect in Round 0 –before our profiles have taken any actions– give us clean estimates of the recommendations that *any* newly-created job profile would receive at that time, regardless of its subsequent application behavior.

A final noteworthy feature of our design is that our paired male and female profiles have current jobs in the same industry-occupation cell, and are both seeking new jobs in that same cell. To the extent that the algorithms respect these declarations, any gendered 'steering' that we detect in job recommendations will likely be within a fairly narrow occupation-industry range.[21] Rather than directing, say, women out of highly-male

[20] In rare cases, we received fewer than 80 job recommendations per profile.

[21] While job profiles created by recruiters are categorized using the same industry-occupation cells available in workers' resumes, we unfortunately do not observe these labels in the job ads that are displayed to our fictitious resumes. Thus we cannot precisely measure the extent to which recommended jobs match

job types and into more female ones, any gendered steering is more likely to occur on subtler margins, such as workers' preferences for work hours, competition, and employers' gendered preferences for beauty and personality types.

3 Results

Our resume creation process started in July 2020 and the last collection of hiring agents' profile views was completed in April 2021. During that period, the 2,240 fictitious profiles we created received 177,320 job recommendations from 81,231 individual job advertisements.[22] Descriptive statistics on our samples of fictitious workers and the jobs recommended to them are provided in Tables B1-B3. Applicant characteristics (in Table B1) reflect the levels we have assigned, based on averages taken from real resumes on the job boards. Reflecting the high wage levels of jobs and resumes on these platforms, the average annual wage of our resume sample is 142,507 RMB, which is about twice the 2020 average wage in urban China.[23] The workers' desired wages are 26.12% higher than the their current wages, and the average years of education are 15.56, indicating that about half of the fictitious workers hold a bachelor's degree.[24]

The characteristics of the job ads that were recommended to our fictitious workers are summarized in Table B2. 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. The average posted wage in recommended jobs was 205,928 RMB; mean

workers' current and desired job types.

[22] 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 published. 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 appears to occur randomly; importantly, it is independent of the gender of fictitious applicants.

[23] 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).

[24] While we attempt to set all workers' desired wage at 20% above the worker's current wage, certain platforms force us to choose a desired wage *range*. This accounts for the 26.12% difference in Table B1, which is calculated from the midpoints of these desired wage ranges.

requested years of education and experience were 15.42 and 2.44 respectively.[25] Overall, the jobs recommended to our fictitious workers were well matched with those workers, as shown in Table B3. In around 90% of cases, the recommended jobs' education and experience requirements were at or below the workers' qualifications, and almost all of the recommended jobs' locations matched the worker's current location. 83.86% of recommended jobs posted wages that exceeded the workers' lowest desired wage.

Appendix B also presents descriptive statistics for each of the four job boards in our sample separately, showing that all four boards serve a highly educated group of workers: Mean requested years of education range from 14.81 to 15.77 across the boards. Mean posted salary levels vary more widely, however, ranging from 148 to 251 thousand CNY per year. Unsurprisingly, the highest-salary boards (3 and 4) tend to serve larger employers than the lowest-salary board (board 1). Since the four job boards in our study have different clienteles, it is possible that they use different types of job recommendation systems. To simplify our presentation, however, all our main results combine data from the four boards. In Appendix D, we replicate those findings separately for each individual job board. While the levels of differentiation between male and female profiles vary substantially across the boards –for example, the set difference rates (see below) are 8.07%, 11.56%, 14.31%, and 15.68%– the ways in which the recommended jobs differ between men and women, and the likely processes that create those differences are strikingly similar.

3.1 The Set Difference Rate

Our first set of main results combines the data from all four Rounds of our experiment (and all four job boards) to describe how the jobs recommended to our identical male and female worker profiles differ from each other. The most basic measure of this difference is the share of job ads in a pair of top-N recommendation lists that are unique to a gender,

[25] Throughout the paper, wages for jobs posting wage ranges are the midpoint of the posted range.

i.e. the *set difference rate*:

$$Set\ Difference\ Rate = \frac{M+F}{2N} = \frac{M}{N} = \frac{F}{N} \tag{1}$$

where N is the number of recommendations collected for each gender, M is the number of jobs that only appear in the male worker's list, and F is the number of jobs that only appear in the female worker's list.[26] Notice that –since M must equal F– the set difference rate does not have any 'directionality' in the sense of favoring men versus women. Also, a positive set difference rate does not necessarily indicate algorithmic bias, since the recommended M and F jobs could have essentially the same characteristics. That said, as we demonstrate in Section 4, the set difference rate is a useful tool for learning which algorithmic processes are active on the boards, even when those processes are not gender biased.[27]

Combining the recommendations received in all the rounds of our experiment, the set difference rate between the jobs recommended to male and female applicants is 12.40%. In other words, out of every 100 jobs recommended to male and female applicants, 87.6 jobs are displayed to both applicants and 12.4 jobs were unique to each gender.[28] Table B4 breaks down this overall gender difference rate by applicant age, and by three job characteristics: the predominant gender in the job type (Female, Neutral, or Male), the job's skill level (Entry, Middle, and High) and the city in which the job is located. We find little variation across age levels and cities, but slightly greater gender differences in gender-neutral jobs compared to male- and female-dominated jobs and greater gender differences in mid-

[26] In set theory, the set difference rate is also known as the *symmetric difference* or the *disjunctive union* between two sets.

[27] In addition to gender bias (e.g. the algorithms' use of profiles' gender and name fields to recommend different types of jobs to men and women), positive set differences within gender pairs in our experiment have two other potential causes. The first is quasi-randomness in the algorithms, such as the arrival of new jobs or applicants on the platform or commonly-used decongestion and diversification processes which 'spread out' good matches across applications and vacancies. These quasi-random processes should not cause systematic differences in the *types* of jobs recommended to men and women, however. Second, because our profiles *apply* to the jobs that were recommended to them, any early gender differences could be magnified across the rounds of our experiment. While we cannot precisely quantify all three drivers of difference rates, our Round 0 results imply that quasi-randomness accounts for a difference rate of no more than 7.34 percent (since part of the Round 0 difference rate is associated with gender bias).

[28] With the exception of a small number of worker profiles that did not receive a full set of recommendations, these numbers are based on 20 recommendations from each of rounds 1-4; thus $N = 80$.

dle and high skill level jobs compared to entry level jobs.

Since jobs displayed at the top of workers' recommendation lists are more likely to be seen and clicked into (Craswell et al., 2008; Richardson et al., 2007), measures of recommendation gaps that account for the ranking of jobs may also be of interest. To that end, Table E1 and Figure E1 replicate the preceding analysis using the *ranking* difference rate. While all the 'ranking' differences are greater than the 'set' differences, the cross-sectional patterns and time trends across experimental rounds are very similar. Motivated by these similarities, we confine our analysis to set difference rates (henceforth 'difference rates') in the remainder of the paper.

3.2 Gender Differences in Recommended Job Characteristics

In this section we use the job characteristics that are consistently recorded in almost all our recommended job ads to test for systematic differences between the types of jobs that are recommended to men versus women. Since the job recommendations that are shared by men and women have identical characteristics, we restrict our sample to the job recommendations unique to the male applicant (M), plus the recommendations unique to the paired female applicant (F) across all rounds of the experiment and estimate the following regression:

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

where Y_{pj} is a characteristic of job j that is recommended to the applicants in gender pair p.[29] The variable of interest is M_{pj}, which takes the value of 1 if the recommended job j is only seen by the male in gender pair p. We control for gender pair fixed effects X_p, so β_1 estimates the average gender gap (male-female) in the characteristic between male-only

[29] To explore how the gender-exclusive jobs comprising the Table 1 sample compare to 'common' jobs that were recommended to both the male and female profiles, Appendix E9 replicates our main results on gender bias (Tables 1, 4 and 5) using the full sample of all recommended jobs, using the common jobs as the omitted category and comparing the male- and female-only jobs to them. While some interesting differences are found –for example, both male- and female-only jobs pay less than common jobs, replicating a pattern found for explicit gender requests in Kuhn and Shen (2013) and Helleseter et al. (2020)– these findings do not affect our estimates of gender gaps in job recommendations.

and female-only recommendations within gender pairs.

Our baseline estimates of equation 2 are reported in Table 1, which shows that jobs recommended to men pay 3,118 RMB or 1.54% (3,118/202,453) more than jobs recommended to women; this difference is statistically significant at the 1% level. Requested education levels are statistically indistinguishable from zero, but the jobs recommended to men are 2.67 percentage points more likely to be in firms with 1000 or more employees and request 0.17 years (or 7.19%) more experience. To assess the contribution of these experience and firm size gaps to the gender wage gap, Appendix E3 estimates the cross-sectional return to requested experience and firm size (and their interaction) in a dataset consisting of all the jobs that were recommended to our profiles. Based on those estimates, the experience and firm size differences between the male-only and female-only jobs predict a gender wage gap of 3,957 RMB, or 1.95 percent, which more than fully accounts for Table 1's actual gender wage gap of 1.54 percent.[30]

Figure E2 explores heterogeneity in Table 1's gender recommendation gaps by applicant age, dominant gender of the occupation-industry cell, and position level. Notably, the gender gap in posted wages is statistically significant and almost identical in both our young and old resume pairs.[31] The gender wage gap is considerably higher and highly statistically significant at 8,045 RMB (3.30 percent) in predominantly male job types; this larger wage gap is likely due, in part, to a larger firm size gap in those job types as well. Finally, as one might expect, gender experience gaps are highest for our older applicant profiles and in highly ranked jobs.

Taken together, the fact that about 88 percent of the jobs recommended to men and

[30] Another recommended job characteristic that could vary by gender is a job's 'freshness', for example the elapsed time since it was posted or last refreshed. Since this is not measured consistently across job boards, Appendix E4 conducts a separate analysis for each of the four boards. We find no gender 'freshness' gaps on any of the boards. In addition, two of the boards display measures of the firm's capital and/or financing details in their job postings, but Appendix E5 shows that these do not differ significantly between jobs recommended to men versus women.

[31] Helleseter et al. (2020) find strong age-gender interactions in the *number* of job ads explicitly requesting women versus men in the universe of job ads on four boards; we interpret these requests as being made by human recruiters. Here we are looking at a very different phenomenon: *wage* gaps in job recommendations made by machines to workers seeking jobs in the same, narrowly-defined industry-occupation cell.

women are the same jobs, plus the wage, experience, and firm size gaps we have documented in the gender-differentiated jobs suggest a positive but modest amount of gender bias in the algorithms on these job boards. In part, these modest gaps may reflect the fact that our profiles are seeking jobs in narrowly defined industry-occupation cells: constrained searches like these are more likely to direct men and women to jobs similar characteristics. As we document below, however, substantially greater differences appear when we look at the words contained in the open-text sections of the job ads. Even within these narrow job types, it appears that men and women are steered in quite different directions by the boards' algorithms.

3.3 Learning from Words 1: Parsing Job Ads into Their Most Common (Distinct) Words

We begin our analysis of the unstructured text of recommended job ads with a corpus consisting of the 81,231 job ads that were recommended to our profiles. Using NLP software we broke this corpus into *chunks*, i.e. short, meaningful phrases ranging from one to nine words, then we normalized and combined the chunks that have the same or close meaning (e.g., leadership vs leading) to make the remaining chunks clearly contrast with each other. We then restricted our attention to chunks that appear more than 100 times and manually refined their categorization, resulting in a final selection of 172 chunks, each represented by a single word or phrase, such as "listening", "marriage leave", or "regular working hours".

Henceforth, we refer to these 172 chunks as *words*; they form the basis of all our analysis of the ads' unstructured text. The 172 most common words that emerged from this process are shown in a word cloud in Figure 2, with a larger size representing a higher frequency.[32] 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. To facilitate our

[32] Figure C1 shows these words in the original Chinese.

discussion of the 172 words in Figure 2, we manually assigned them to six categories, described below. A complete list of all the words, by category, is provided in Table C2.

(1) *Standardized (PIACC) Skills.* While a variety of methods have been developed to categorize the 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)[33] PIACC skills are divided into seven subsets, specifically literacy, numeracy, information and communication technology (ICT), problem-solving, influencing, co-operation, and self-organization.

(2) *Benefits.* In Chinese job boards, commonly advertised benefits are often tagged, and their expressions are quite uniform across job types and platforms. We classify these benefits into five types: compensation, leave and vacation, facilities and transportation, insurance, and other benefits.

(3) *Work Timing and Location.* These words refer to work schedules, the need to travel for work, breaks, and overtime.

(4) *Company and Rank.* These words include descriptions of the position's rank (such as senior or middle), company culture (such as "atmosphere" and "employee care"), and company size and type (such as "top 500" or "startup").

(5) *Other Qualifications.* These words include a desire for a specific college major, elite schools, 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.

[33] Christl and Köppl-Turyna (2020) and Pető and Reizer (2021), among others, have used the PIAAC classification to study gender skill differentials.

3.4 Learning from Words 2: Which Words are Over-Represented in Jobs Shown Only to Men and Women?

If the job recommender systems used by our job boards are gender-neutral, the 172 words listed in Table C2 should appear with roughly equal frequency in the jobs recommended exclusively to the male and female job profiles. In this section we identify which individual words are over-represented in jobs recommended to men, and which are over-represented in jobs recommended to women. We also test the null hypothesis that all 172 words are equally likely to be recommended to both genders.

Our main approach uses the sample and regression specification in equation 2, but replaces the outcome variable with a dummy for the appearance of a word in the recommended job. We run this regression for each of Table C2's 172 words; the regressor of interest is whether the job was recommended only to the male profile in the pair. Thus negative (positive) coefficients indicate that the word was over-represented in jobs recommended to women (men). To account for the fact that we are simultaneously testing 172 hypotheses, we use the Romano-Wolf correction (Romano and Wolf, 2005*a*,*b*) to control the familywise error rate (FWER), i.e. the probability of rejecting at least one true null hypothesis.[34] To also discipline the false discovery rate (FDR), we calculate Anderson (2008) q-values for each word as well, then we define our list of *over-represented words* as the 58 words whose p- and q-values are *both* below 5 percent.[35]

Using this criterion, Table 2 displays the 36 words (out of 172) that are significantly over-represented in female-only jobs (left column) and the 22 words that are significantly over-represented in male-only jobs (right column). Table 2 also reports each word's regression coefficient in parentheses. To simplify the presentation, the panels of Table 2 list the words according to the six categories discussed in the previous section. Starting

[34] As described in Clarke et al. (2020), the Romano-Wolf correction is considerably more powerful than earlier multiple-testing procedures such as Bonferroni (1935) and Holm (1979) because it uses resampling methods to account for the dependence structure of the test statistics.

[35] The overlap between the words satisfying the p- and q-criteria is extremely high. Specifically, the 58 words satisfying $p < .05$ condition are a subset of the 62 words satisfying the $q < .05$ condition.

first with the standardized (PIACC) skills, we can see that literacy skills, such as *listening*, *writing*, *speaking* and *documentation*, and interpersonal 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 *collecting*. Male applicants see more jobs that require problem-solving skills such as *planning*, *decision-making*, and *engineering*, and influencing skills such as *leadership*, *charge* and *supervise*. These findings coincide with previous studies of the gender-skills gap (Pető and Reizer, 2021; Black and Spitz-Oener, 2010) which document that women are more likely to execute tasks and plans (in contrast to making plans or decisions).

Turning to the *Benefits* panel, only-to-women jobs are more likely to mention *marriage leave*, *maternity leave*, *parental leave*, *social security*, *maternity insurance* and *medical insurance* while only-to-male jobs emphasize *commuting friendly* and providing *shuttle*, *commission*, *injury insurance*, *allowance*, *free meal*, *reward* and *stock*. In 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 findings 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., 2023). Under *Company and Rank*, workplace *atmosphere* and *training* are 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 a *certificate*. Only-to-men jobs are more likely to 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 *entrepreneurial*, and able to work under *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.

3.5 Learning from Words 3 : 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 do these words reinforce commonly-held gender stereotypes? In this Section we exploit four data sources that are external to our job boards to assess which, if any, of our over- and under-represented words are associated with widely-held gender stereotypes. We then use stereotype scores from these external sources to quantify the overall amount of stereotypically male and female content in the jobs recommended to men versus women.

Our first external data source comprises three published papers that have identified gender-stereotypical words in job ads.[36] Gaucher et al. (2011) assembled a list of masculine and feminine words from published lists of agentic, communal, masculine, and feminine words (Appendix A in their paper) and showed that including these words in job ads affects readers' perceptions of gender representation in the 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 allowed the authors to train text analysis and machine learning techniques to predict the effect of observing a particular word on the probability the ad explicitly requests only male or female applicants.[37] Our first external list of male and female words is the subset of our 172 most common words that appear in any of the lists compiled in these three papers. The resulting words are listed in Appendix C3.

[36] There is a large literature on gendered language in linguistics (Fitzpatrick et al., 1995; Gastil, 1990; Lindqvist et al., 2019), political science (Roberts and Utych, 2020) and psychology (Bem, 1981; Hoffman and Hurst, 1990; Rudman and Kilianski, 2000). The vast majority of this literature focuses on contexts other than jobs, however, such as behavior in daily life and support for public policies.

[37] 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 of whether the job ad text is predictive of an employer's explicit male or female preference using a multinomial logistic regression classifier.

Our second and third approaches are based on two surveys in which we showed respondents our full list of 172 most common words, and elicited respondents' beliefs about whether a recruiter posting an ad containing the word was most likely seeking a man or a woman for the position (or had no preference). Our English-language survey was conducted on Amazon Mechanical Turk (MTurk); the Chinese version was conducted on wenjuanxing.com, a platform that provides professional online questionnaire survey services. Details of survey methodologies and the resulting word lists are provided in Appendices D3 and D4. In both surveys, the question was "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?". For each survey, stereotypically female (male) words were defined as words that were significantly associated with seeking women (men) at the 5 percent level or more.[38]

Fourth, we utilized a large language model (LLM) to identify words with gender associations.[39] Specifically, we prompted ChatGPT 4.0 with the request: "We are interested in investigating gendered words in the labor market. Can you categorize each word in the following six categories as neutral, male, or female?" The lists of male- and female-associated words provided by ChatGPT are displayed in Appendix C6.

To summarize the results of the preceding exercises, Table 3 reproduces Table 2's list of over-represented words and color codes the words to indicate their stereotype direction and intensity. Specifically, if a word is highlighted with dark red (like *assist* and *patient*), it was identified as stereotypically female by all four of our external sources. Words in bright red (like *administrative* and *facial*) are defined as stereotypically female in three approaches; those in light red are recognized as stereotypically female words two approaches, and pink indicates that the word was stereotypically female in just one approach. Male words are marked with blue colors, in which dark blue, bright blue, light

[38] The lists of gendered words from the two surveys are provided in Appendices D3 and D4, respectively.

[39] LLMs use machine learning methods to process large volumes of text. From their training on extensive corpora of text, LLMs internalize the structure and logic of human language to produce impressively human-like responses to questions. There have been some studies in social sciences using GPT to simulate human experiments (Aher et al., 2023; Argyle et al., 2023; Bybee, 2023; Brand et al., 2023; Hagendorff, 2023; Horton, 2023).

blue and pale blue represent stereotypical male words from four, three, two and one approaches, respectively. For example, *leadership* and *night work* are considered male in all four of our approaches. Overall, the dominance of red colors in the left panel and blue colors in the right panel of Table 3 clearly demonstrates that the words we have identified as over-represented in only-to-male and only-to-female jobs in our study are indeed correlated with commonly held gender stereotypes. 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.

3.6 Learning from Words 4: Quantifying Gender Differences in Stereotypical Ad Content

To quantify this stereotype-reinforcing effect of recommendation algorithms, we first assign stereotypical femaleness and maleness scores to all the words in Table 3, equal to the number of external datasets (zero to four) that classified the word as female (male). We then define the stereotypically female content of a *job ad* as:

$$S^f = \sum_{w \in ad} s^f_w, \tag{3}$$

where s^f_w represents the female stereotype score (zero to four) of each word in the ad. Our index of stereotypically male job ad content is defined analogously. Finally, we standardized S^f and S^m to have means of zero and standard deviations of 1 and replicated our main regressions (equation 2) with these standardized measures of stereotype intensity on the left hand side. The results are presented in Table 4, which shows that jobs targeted exclusively at women have 0.58 standard deviations more stereotypically female content (S^f) than ads shown exclusively to men. Conversely, only-to-men jobs are only 0.13 standard deviations more stereotypically male than only-to-women jobs.[40] To check if this

[40] As a robustness check, Table E7 replicates Table 4 using the entire list of 172 most common words on our four datasets (without restriction to being over-represented in either only-to-male and only-to-female job ads). While we do not expect this to change our results —because we did not restrict our word list with respect to the *direction* of over-representation—- it is not unreasonable to ask if this restriction affects the

pattern could be an artifact of the particular external sources we used to define gender-stereotypical words, Table C7 replicates Table 4, calculating stereotype scores using each of our four external sources individually: previous literature, MTurk Survey, Chinese Survey, and ChatGPT. The results are remarkably similar: all the S^f coefficients are significantly negative, all the S^m coefficients are significantly positive, and absolute value of the former is more than three times the latter in all cases.

Interestingly, this stark gender difference is consistent with findings from psychology and linguistics which suggest that cultural defaults affect how humans categorize items. Specifically, if a cultural default exists (for example, if the word 'worker' is more readily associated with men than women), then according to Smith and Zarate (1992) "a departure from the expected attribute value will be likely to attract attention and be the basis for categorizing the target". For example, Smith and Zarate's experiments show that people are more likely to categorize Black men as Black rather than as male, while White women are categorized as female, not as White. The algorithms on our boards may have internalized these tendencies, treating stereotypically male words as generic job characteristics and therefore as less informative of which gender is better matched to the job than female words.[41]

3.7 The Evolution of Difference Rates and Gender Gaps Across Experimental Rounds

Having documented the overall differences between the jobs that were recommended to women versus men, we now describe how these differences evolved across the rounds of

Table 4 results. Except for a difference in coefficient magnitudes related to the re-scaled dependent variable, Table E7's estimates are very similar to Table 4's.

[41] In Figure E2 (parts e and f) we explore heterogeneity in these stereotypical content gaps by applicant age, by femaleness of the occupation-industry cell, and by position level. Most dramatically, gender gaps in stereotypically female ad content are greatest in predominantly female job types. As Table E5 shows, this is related to the fact that jobs recommended exclusively to women in female-dominated types of work have extremely high levels female-stereotyped content (0.98 standard deviations more than an average recommended job). Similar patterns are present for stereotypically male content, but the magnitudes are much smaller, consistent with the notion that female content plays a much stronger role in recommendations than male content.

our experiment.[42] We begin with our most basic difference measure, the (set) difference rate; its evolution across experimental rounds is summarized in Figure 3. In Round 0 (before workers apply to any jobs or take any other actions), the set difference rate is 9.74 percent. Two weeks after the release of workers' profiles, this rate increases to 12.15% in Round 1.1. After workers apply to those 10 jobs, the difference rate rises to 12.90% in Round 1.2. The difference rate continues to rise both within and across rounds, reaching 17.60% in round 3.2. A linear regression of the difference rate on a round indicator shows a highly significant increase ($p = .000$).[43] The especially large increase between Rounds 0 and 1.1 (2.41 percentage points, or 25 percent) is also highly significant ($p = .000$).

Figures 4(a) to (d) graph the evolution of Table 1's wage, education, experience, and firm size gaps across the Rounds of our experiment. In contrast to the difference rate trends, none of the time trends in these gender gaps are statistically significant (p = .688, .060, .738 and .490 for wages, education, experience and firm size respectively). The same is true for the gender gap in the amount of stereotypically male content: p = 0.654 in Figure 4(f). Consistent with our finding that the jobs recommended to men versus women are differentiated mostly by the amount of stereotypically *female* content, however, we do see a time trend for female content (p = 0.014 in Figure 4(e)). Mirroring the disproportionate jump in the difference rate between rounds 0 and 1.1, this trend is concentrated there: In Round 0, jobs recommended (only) to men had 0.45 standard deviations less stereotypically female content than jobs recommended to women; this jumps downward to about 0.58 deviations and remains at around that level for the rest of the experiment. In sum, we see some 'growing apart' in the characteristics of jobs recommended to men versus women, but only for the amount of stereotypically female content, and only between Rounds 0 and 1.1.

We were surprised by the lack of growth in all these gender gaps after Round 1.1

[42] Notably, in this Section and in all our analyses of trends across experimental Rounds we make a small change to our estimation sample: Instead of using all 20 recommendations we collected from each profile in Round 0, we use only the first ten in the list. While this has only a minimal effect on our estimates, it ensures that our Round 0 observations are strictly comparable to the observations collected in Rounds 1.1, 1.2, etc.

[43] Specifically, our round indicator equals 0, 1, 2, 3, 4, 5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively. The sample size for these regressions is 7,746 pairwise difference rates).

because (a) we see substantial gender gaps in recommended jobs' stereotypical content in Round 0, and (b) our profiles apply to their top ten recommended jobs throughout the rest of the experiment. If algorithms use the jobs workers previously applied to to make recommendations, we would expect these Round 0 gender gaps to be magnified over the course of the experiment. We explore this idea further in Section 4, where we attempt to isolate which types of algorithmic processes are active on these boards.[44]

4 Mechanisms

Having demonstrated that job recommender algorithms perpetuate gender stereotypes, we now ask which specific processes and pieces of information are responsible for this outcome. Understanding these mechanisms is useful for at least two reasons. First, the direct use of certain types of information, such as race and gender, is prohibited in many jurisdictions and considered as unfair by many people (Kleinberg et al., 2018; Kuhn and Osaki, 2023). Knowing whether such information is being used could therefore be useful to policymakers and enforcement agencies. Second, isolating the likely source of stereotypical recommendations can help algorithm designers mitigate such recommendations, if desired.

In this Section, we study mechanisms in the order in which they become available to recommender systems during the course of our experiment. For each potential mechanism, we first describe how it works, based on survey articles (Al-Otaibi and Ykhlef, 2012; Hong et al., 2013; Siting et al., 2012) and on our personal experience with job boards. Then we discuss the evidence indicating whether each process (a) is operative on our job boards, and (b) contributes to the gender-recommendation gaps we documented in Section 3. As in Section 3, we group all four of our job boards together for these analyses.

[44] One additional factor that could cause changes in recommendations across experimental rounds is updates to the recommendation algorithms themselves. Based on our experience with job boards, we expect such changes to be infrequent. To test for this, we replicated Tables 1 and 4 in Table E6, controlling for job board $\times$ calendar week fixed effects. The results are very similar, suggesting little impact of algorithm updates.

Appendix D replicates this section's main results separately for each job board and finds very similar patterns, suggesting that our four job boards use a similar mix of algorithmic processes in their job recommender systems.

4.1 Round 0: Content-Based Processes

Consider Worker A, who has just published their profile on a job board. An algorithm charged with recommending suitable jobs for Worker A at this time faces a *cold-start problem*: Worker A has taken no actions (such as viewing, clicking, or applying to ads) that the algorithm can use to infer the worker's preferences, and no recruiters have yet had a chance to react to Worker A's resume. Since the only worker-specific information available to the algorithm is the contents of Worker A's resume, an essential first step in all possible processes is to match the content of Worker A's resume with other available content, such as existing job ads or other workers' resumes. There are three broad ways the algorithm could proceed, all of which are illustrated in Figure 5.

The most straightforward way to recommend jobs to new worker profiles is to search for job ads whose content matches the worker's resume (Channel 1 in Figure 5). The oldest and simplest way to do this mimics the manual, keyword based search options on the earliest job boards by matching the coded data fields in Worker A's profile to vacant jobs using a set of human-approved rules. For example, a typical rule would require the worker's experience to satisfy the job's experience requirement.[45] In addition to this *rules-based* approach, modern algorithms now also use natural language processing (NLP) methods to compute similarity scores between the complete text of a worker's resume and the entire text of a job ad. Importantly, however, both these resume-job matching methods can recommend systematically different types of jobs to our male and female profiles in Round 0 *only if* the matching methods use the worker's gender as an input. This is because our male and female profiles are identical except for gender.

[45] In more detail, on a job board we are familiar with, programmers dig factors and assign different weights to compute match degrees between workers' characteristics and jobs' requirements. Models are then trained to learn the weights, and the final weights are adjusted and approved by humans.

The other way for algorithms to make worker-specific recommendations in Round 0 is to use resume content to match Worker A to similar *workers* on the platform. For example, Worker A might be matched to Worker B because both mentioned "leadership" in their resumes. Then, algorithms could recommend the jobs Worker B applied to to Worker A (Channel 2). Alternatively, algorithms could recommend jobs posted by recruiters who previously reacted positively to Worker B (Channel 3). Like Channel 1, however, these Channels can only yield systematic gender gaps if the worker-worker content-matching algorithm that initiates this process uses our profiles' genders as inputs. In short, our experimental design allows us to test for a particularly strong form of gender bias in algorithm design: the explicit use of gender by the recommendation algorithm.

To conduct this test, Table 5 replicates Tables 1 and 4 using data from Round 0 only. Overall, the Round 0 results are surprisingly similar to the full-sample results, given that Round 0 recommendations only comprise about 15 percent of all the gender-specific recommendations in our data.[46] While the gender wage gap in job recommendations becomes statistically insignificant in Round 0, the tendency to steer men to larger firms and to jobs requiring more experience are highly statistically significant and very similar to the full sample, as are the large gender differences in both stereotypically male and stereotypically female job ad content.

In sum, our analysis of Round 0 mechanisms yields three conclusions: First, the algorithms on our job boards must be using content-based matching (either between workers and jobs, or between workers and workers), because content-based matches are an essential first step in all Round 0 processes. Second, compared to women, these content-based processes steer men to higher-ranked jobs in larger firms, and to job ads with more (less) stereotypically male (female) content. Finally, these content-based processes must be using the worker's declared gender as an input. The latter finding illustrates how algorithm audits can reveal whether black-box recommender systems are making explicit use of demographic markers like race or gender.

[46] Using the last column of Tables 4 and 5, the Round 0 share is 3,289/22,023 = 14.9%. While (by design) the 20 jobs in Round 0 recommendations comprise 25 percent of all recommendations collected, gender-specific recommendations are much less prevalent in Round 0, accounting for this 15 percent share.

We conclude our discussion of content-based processes (Channels 1-3) by noting that their potential influence is not confined to Round 0 of our experiment; in fact we expect these processes to operate in Rounds 1-3 as well, alongside the newly available processes we describe below. While in principle this could complicate our efforts to identify which new processes are at work in Rounds 1-3, the fact that content-based processes do not learn from the reference worker's actions –or from recruiters' reactions to the reference worker– mitigates this concern. Specifically, because the worker's resume is fixed across rounds, and because the aggregate content it's being matched to is likely stable as well, content-based processes cannot easily create increasing gender gaps across experimental rounds.[47] In contrast, growing gaps are a likely consequence of the newly-available, *adaptive* processes in Rounds 1-3, as described below.

4.2 Round 1.1: Recruiters' Resume Searches

Round 1.1 of our recommendation harvesting occurs two weeks after our resumes have been published, but before our resumes have taken any other actions such as viewing, clicking, or applying to a job ad. As illustrated in Figure 6, only one new process is available to generate recommendations at this time: Channel 4, which exploits the fact that recruiters on all our boards are free to search workers' published resumes during Interval 1 (the two weeks between Rounds 0 and 1.1).[48] A direct and natural way to exploit recruiters' resume search activity is to encourage workers to apply to recruiters who have already viewed, clicked, or downloaded their resumes (Köchling and Wehner, 2020). In addition, Round 1.1 algorithms can also use job-job similarity scores to direct workers to jobs that resemble the jobs that 'found' the worker.

[47] Aside from the fact that alternative processes are available in Rounds 1-3, we expect content-based recommendations to play a smaller role in Rounds 1-3 for two additional reasons. First, content-based processes have less 'bite' in Rounds 1-3 than Round 0 because they must rely on the new inflow of job ads to make novel recommendations to our profiles (Gregg and Petrongolo, 2005). Second, the relative quality of content-based recommendations should fall over time because these processes do not learn from the behavior of job seekers and recruiters.

[48] Because our Round 0 recommendations are collected immediately after publishing each profile, there is virtually no chance that resume searches by recruiters can affect these Round 0 recommendations.

To assess the importance of Channel 4, we turn to three types of evidence. The first is the sharp rise in the difference rate between Rounds 0 and 1.1, already noted in Figure 3. Since content-based matching should yield roughly constant difference rates over time, and since recruiter resume searches are the only new process to become available during this interval, this 25 percent increase (from 9.74 to 12.15, which is larger than all other between-period increases and highly statistically significant) strongly suggests that recruiters' resume searches play a role in directing different jobs to different workers. Second, to see if recruiters' resume searches raise the gender gap in the *types* of jobs recommended to men versus women, Table 6 combines recommendation data from Rounds 0 and 1.1 to quantify whether the gender gaps in job characteristics increased significantly between those two periods.[49] Interestingly, we see no statistically significant growth in our 'hard' measures of job characteristics: wages, education requirements, experience requirements, and firm size. We do, however, see increased gender stereotyping in the words contained in recommended jobs: Jobs recommended to men contained 0.45 standard deviations less female content than jobs recommended to women in Round 0; this rose to 0.58 units in Round 1.1. Jobs recommended to men contained 0.08 standard deviations more male content than jobs recommended to women in Round 0; this more than doubled to 0.18 units in Round 1.1.

Finally, panel A of Table 7 pursues the intuition that –if recruiters' inspections of a worker's resume are affecting job recommendations to that worker– the recommendations received by our paired male and female resumes should diverge more strongly among pairs whose resumes have been viewed more often. In more detail, the observations in Table 7 are the 1,120 gender pairs in our audit study. The regressor of interest is the total number of views a gender pair's resumes received during Interval 1 (the two-week period between Rounds 0 and 1) and the outcomes are the pair's set difference rate and the gender gaps in recommended job characteristics in Round 1.1.[50] According to column 1, more resume views are strongly associated with larger set difference rates: One more

[49] As we did in Figure 4, Table 6 uses only the first ten recommendations from Round 0 to ensure comparability between Rounds 0 and 1.1.
[50] Appendix F shows similar effects when resume views are disaggregated by profile gender.

profile view is associated with .0961 more gender-specific jobs per 100 recommendations, an elasticity of about 0.21.[51] The remaining columns of Table 7A tell a very different story, however: the number of resume views is not significantly associated with the gender gap in any of our recommended jobs' characteristics (both 'hard' and 'soft').

Summing up, we find strong evidence (based on difference rate patterns) that Channel 4 is *active* on our job boards, but only limited evidence that this process accentuates the gender gap in the types of jobs recommended to men and women. While the fact that job recommendation algorithms use information based on recruiters' reactions to a worker's resume may not seem surprising, we note that the preferences of the 'other' side of the market do not play a role in many important markets. For example, Amazon's consumer recommendation algorithms do not consider whether a pair of socks prefers to be purchased by consumer A or consumer B. The use of recruiters' revealed preferences to recommend jobs to *workers* thus signals that our job boards have not simply borrowed standard customer-search tools from retail markets; in contrast, they have incorporated the fact that labor markets are two-sided matching markets into their algorithms.

4.3 Round 1.2: Application-Based Processes

Round 1.2 occurs immediately after our profiles have submitted their first round of job applications; thus our Round 1.2 recommendations are the first ones that can learn from our profiles' application behavior. One such process (Channel 5, or "You Previously Applied To...") uses job-to-job similarity scores to recommend jobs to Worker A that are similar to the ones Worker A recently applied to.[52] For real workers –especially those who found the board's Round 0 suggestions unhelpful (or excessively stereotypical)– this channel provides an opportunity to 'teach' the board's algorithms to accommodate their individual preferences. In contrast –since the profiles in our experiment always follow the

[51] This is based on a mean of number of 15.83 profile views per pair during during Interval 1 and a mean difference rate in Round 0 of 7.34.

[52] In their interactions with real applicants, job boards can use several indicators of past worker interest, including views, clicks and applications to learn the types of jobs a worker is interested in. In our experiment, where the profiles' only activity is to apply to jobs, application behavior is the only available indicator.

board's recommendations– we would expect Channel 5, if it is operative, to magnify any gender gaps that were present in preceding experimental rounds.[53]

A second new mechanism that becomes available in Round 1.2 (Channel 6, or "Workers Who Applied to this Job Also Applied to...") is called Item-Based Collaborative Filtering (IBCF). IBCF is one of the most widely used recommendation algorithms; it is used to recommend Amazon products, Netflix movies, and iTunes music. Notably, IBCF does not use *any* content-based measures of similarity: In our context IBCF finds worker A's co-applicants at previously-applied jobs, then recommends the other jobs those co-applicants applied to, to worker A (Jannach et al., 2016). Therefore, *if* the real workers on our job boards disproportionately apply to gender-stereotypical jobs, IBCF (Channel 6) should have a similar effect to Channel 5: the gender gaps in recommended job characteristics should increase across successive rounds of our experiment.

To assess the importance of Channels 5 and 6, we focus on three specific points in our experiment where *only* application-based processes should be at work: *within* the Rounds of our experiment (i.e between Rounds 1.1 and 1.2, etc.), which are only a few seconds apart. For example, between Rounds 1.1 and 1.2, our profiles submit 10 new applications. Because the boards' algorithms already 'know' which other workers previously applied to those 10 jobs, Channels 5 and 6 ("more like this" and IBCF) should both be feasible. That said, the few seconds between Rounds 1.1 and 1.2 leave essentially no time for recruiters to react to those ten new applications.

To implement this idea, Table 8 uses only recommendations from Rounds 1 to 3, then splits this sample into the first 10 jobs and the second 10 jobs within each Round. We then compare the difference rates and gender gaps in job characteristics between first and last 10 jobs within each Round. We find that the difference rate in the last 10 jobs is 1 percent higher than in the first 10 jobs. This is consistent with a causal effect of past applications on the list of jobs seen by our profiles, but –as was discussed in Section 3.1– quasi-randomness

[53] These mechanical effects could be dampened, but not eliminated, by any *diversification* algorithms operative on a board. Diversification processes have been added to recommender systems in a variety of contexts (including job boards) to improve overall user satisfaction (Szpektor et al., 2013; Wu et al., 2016; Kunaver and Požrl, 2017).

and dispersion/decongestion processes could also account for these changes. Furthermore, none of gender gaps in recommended jobs' *characteristics* grow between these subrounds.[54] Contrary to our expectations, there is no evidence that application-based processes cause the types of jobs recommended to men versus women to diverge across the rounds of our experiment.

Overall, our evidence that application-based processes are operative is weak at best, and we find no evidence that such processes cause gender gaps in the types of jobs recommended to men versus women. A possible explanation of these findings is the simple fact that our profiles' past applications do not differ *very much*; in fact over 85 percent of past applications overlap because their job recommendations overlapped. Job boards may treat workers with this much application overlap as essentially the same. Another possibility is that –given that most recommended jobs already satisfy workers' requested wage, experience, education, occupation-industry and city criteria– the most important remaining job characteristic to workers is *whether the firm wants them*.[55]

4.4 Round 2.1: Recruiters' Reactions to Applications

A final potential way to make job recommendations (Channel 7) first becomes available in Round 2.1 of our experiment, which is the first time the algorithms can see how recruiters reacted to Worker A's previous job applications.[56] Specifically, algorithms can look at how recruiters reacted to Worker A's Round 1 applications during Interval 2, and use that information to recommend jobs in Round 2.1. Our first piece of evidence on this channel comes from Table F4, which shows that the mean number of profile views (per

[54] While the standard errors for some of these comparisons are large, Table 8 rules out even fairly small effects relative to the baseline gender gaps in experience and stereotypical ad content.

[55] Related, the boards may have decided to link their recommendations primarily to the employers' preferences rather than the workers' because employers are the paying customers. Finally, the boards may have decided to rely mostly on workers' keyword queries rather than their application behavior to customize their job recommendations. Our profiles do not make any keyboard queries.

[56] Because the recommendations issued in Round 1.2 occur immediately after workers have submitted their Round 1.1 applications, it is highly unlikely that recruiter reactions to those applications can affect the Round 1.2 recommendations.

pair) declines monotonically over the course of the experiment, from 15.83 in Interval 1 to 12.57 in Interval 4.[57] Notably, this decline occurs despite the fact that, starting in Round 2.1, recruiters –who always have the option to search for resumes– now have a new source of candidates to consider: workers who apply to their job postings. On the surface, this suggests a relatively modest role for Channel 7, which relies on those applications.

Our second piece of evidence is presented in panels B and C of Table 7, which replicate panel A for resume views during Intervals 2 and 3, finding a similar pattern: more profile views during these Intervals are again associated with a larger subsequent difference rate between men's and women's recommendations, with elasticities of 0.09 and 0.10 respectively (compared to 0.21 in Interval 1). Also like Interval 1, these resume views do not appear to magnify gender gaps in the *types* of jobs that are recommended.[58] Thus, it appears that recruiters' resume views continue to affect job recommendations in Rounds 2 and 3 of the experiment, but (as in Round 1) do not introduce gender biases. Note however that our data do not allow us to distinguish resume views that resulted from recruiter searches from views that resulted from workers' applications, so Tables 7B and C do not provide definitive evidence in favor of Channel 7. This noted, our results in the section strongly suggest that recruiters' reactions to a given worker's resume continue to raise within resume-pair difference rates in Rounds 2 and 3 of the experiment. They do so, however, without magnifying gender gaps in the characteristics of the recommended jobs.

5 Discussion

Personalized recommender systems have become indispensable tools that help people find items, friends, romantic partners, and information that suit their interests and pref-

[57] These declines are also evident in the views of the male versus female profile separately. In all cases, women's profiles were viewed about 10 percent less often than men's, indicating a bias among human recruiters (or in the boards' resume-search algorithms) that does not appear to create gender gaps in the boards' job recommendations to workers.

[58] There is one exception to this: a small and marginally significant wage gap in Interval 2.

erences. Depending on the algorithms they use, 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 show different jobs to identical male and female job seekers, though most of the recommended jobs overlap between the two genders. In the jobs that differ, we find that women are steered towards jobs that pay 1.54 percent less, are in smaller firms, and require 7.19 percent less experience, suggesting a lower rank in firms' hierarchies. In addition, we find that the recommender systems steer both men and women towards job ads that contain words that are stereotypical for their gender; this effect is especially strong for stereotypically female content, which is 0.58 standard deviations more prevalent in jobs directed at women than at men.

We also explore the mechanisms responsible for these gaps and find strong evidence that *content-based* processes (which use rules and natural language processing to match resumes to job ads) are their primary driver. Intuitively, this is because these gender-recommendation gaps are present very early in our experiment (before any other processes are feasible), and because the gaps do not grow appreciably in later rounds. Due to the design of our experiment, we can also conclude that these content-based methods must be using the worker's gender as a direct input. Turning to *action-based* processes, we find strong evidence that the boards use *recruiters' views* of a worker's resume to recommend jobs to that worker, but only mixed evidence that these reactions contribute to the gender gaps we see. Finally, we detect little evidence that *workers' previous applications* affect the recommendations they receive, though we caution that we do not manipulate our profiles' application behavior to directly test for such effects.[59]

[59] To test for such effects directly, future investigators could construct identical female resumes, some of

As the first audit of job recommender systems that we are aware of, our paper has some noteworthy limitations. The first is that our outcome variables are confined to a very early stage of the job search process: Which jobs are *displayed* to male versus female job seekers? While seeing a job is likely an important precursor to getting it, extensions of our approach could merge it with data from later stages of the recruiting process (especially hiring) to examine these effects. Second, we have not discussed which types of algorithmic changes might reduce stereotyping, nor have we discussed whether such changes would have undesirable side effects. While other studies have found that de-biasing the job matching process can integrate jobs and workplaces with no detectable efficiency costs (Kuhn and Shen, 2023; Card et al., 2024), experimenting with recommender systems is needed to assess whether, for example, de-biasing reduces mean match relevance on other dimensions workers care about.

which follow the boards' recommendations when applying, while others apply only to jobs with specific characteristics, such as advertised hours flexibility or a stated leadership role. If applications guide recommendations, these two resume types should receive similar recommendations before they submit any applications, but diverge in the expected ways in later experimental rounds.

References

Abraham, L., Hallermeier, J. and Stein, A. (2024), 'Words matter: Experimental evidence from job applications', *Journal of Economic Behavior & Organization* **225**, 348–391.

Aher, G. V., Arriaga, R. I. and Kalai, A. T. (2023), Using large language models to simulate multiple humans and replicate human subject studies, *in* 'International Conference on Machine Learning', PMLR, pp. 337–371.

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.

Anderson, M. L. (2008), 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American Statistical Association* **103**(484), 1481–1495.

Andrew, C. (2020), Glassdoor's job and hiring trends for 2020, Report, Glassdoor Economics Research. Accessed Nov 1, 2023.
URL: *https://research.glassdoor.com/site-us/wp-content/uploads/sites/2/2019/11/Job_Hiring_Trends_2020-FINAL-1-1.pdf*

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

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C. and Wingate, D. (2023), 'Out of one, many: Using language models to simulate human samples', *Political Analysis* **31**(3), 337–351.

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

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.

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.

Board, C. (2022), How retailers can keep up with consumers, Report, The Conference Board-Lightcast Help Wanted OnLine. Accessed Dec 1, 2022.
URL: *https://www.conference-board.org/topics/help-wanted-online*

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**.

Bonferroni, C. E. (1935), 'Il calcolo delle assicurazioni su gruppi di teste', *Studi in onore del professore salvatore ortu carboni* pp. 13–60.

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

Brand, J., Israeli, A. and Ngwe, D. (2023), 'Using gpt for market research', *Harvard Business School Marketing Unit Working Paper* **23**(062).

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.

Burks, S. V., Cowgill, B., Hoffman, M. and Housman, M. (2015), 'The value of hiring through employee referrals', *The Quarterly Journal of Economics* **130**(2), 805–839.

Burn, I., Button, P., Corella, L. M. and Neumark, D. (2022), 'Does ageist language in job ads predict age discrimination in hiring?', *Journal of Labor Economics* **40**(3), 613–667.

Bustelo, M., Diaz, A. M., Lafortune, J., Piras, C., Salas, L. M. and Tessada, J. (2023), 'What is the price of freedom? estimating women's willingness to pay for job schedule flexibility', *Economic Development and Cultural Change* **71**(4), 1179–1211.

Bybee, J. L. (2023), 'The ghost in the machine: Generating beliefs with large language models', *arXiv preprint arXiv:2305.02823* .

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.

Card, D., Colella, F. and Lalive, R. (2024), 'Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions', *The Review of Economic Studies* (forthcoming).

Castilla, E. J. and Rho, H. J. (2023), 'The gendering of job postings in the online recruitment process', *Management Science* **69**(11), 6912–6939.

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.

Chong, D. (2020), A deep dive into netflix's recommender system, Report, Towards Data Science. Accessed Feb 1, 2023.
URL: *https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48*

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.

Clarke, D., Romano, J. P. and Wolf, M. (2020), 'The romano–wolf multiple-hypothesis correction in stata', *The Stata Journal* **20**(4), 812–843.

Coffman, K., Collis, M. R. and Kulkarni, L. (2023), 'Stereotypes and belief updating', *Journal of the European Economic Association* p. jvad063.

Cortés, P., Pan, J., Pilossoph, L., Reuben, E. and Zafar, B. (2023), 'Gender differences in job search and the earnings gap: Evidence from the field and lab', *The Quarterly Journal of Economics* **138**(4), 2069–2126.

Cowgill, B. and Tucker, C. E. (2020), 'Algorithmic fairness and economics', *Columbia Business School Research Paper* .

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.

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.

Flory, J. A., Leibbrandt, A., Rott, C. and Stoddard, O. (2021), 'Increasing workplace diversity: Evidence from a recruiting experiment at a fortune 500 company', *Journal of Human Resources* **56**(1), 73–92.

Friebel, G., Heinz, M., Hoffman, M. and Zubanov, N. (2023), 'What do employee referral programs do? measuring the direct and overall effects of a management practice', *Journal of Political Economy* **131**(3), 633–686.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2022), 'Predictably unequal? the effects of machine learning on credit markets', *The Journal of Finance* **77**(1), 5–47.

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.

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.

Gee, L. K., Jones, J. and Burke, M. (2017), 'Social networks and labor markets: How strong ties relate to job finding on facebook's social network', *Journal of Labor Economics* **35**(2), 485–518.

Gregg, P. and Petrongolo, B. (2005), 'Stock-flow matching and the performance of the labor market', *European Economic Review* **49**(8), 1987–2011.

Hagendorff, T. (2023), 'Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods', *arXiv preprint arXiv:2303.13988* .

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.

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.

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.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian journal of statistics* pp. 65–70.

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

Horton, J. J. (2023), Large language models as simulated economic agents: What can we learn from homo silicus?, Technical report, National Bureau of Economic Research.

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.

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. (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 Osaki, T. T. (2023), When is discrimination unfair?, Technical report, National Bureau of Economic Research.

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. and Shen, K. (2023), 'What happens when employers can no longer discriminate in job ads?', *American Economic Review* **113**(4), 1013–1048.

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.

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.

MacKenzie, I., Meyer, C. and Noble, S. (2013), How retailers can keep up with consumers, Report, McKinsey Insights. Accessed Nov 1, 2023.
URL: *https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers*

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

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*

Pallais, A. and Sands, E. G. (2016), 'Why the referential treatment? evidence from field experiments on referrals', *Journal of Political Economy* **124**(6), 1793–1828.

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.

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.

Romano, J. P. and Wolf, M. (2005*a*), 'Exact and approximate stepdown methods for multiple hypothesis testing', *Journal of the American Statistical Association* **100**(469), 94–108.

Romano, J. P. and Wolf, M. (2005*b*), 'Stepwise multiple testing as formalized data snooping', *Econometrica* **73**(4), 1237–1282.

Roussille, N. (2021), 'The central role of the ask gap in gender pay inequality'.

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

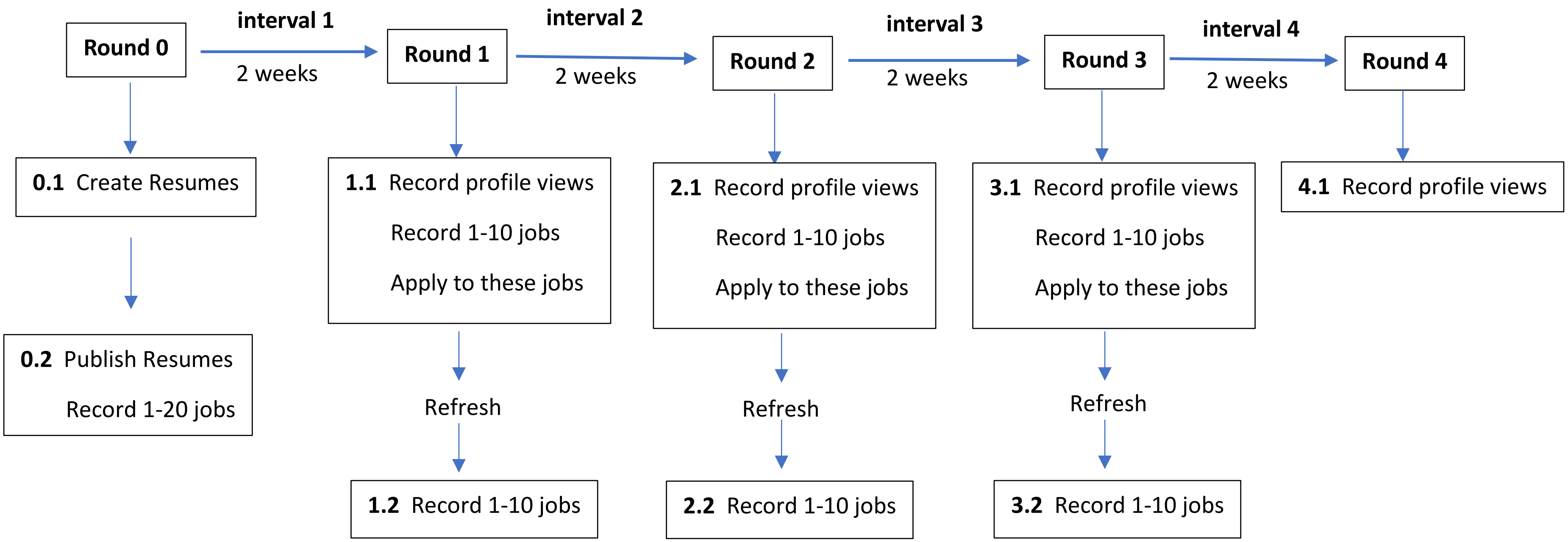
Siting, Z., Wenxing, H., Ning, Z. and Fan, Y. (2012), Job recommender systems: a survey, *in* '2012 7th International Conference on Computer Science & Education (ICCSE)', IEEE, pp. 920–924.

Smith, E. R. and Zarate, M. A. (1992), 'Exemplar-based model of social judgment.', *Psychological Review* **99**(1), 3.

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.

Figure 1: Experimental Timeline



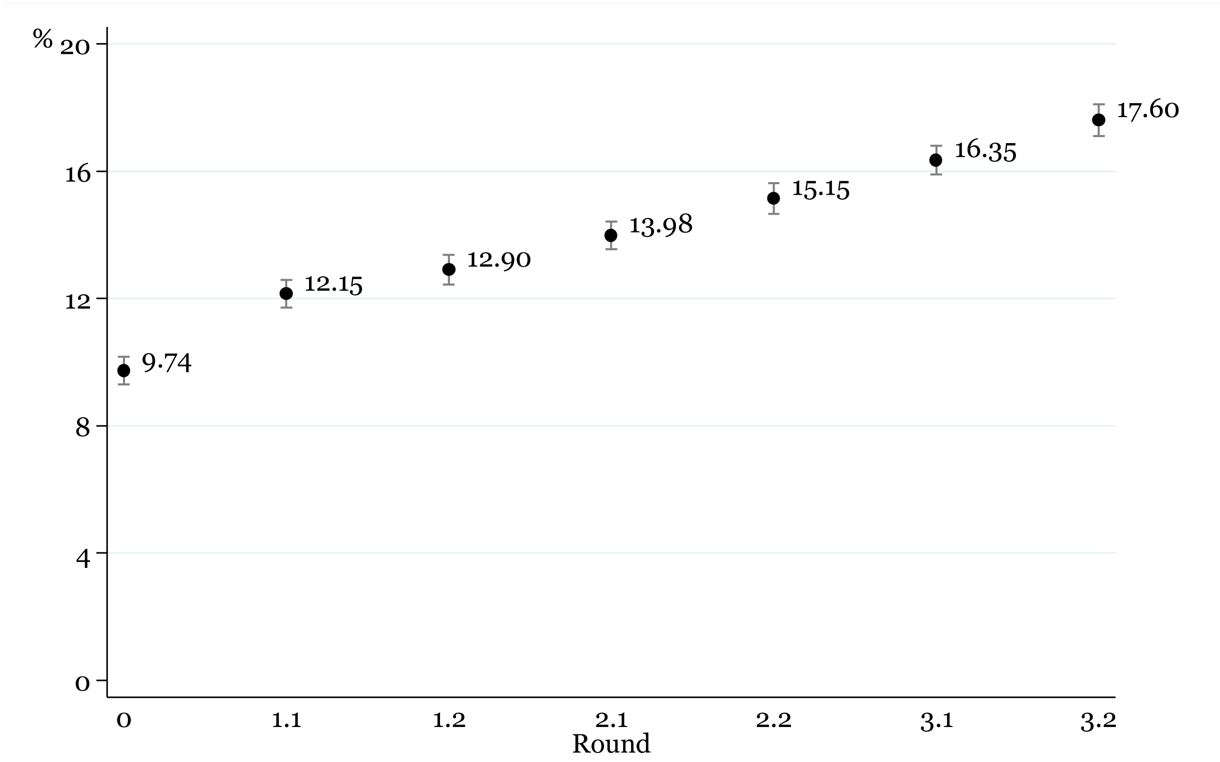
Notes:

1. The two profiles in each gender pair follow the same timeline.
2. In rounds 1.1, 2.1, and 3.1, fictitious workers apply to the top 10 jobs in their customized list of job recommendations.

Figure 2: Word Cloud from Job Ads

Note: The word cloud is based on the extracted words in the job descriptions from 81,231 recommended job advertisements, and the size corresponds to the word frequency. The Chinese version is shown in Appendix D Figure D1.

Figure 3: Set Difference Rate by Experimental Rounds

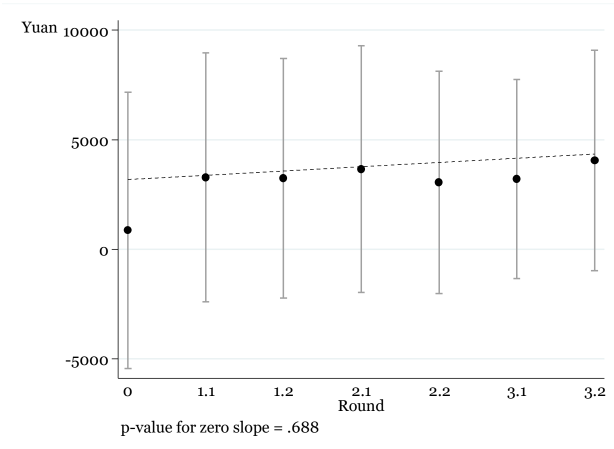


Notes:

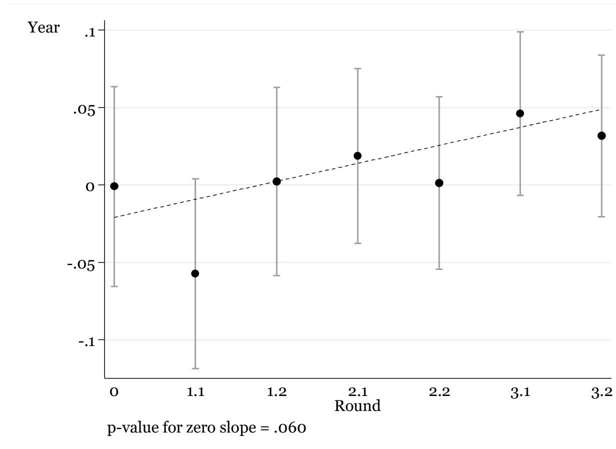
1. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
2. Whiskers show 95% confidence intervals.
3. The increase in the difference rate between Rounds 0 and 1 is 2.41 percentage points, which is significant at the 1% level ($t = 7.69$, $p = 0.000$).
4. A regression of the pair-level difference rate on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively) yields a coefficient of 1.223 ($p = 0.000$; $N = 7{,}746$).

Figure 4: Gender Differences in Job Characteristics by Experimental Rounds

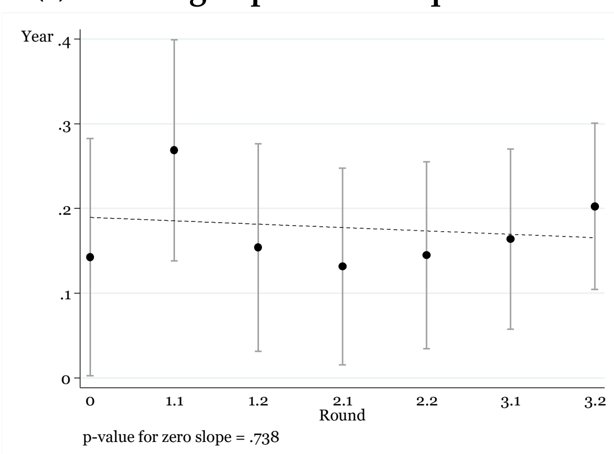
(a) Posted Wage



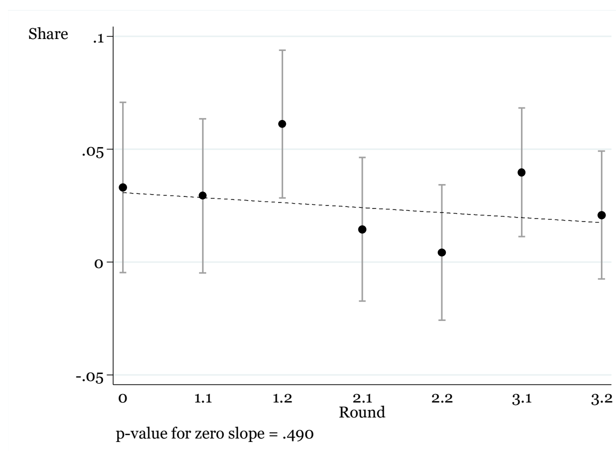
(b) Education Requirements



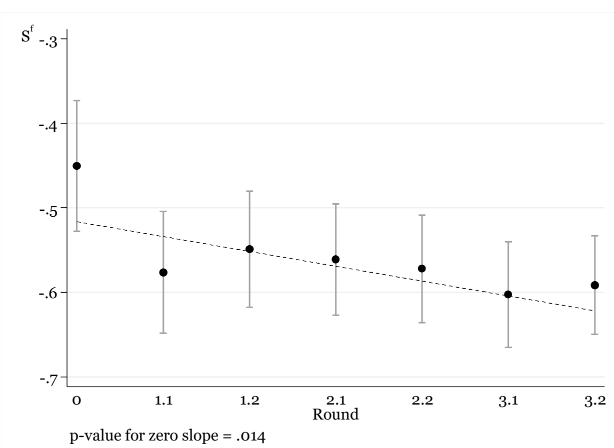
(c) Working Experience Requirements



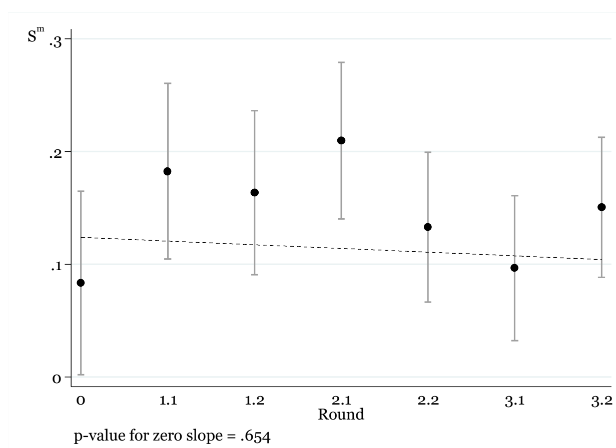
(d) Firm Size (≥1000)



(e) Stereotypically Female Content (S^f)



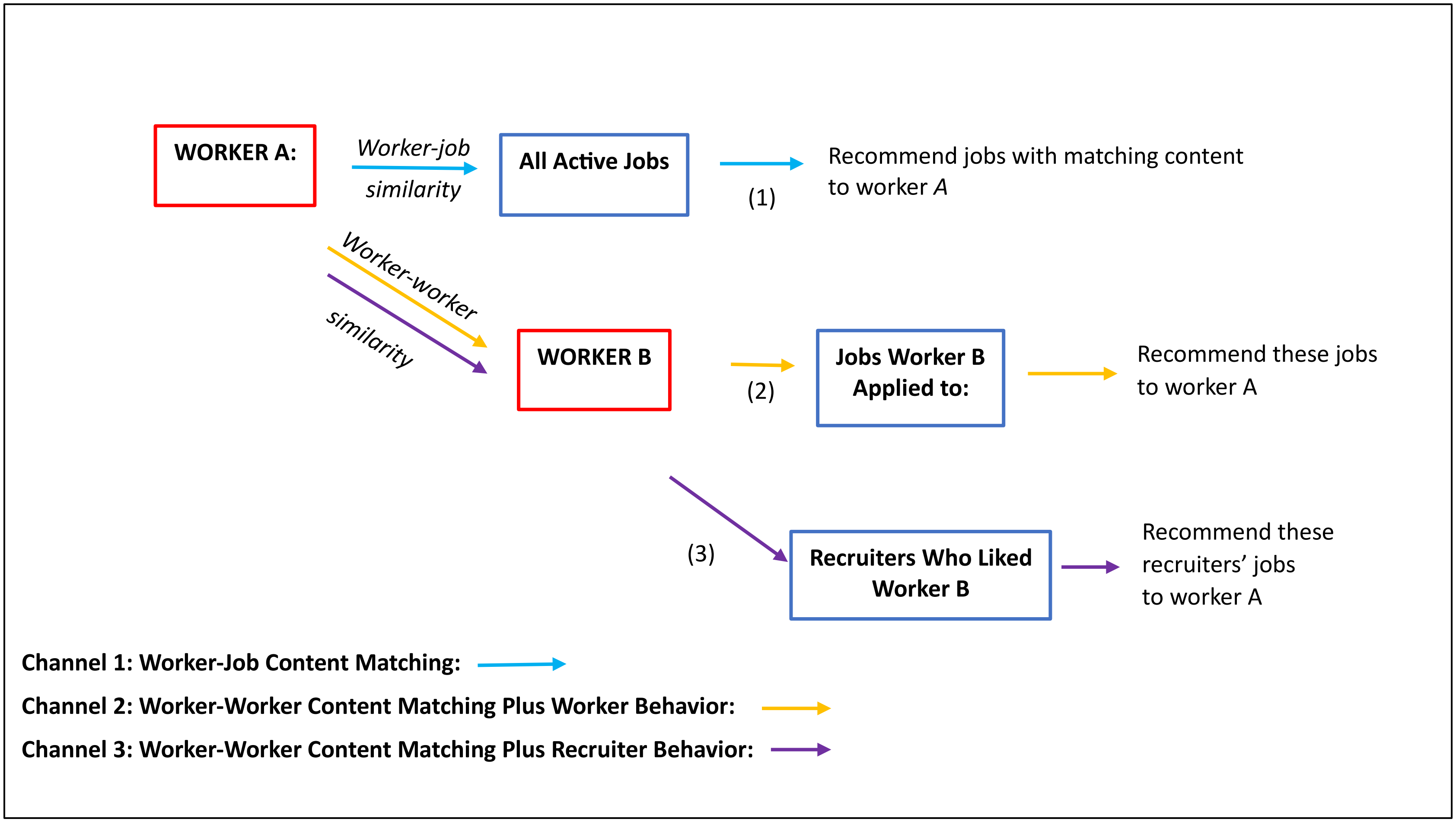
(f) Stereotypically Male Content (S^m)



Notes:

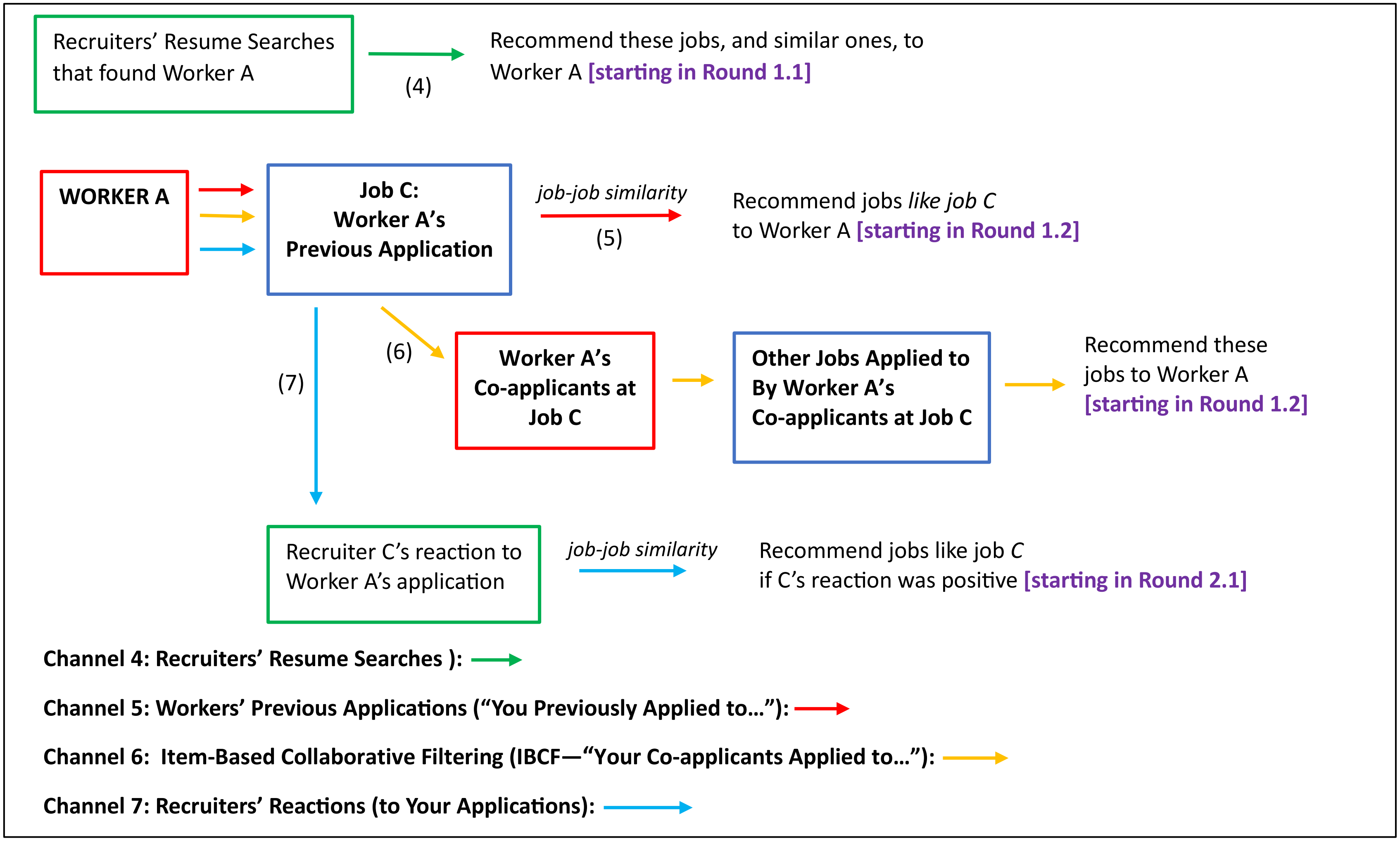
1. This figure presents the gender differences in job characteristics between male-only and female-only jobs across each round.
2. Figure 4 runs the regressions in Tables 1 and 4 separately for each of rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2, showing the resulting *Male* coefficients and 95% confidence intervals. To ensure comparability across Rounds, the sample for Round 0 includes only the first ten ads in each worker's recommendation list.
3. All panels show regression lines of the gender gap in the outcome on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively). Sample sizes vary between 6,160 and 6,910 depending on the outcome. p-values for a slope of zero are shown below each panel.

Figure 5: Recommendation Algorithms Available to a Newly Created Profile (Round 0)



Note: Red boxes refer to workers, blue boxes refer to jobs or recruiters. Arrows show the flow of an algorithm. For example, Channel 1 starts with Worker A's job profile, then searches the current stock of jobs for ones that contain the same or similar content.

Figure 6: Additional Algorithms available to an "Experienced" Profile (Rounds 1-3)



Note: Red boxes refer to workers, blue boxes refer to jobs, green boxes refer to recruiters. Arrows show the flow of an algorithm. For example, Channel 5 starts with Worker A, finds the jobs Worker A applied to, then recommends job ads with similar content.

Table 1: Gender Differences in Characteristics of Recommended Jobs

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)
Male	3,118***	0.0175	0.1656***	0.0276***
	(1,023)	(0.011)	(0.022)	(0.006)
N	21,262	19,900	21,922	22,023
R^2	0.609	0.449	0.390	0.164

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. The regression sample is all gender-specific jobs, i.e. the jobs that are only recommended to one worker in a gender pair, combining all rounds of the experiment. Thus, *Male* indicates that only the male worker in the pair saw the job.
2. The total number of gender-specific jobs is 22,023. Columns (1)-(3) use fewer observations because of missing information for some ads.
3. Firm size is recorded in different intervals on different boards, but 1000 is a cut point on all four boards. Overall, 36.93% of jobs in this regression sample were in firms with 1000 or more workers.
4. All regressions control for profile pair fixed effects.

Table 2: Over-Represented Words in Jobs Recommended to Women versus Men

	Words that are over-represented in jobs recommended to women	Words that are over-represented in jobs recommended to Men
Skills	listen (-0.0187), speak (-0.0601), write (-0.0556), documentation (-0.0180), data (-0.0397), chat tools (-0.0308), cooperation (-0.0425), communication (-0.0380), assist (-0.0812), negotiation (-0.0221), administrative (-0.0354), collect (-0.0586)	decision-making (0.0184), planning (0.0338), engineering (0.0173), leadership (0.0471), charge (0.0123), supervise (0.0310)
Benefits	marriage leave (-0.0725), parental leave (-0.0188), maternity leave (-0.0619), medical insurance (-0.0229), social security (-0.0281), maternity insurance (-0.0117)	commission (0.0262), stock (0.0212), allowance (0.0337), reward (0.0224), meal (0.0268), shuttle (0.0260), commute friendly (0.0356), injury insurance (0.0070)
Work Timing and Location	eight-hour working (-0.0204), flexible (-0.0438), weekly break (-0.0571), regular hour (-0.0284)	nightwork (0.0032), work overtime (0.0174), long travel (0.0069)
Company	training (-0.0476), atmosphere (-0.0288)	public company (0.0197)
Other Qualifications	certificate (-0.0125), new grad (-0.0195), non-experience (-0.0060)	science&engineering (0.0193), no crime history (0.0181)
Personality, Age, and Appearance	careful (-0.0930), patient (-0.0264), active (-0.0183), outgoing (-0.0342), generous (-0.0109), punctual (-0.0307), figure (-0.1835), temperament (-0.0985), facial (-0.0152)	entrepreneurial (0.0092), pressure (0.0426)

Note: Table 2 displays the 58 words (out of 172) that are significantly over-represented in male-only or female-only jobs. Over-representation is measured using the regressions in equation (2), where the outcome variable is a dummy for the word was present in the job ad. Regression coefficients are reported in parentheses, with negative (positive) coefficients indicating the word was over-represented in jobs recommended to women (men). To correct for multiple hypothesis testing, we include only words whose Romano-Wolf (2005*a*,*b*) *p*-values *and* Anderson (2008) *q*-values are both below 5 percent.

Table 3: Over-Represented Words in Job Ads and Gender Stereotypes

	Words that are over-represented in jobs recommended to women	Words that are over-represented in jobs recommended to Men
Skills	listen, speak, write, documentation, data, chat tools, cooperation, communication, assist, negotiation, administrative, collect	decision-making, planning, engineering, leadership, charge, supervise
Benefits	marriage leave, parental leave, maternity leave, medical insurance, social security, maternity insurance	commission, stock, allowance, reward, meal, shuttle, commute friendly, injury insurance
Work Timing and Location	eight-hour working, flexible, weekly break, regular hour	nightwork, work overtime, long travel
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	entrepreneurial, pressure

Note:

1. Stereotypically female (male) words are highlighted in red (blue). Color intensity indicates the number of external sources (1-4) that classify the word as stereotypical. For example, word w has a female stereotype score (s_w^f) of 4 if all four external sources define it as female (e.g. *patient*). Thus, *administrative*, *careful*, and *flexible* have female stereotype scores of 3, 2, and 1. Similarly, *leadership*, *supervise*, *no crime history*, and *decision-making* have s_w^m scores of 4, 3, 2, and 1 respectively.

Table 4: Gender Differences in the Stereotypical Content of Job Ads

	(1) Index of Stereotypically Female Content (S^f) (standardized)	(2) Index of Stereotypically Male Content (S^m) (standardized)
Male	-0.5760***	0.1322***
	(0.012)	(0.013)
N	22,023	22,023
R^2	0.297	0.117

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Sample and regression specification are the same as Table 1: Sample is all only-to-male jobs plus all only-to-female jobs. *Male* indicates the ad was only seen by the male profile in a gender pair. All regressions include pair fixed effects.
2. Our index of stereotypically female *ad* content is calculated as: $S^f = \sum_{w \in \text{ad}} s_w^f$, where s_w^f is the female stereotype score of each word in the ad, defined in Table 3. Stereotypical male ad content, S^m, is constructed analogously. In Table 4, S^f and S^m are standardized to have a mean of zero and standard deviation of 1. Thus, column 1 indicates that (compared to the ads that only the female profile saw) the ads displayed only to male profiles contained words that were .576 standard deviations less stereotypically female.
3. In the sample of all job ads, the (unstandardized) means of S^f and S^m were 5.79 and 8.37 respectively. Thus, a randomly selected job ad contained 2.58 more stereotypically male words than female words. In ads seen only by one member of a gender pair (i.e. the Table 4 sample), these means were 6.67 and 9.09 respectively.

Table 5: Gender Differences in Characteristics of Recommended Jobs in Round 0

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	2,389	0.0154	0.1699***	0.0325**	-0.5493***	0.1166***
	(2,502)	(0.028)	(0.058)	(0.016)	(0.032)	(0.034)
N	3,177	2,934	3,278	3,289	3,289	3,289
R^2	0.741	0.681	0.559	0.405	0.452	0.313

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Columns 1 to 4 replicate the regressions in Table 1 based on the sample of 20 jobs recommended to fictitious applications in Round 0. *Male* indicates that the job ad was displayed only to the male profile in a gender pair.
2. Columns 5 and 6 replicate Table 4 based on the sample of 20 jobs recommended to fictitious applications in Round 0. S^f and S^m are standardized to mean 0 and standard deviation 1.
3. Firm size is recorded in different intervals on different boards, but 1000 is a cut point on all four boards. Overall, 39.34% of jobs in this regression sample were in firms with 1000 or more workers.
4. All columns control for pair fixed effects.

Table 6: The Growth of Gender Differences in Recommended Job Characteristics Between Rounds 0 and 1.1

Male coefficient in:	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
1. Round 0 (β_1)	1,559.2112	0.0123	0.1420*	0.0324*	-0.4507***	0.0828*
	(3,266.027)	(0.035)	(0.074)	(0.019)	(0.040)	(0.043)
2. Round 1.1 (β_2)	3,528.4602	-0.0568*	0.2686***	0.0302*	-0.5756***	0.1832***
	(2,909.162)	(0.031)	(0.066)	(0.017)	(0.036)	(0.038)
3. Growth in the *Male* Coefficient between Rounds 0 and 1.1 ($\beta_2 - \beta_1$)	**1,969 (p = 0.653)**	**-0.0691 (p = 0.139)**	**0.1266 (p = 0.200)**	**-0.0022 (p = 0.932)**	**-0.1249** (p = 0.020)**	**0.1003* (p = 0.081)**
N	4,710	4,419	4,872	4,899	4,899	4,899

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Table 6 extends equation 2 to estimate the following regression: $Y = \beta_0 + \beta_1(Male * R0) + \beta_2(Male * R1.1) + \beta_3 R1.1 + \beta_4 X_p + \varepsilon$, where observations are gender-specific job recommendations made in Rounds 0 and 1. R0 and R1.1 are Round indicators and *Male* indicates the job was only displayed to the male profile in a pair. All regressions include pair fixed effects, X_P. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list.
2. Rows 1 and 2 show the gender gaps in each outcome in Rounds 0 and 1.1 respectively (β_1 and β_2).
3. Row 3 shows the increase in the gender gap ($\beta_2 - \beta_1$) and the p value from the F test of $\beta_2 - \beta_1 = 0$.

Table 7: Effects of Profile Views during Intervals 1, 2, and 3 on Subsequent Gender Recommendation Gaps

	Gender Gap (*male – female*) in:						
	(1) Difference Rate (%)	(2) Posted Wage (RMB)	(3) Requested Education (years)	(4) Requested Experience (years)	(5) Firm Size (≥1000)	(6) Stereotypically Female Content (Sf)	(7) Stereotypically Male Content (Sm)
A. Interval 1							
Views	0.0961***	-221	0.0043	0.0006	-0.0007	-0.0049	-0.0058
	(0.017)	(231)	(0.004)	(0.006)	(0.002)	(0.004)	(0.004)
N	1,118	1,073	1,016	1,089	1,090	1,090	1,090
R^2	0.323	0.005	0.005	0.010	0.010	0.064	0.041
B. Interval 2							
Views	0.0741***	404*	0.0008	-0.0021	0.0009	-0.0031	-0.0043
	(0.018)	(238)	(0.003)	(0.006)	(0.002)	(0.003)	(0.004)
N	1,100	1,078	1,049	1,088	1,089	1,089	1,089
R^2	0.346	0.011	0.005	0.006	0.007	0.028	0.032
C. Interval 3							
Views	0.1037***	201	-0.0031	0.0040	0.0013	-0.0019	0.0030
	(0.019)	(222)	(0.003)	(0.005)	(0.001)	(0.003)	(0.003)
N	1,095	1,082	1,054	1,090	1,090	1,090	1,090
R^2	0.308	0.008	0.007	0.009	0.018	0.057	0.029

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Observations are profile pairs, consisting of an identical male and female profile.
2. The regressor, *Views*, is the total number of profile views for the pair (male plus female) during each Interval. The dependent variable in column 1 is the number of gender-specific jobs per 100 job recommendations received by the two applicants in each gender pair (difference rate*100) immediately after the Interval. In columns 2 to 6 the outcomes are the gender gaps (male – female) in those recommended jobs' characteristics.
3. Panel A regresses gender recommendation gaps during Round 1 (Rounds 1.1 and 1.2 combined) on the number of views the pair received during the preceding two weeks (interval 1). Panels B (C) regress gender gaps during Round 2 (3) on the number of views the pair received during interval 2 (3). All regressions control for the pair's age, the gender type of the pair's current (and sought) job, and job board fixed effects.
4. Mean profile views are 15.83, 14.66, and 13.38 in Intervals 1-3 respectively.
5. In column 1, elasticities of the difference rate with respect to profile views are 0.20, 0.09, and 0.10 in Intervals 1-3 respectively.

Table 8: Gender Gaps in Job Recommendations Within Rounds 1-3

		Gender Gap (*male – female*) in:					
	(1) Difference Rate (%)	(2) Posted Wage (RMB)	(3) Requested Education (years)	(4) Requested Experience (years)	(5) Firm Size (≥1000)	(6) Stereotypically Female Content (Sf)	(7) Stereotypically Male Content (Sm)
First: 1-10 (β_1)	13.97	3,485**	0.0052	0.1838***	0.0292***	-0.5807***	0.1602***
	(0.146)	(1,574)	(0.017)	(0.034)	(0.009)	(0.019)	(0.020)
Last: 11-20 (β_2)	15.06	3,159**	0.0200	0.1664***	0.0270***	-0.5727***	0.1492***
	(0.164)	(1,515)	(0.016)	(0.033)	(0.009)	(0.019)	(0.020)
Difference ($\beta_2 - \beta_1$)	**1.09*** (p = 0.000)**	**-326 (p = 0.882)**	**0.0148 (p = 0.527)**	**-0.0173 (p = 0.714)**	**-0.0022 (p = 0.867)**	**0.0079 (p = 0.767)**	**-0.0110 (p = 0.696)**
N	2,236	18,766	17,602	19,355	19,445	19,445	19,445

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Table 8 is based on a similar regression to Table 6 except that the two periods are first 10 recommendations in a Round (i.e. in Rounds 1.1, 2.1 and 3.1) and last 10 recommendations in a Round (Rounds 1.2, 2.2 and 3.2).
2. Rows 1 and 2 show the gender gaps in each outcome in first 10 jobs and last 10 jobs respectively (β_1 and β_2).
3. Row 3 shows the increase in the gender gap ($\beta_2 - \beta_1$) and the p value from the F test of $\beta_2 - \beta_1 = 0$.

Appendix A: Experimental Design

A1: Job Type Selection

As noted in Section 2.2, we selected 35 industry-occupation cells (*job types*) on each job board based on three criteria: the number of active job openings, the job type's dominant gender (female, gender-balanced, or male), and skill level (entry, middle, and high). The complete list of resulting jobs is provided in Tables A1.1-A1.4, along with the type's modal requested education level and major, and the workers' mean current wage.

Table A1.1: Selected Job Types in Job Board 1

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

Table A1.1, continued

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

Notes:

1. Gender (M, N, F) represents the job type's dominant gender.
2. Current wages (·,·) refer to the resume's current wages for (young, older) workers in 10,000 RMB, respectively.

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

Table A1.2, continued

F	E-Business	Web Customer Service	Entry	College	Marketing	(6, 8)
	Banking	Telephone Customer Service	Entry	College	Marketing	(6, 8)
	E-Business	Customer Service Manager	High	College	Marketing	(12, 14)
	Banking	Customer Service Manager	High	College	Marketing	(12, 14)
	E-Business	Accountant	Middle	Bachelor	Accounting	(9, 14)
	Internet	HR Specialist/Assistant	Entry	College	Econ&Management	(6, 9)
	Professional Services	HR Specialist/Assistant	Entry	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	Entry	College	Econ&Management	(7, 9)
	Internet	Administration Specialist/Assistant	Entry	College	Econ&Management	(7, 8)
	Internet	Administration Manager/Supervisor	High	College	Econ&Management	(11, 14)

Notes:

1. Gender (M, N, F) represents the job type's dominant gender.
2. Current wages (·,·) refer to the resume's current wages for (young, older) workers in 10,000 RMB, respectively.

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	Middle	Bachelor	Computer Science	(17, 24)
	Machine Manufacturing	Mechanical Engineer	Middle	Bachelor	Machinery	(16, 20)
	Computer Software	Software Engineer	Middle	Bachelor	Computer Science	(18, 24)
	Computer Software	Senior Software Engineer	High	Bachelor	Computer Science	(22, 26)
	Internet/E-Business	Operations Specialist	Entry	College	Computer Science	(10, 12)
	Internet/E-Business	Operations Manager/Supervisor	High	Bachelor	Computer Science	(13,19)
	Real Estate Development	Architect	Middle	Bachelor	Architecture	(14, 22)
N	Pharmaceuticals/Biotechnology	Sales Representative	Entry	College	Marketing	(10, 14)
	Securities/Investment Funds	Sales Representative	Entry	College	Marketing	(11, 14)
	Pharmaceuticals/Biotechnology	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 18)
	Internet/E-Business	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(12, 18)
	Securities/Investment Funds	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(12, 18)
	Pharmaceuticals/Biotechnology	Sales Director	High	Bachelor	Marketing	(17, 24)
	Internet/E-Business	Sales Director	High	Bachelor	Marketing	(16, 24)
	Commodity	Sales Director	High	Bachelor	Marketing	(16, 24)
	Internet/E-Business	Product Manager/Supervisor	High	Bachelor	Econ&Management	(14, 22)
	Computer Software	Product Manager/Supervisor	High	Bachelor	Econ&Management	(14, 22)
	Internet/E-Business	Project Manager/Supervisor	High	Bachelor	Econ&Management	(14, 22)
	Computer Software	Project Manager/Supervisor	High	Bachelor	Econ&Management	(14, 22)
	Commodity	Marketing Manager/Supervisor	High	Bachelor	Marketing	(13, 22)
	Wholesale/Retail	Marketing Manager/Supervisor	High	Bachelor	Marketing	(13, 22)
	Real Estate Development	Legal manager/Supervisor	High	Bachelor	Law	(14, 24)

Table A1.3, continued

F	Internet/E-Business	Legal manager/Supervisor	High	Bachelor	Law	(14, 24)
	Internet/E-Business	Human Resources Specialist/Assistant	Entry	College	Econ&Management	(8, 12)
	Real Estate Development	Human Resources Specialist/Assistant	Entry	College	Econ&Management	(8, 12)
	Internet/E-Business	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(13, 19)
	Real Estate Development	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(13, 19)
	Internet/E-Business	Human Resources Director	High	Bachelor	Econ&Management	(16, 25)
	Real Estate Development	Human Resources Director	High	Bachelor	Econ&Management	(16, 25)
	Internet/E-Business	Accountant	Middle	Bachelor	Accounting	(12, 18)
	Securities/Investment Funds	Financial Manager	High	Bachelor	Finance	(14,19)
	Internet/E-Business	Administration Specialist/Assistant	Entry	College	Econ&Management	(8, 12)
	Real Estate Development	Executive Assistant/Secretary	Entry	College	Econ&Management	(10, 13)
	Internet/E-Business	Administration Manager/Supervisor	Middle	Bachelor	Econ&Management	(14,19)
	Internet/E-Business	Administration Vice President	High	Bachelor	Econ&Management	(50, 60)

Notes:

1. Gender (M, N, F) represents the job type's dominant gender.
2. Current wages (·,·) refer to the resume's current wages for (young, older) workers in 10,000 RMB, respectively.

Table A1.4: Selected Job Types in Job Board 4

Gender	Occupation	Skill Level	Education Level	Major	Current Wages
M	WEB Front-end Developer	Middle	Bachelor	Computer Science	(19, 25)
	Operation and Maintenance Engineer	Middle	Bachelor	Computer Science	(18, 24)
	Operation and Maintenance Director	High	Bachelor	Computer Science	(19, 26)
	Pattern Recognition	Middle	Bachelor	Computer Science	(19, 25)
	Machine Learning	Middle	Bachelor	Computer Science	(19, 25)
	Operations Assistant	Entry	College	Computer Science	(7, 10)
	Operations Specialist	Middle	College	Computer Science	(11, 12)
	Operations Manager/Supervisor	High	Bachelor	Computer Science	(14, 19)
	Test Engineer	Middle	Bachelor	Computer Science	(16, 23)
	Test Manager	High	Bachelor	Computer Science	(19, 25)
	Full Stack Engineer	Middle	Bachelor	Computer Science	(17, 25)
N	Sales Representative	Entry	College	Marketing	(8, 12)
	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 17)
	Sales Director	High	Bachelor	Marketing	(18, 25)
	Product Assistant	Entry	College	Econ&Management	(10, 11)
	Product Manager	High	Bachelor	Econ&Management	(16, 23)
	Project Assistant	Entry	College	Econ&Management	(10, 11)
	Project Manager	High	Bachelor	Econ&Management	(16, 23)
	Data Analyst	Middle	Bachelor	Statistics	(13, 19)
	Design Assistant	Entry	College	Arts	(8, 11)
	Designer	Middle	College	Arts	(13, 19)
	Design Manager	High	Bachelor	Arts	(16, 23)
	Strategy Consultant	Middle	Bachelor	Econ&Management	(13, 19)

Table A1.4, continued

F	Human Resources Specialist/Assistant	Entry	College	Econ&Management	(10, 11)
	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Human Resources Director	High	Bachelor	Econ&Management	(17, 26)
	Accountant	Middle	Bachelor	Accounting	(13, 17)
	Training Specialist	Entry	College	Econ&Management	(10, 12)
	Customer Service	Entry	College	Marketing	(7, 8)
	Customer Service Manager	High	College	Marketing	(13, 17)
	Media Specialist	Entry	College	Marketing	(7, 8)
	Media Manager	High	Bachelor	Marketing	(11, 16)
	Administration Specialist/Assistant	Entry	College	Econ&Management	(10, 12)
	Administration Manager/Supervisor	Middle	Bachelor	Econ&Management	(13, 18)
	Administration Director	High	Bachelor	Econ&Management	(16, 25)

Notes:

1. Gender (M, N, F) represents the job type's dominant gender.
2. The industry in job board 4 is set as "all industries".
3. Current wages (·,·) refer to the resume's current wages for (young, older) workers in 10,000 RMB, respectively.

A2: Fictitious Resumes

For each of the job types listed in Tables A1.1-A2.4, we scraped 50 job ads and 50 resumes as the information pool for our fictitious profiles. As noted in Section 2.3, for each job type on each board we created four profiles: a younger pair of identical male and female workers and an older pair. In all cases our profiles provided only the basic information required by each job board to register as a valid job seeker; this information falls into four categories on all the boards: *Personal information* includes the worker's name, birth date, years of working experience, current wage, city, employment status, phone number, and email address. *Education* includes the highest education level, years attended, university name and major; *experience* includes the current company name, occupation, industry, job title, and job description. Finally, workers are asked about their *job search goals*, including the desired wage, location, industry, and occupation. In the rest of this section, we detail how these main components of our resumes were generated. A summary of these design choices is provided in Table A2.1.

A2.1 Personal Information

Name: The name pool for our applicants is the 20 most common last names, the top 15 male first names and top 15 female first names based on statistics from 2015 Chinese Census 1% Population Sample, as listed below:

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).

For each applicant, a last name and plus a first name corresponding to the applicant's gender were randomly drawn from the name pool.

Birth Date: Instead of varying workers' ages directly, we varied their graduation year, which was 2017 for our young gender pairs and 2007 for our older pairs. Once a worker's graduation year is fixed, their age is jointly determined by their graduation year and education level. For example, our 'young' workers with (three-year) college degrees are age 25 (born in 1995), while young workers with a bachelor's degree are 26. Our older workers are 35 or 36 years old depending on their education level. Fixing graduation years in this way has the advantage of equalizing work experience levels between our more- and less-educated applicants.

Years of Work Experience: To simplify the profiles, all our workers started to work just after they graduated from the university/college of their highest degree. Therefore, our young profiles all have 3 years of experience and our older profiles have 13 years.

Current Wage: To generate realistic current wages for our profiles, we created hiring-agent profiles on each of our four job boards in March 2020, and –for each of our 35 job types-- searched for workers that were currently working in those jobs. Our search criteria specified "1 to 3 years" of working experience for our young profiles, and "5 to 10 years" for our older profiles. For both experience levels in every job type, we recorded the current wages of the first 50 workers who appeared in our search results and used the mean of these wages as our fictitious worker's wage.

City: All the four job boards are nationally recognized and serve most of the regions in China, but over half of their job postings are in China's four first-tier cities: Beijing, Shanghai, Shenzhen, and Guangzhou. To ensure a sufficient sample of job recommendations for our experiment, we located all our fictitious workers in those four cities.

Employment Status: All our fictitious workers are currently employed.

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

A2.2 Education

As noted, our workers had either three or four years of post-secondary education ("college" or "bachelor's degree"). To assign these levels we created worker accounts on all four job boards in February 2020 and used them to look at job postings in all 35 of each board's job types. For each job type, we collected the first 50 job advertisements in February 2020. We then used the modal education requirement in this sample (which was always either college or bachelors) as the education level for all profiles in that job type. Start- and end dates of higher education were then assigned accordingly to our young and old profiles.

The two workers in a gender pair have the same educational background and attended the same college or university. The institution's name was randomly drawn from the Chinese Ministry of Education's *2019 Higher Education Institution List*, restricting attention to the provinces surrounding the worker's current location.[1] Workers' majors also match the job type. For example, Computer Science/Software is assigned for IT jobs, Mathematics/Statistics for data positions, and economics/ management/marketing/ majors for other jobs.

A2.3 Recent Job History

Since all our workers are currently employed, their recent jobs are their current jobs. For young workers, their current jobs started in August in their graduation year (2017); for old workers, their current jobs started five years ago (in March 2015), so they have 5 years of tenure in their current / recent position. Our workers' current occupation and industry is that of their current job type, which is also the job type they are seeking. The job titles in workers' resumes are simply the occupation of their job type.

We made up current employer names to minimize the disturbance to real firms and workers. All company names consist of three parts, beginning with the company's location, which is equal to worker's current city. Next, we used an online business name generator to create the 100 company names listed below.

[1], and we excluded the provinces that have ethnic minority groups, such as Xinjiang, Yunnan, Qinghai, Tibet and Guangxi.

Company Names

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

After randomly assigning company names to all gender pairs, the final part of the company name is equal to the industry associated with the worker's job type. Thus, for example, a typical company name would be Beijing Dongya Internet Technology Company.

Workers' current occupation and industry are given by their job type; the worker's job title is also equal to the occupation associated with their job type.

A2.4 Applicants' Job Search Goals

All our workers are looking for full-time jobs and list a desired wage equal to 120% of their current wage. Their desired city, industry, and occupation are the same as their current ones.

Table A2.1: Resume Information Generation

	Method	Notes
Personal Information		
Name	Randomly assigned to each worker	Appendix A2.1
Birth Date	Young workers graduated in 2017, and older workers 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 resumes we collected from the platforms.	Assigned based on job type and worker experience.
City	Beijing, Shanghai, Shenzhen, Guangzhou	
Employment Status	Currently employed.	
Phone Number & Email	Uniquely assigned for each worker.	
Education		
Highest degree	Bachelor's degree or junior college, depending on the 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.

Table A2.1, continued

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 pair of workers.	Appendix A2.3
Occupation	Matches job type	
Industry	Matches job type	
Job Title	Matches occupation	
Job Description	Matches occupation	
Goals		
Desired Wage	Current wage*1.2	
Desired City	Matches worker’s city	
Desired Industry	Matches job type	
Desired Occupation	Matches job type	

Appendix B: Descriptive Statistics

Table B1: Descriptive Statistics, Applicant Sample

	All Job Boards	Job Board 1	Job Board 2	Job Board 3	Job Board 4
Current Wage (RMB/year)	142,507	100,571	132,086	177,943	159,429
	(65,142)	(44,276)	(45,685)	(80,082)	(55,863)
Desired Wage (RMB/year)	179,732	133,500	163,114	221,057	201,257
	(81,819)	(69,294)	(57,146)	(98,184)	(67,178)
Education (years)	15.5643	15.3143	15.5143	15.8000	15.6286
	(0.496)	(0.465)	(0.500)	(0.400)	(0.484)
Sample Size	2,240	560	560	560	560

Notes:

1. Education levels in the resumes have been converted to years, as follows: College degree = 15 years; bachelor’s degree = 16 years.
2. Standard deviations are in parentheses.

Table B2: Descriptive Statistics, Recommended Job Sample

A. By Job Board:

	All Job Boards	Job Board 1	Job Board 2	Job Board 3	Job Board 4
Posted Wage?	0.9588	0.9488	0.9752	0.9365	0.9775
	(0.199)	(0.220)	(0.156)	(0.244)	(0.148)
Wage, if posted (RMB/year)	205,928	148,422	174,320	250,993	239,423
	(128,428)	(88,748)	(89,832)	(143,656)	(141,515)
Required Education (years)	15.4219	14.8113	15.3321	15.7668	15.6676
	(1.121)	(1.608)	(0.814)	(0.614)	(0.952)
Required Experience (years)	2.4395	2.0741	2.0929	3.1624	2.3341
	(2.106)	(1.841)	(2.030)	(2.341)	(1.970)
Firm Size (⩾1000)	0.3487	0.2297	0.3102	0.4071	0.4331
	(0.477)	(0.425)	(0.466)	(0.494)	(0.496)
Sample Size	81,231	20,615	16,981	22,078	21,557

Notes:

1. Observations are 81,231 unique job advertisements collected in the experiment.
2. Wage is the midpoint of the posted wage range.
3. Education levels in job ads are converted to years of education as follows: middle school = 9; tech or high school = 12; college = 15; bachelor's degree = 16; masters/MBA = 18; doctoral degree or equivalent = 23 years.
4. Company size is self-reported by hiring agents.
5. Standard deviations are in parentheses.

B. By Recommendation:

	(1) All Recommended Jobs	(2) Common Jobs	(3) Recommended to Women Only	(4) Recommended to Men Only
Posted Wage?	0.9592	0.9558	0.9703	0.9688
	(0.198)	(0.206)	(0.170)	(0.174)
Wage, if posted (RMB/year)	205,704	206,162	202,581	206,050
	(127,774)	(131,146)	(117,696)	(116,423)
Required Education (years)	15.4234	15.4082	15.4626	15.4755
	(1.115)	(1.147)	(1.018)	(1.008)
Required Experience (years)	2.4373	2.4527	2.3092	2.4652
	(2.103)	(2.119)	(2.029)	(2.072)
Firm Size (⩾1000)	0.3493	0.3427	0.3552	0.3833
	(0.477)	(0.475)	(0.479)	(0.486)
Sample Size	83,793	63,224	9,808	10,899

Notes:

1. Common Jobs are recommended to both the Male and Female profile in a pair.
2. The male only and female only jobs are defined at the pair level. Observation counts in columns 2-4 sum to more than column 1 because a job recommended only to men (women) in one gender pair could be recommended only to women (men) in another.

Table B3: Matching Rates between Recommended Jobs and Workers' Characteristics

	All Job Boards	Job Board 1	Job Board 2	Job Board 3	Job Board 4
Desired Wage Match	83.86%	94.59%	87.69%	77.66%	75.72%
Education Match	88.46%	87.66%	98.23%	87.90%	80.14%
Experience Match	92.25%	94.33%	91.80%	85.37%	97.48%
Location Match	97.10%	95.83%	97.66%	95.01%	99.92%
Sample Size	177,320	44,800	43,880	44,320	44,320

Notes:

1. Table B3 summarizes the extent to which the recommended jobs match the worker's characteristics. The sample is 177,320 job recommendations received by 2,240 fictitious applicants.
2. Desired wage match equals 1 if the upper bound of the recommended job's posted wage range exceeds the lower bound of the worker's desired wage.
3. Education (experience) match equals 1 if the job's requirement is less than or equal to the worker's qualifications.
4. Location match equals 1 if the job's city is consistent with the worker's city.
5. The recorded number of job recommendations is slightly smaller than the designed number 2,240*80 = 179,200 for at least two reasons. One is that job boards occasionally froze the fictitious worker accounts we created; in these cases we terminated the experiment for the whole gender pair if one member was blocked. Second, some recommended job links were blank, so we could not scrape their characteristics. Overall, fewer than 0.5 percent of recommendations are missing; missing data appear to occur randomly, and independently of the gender of fictitious applicants.

Table B4: Difference Rate in Job Recommendations

	Difference Rate (S.D.)	Between-Group Differences
All Recommendations:	0.1240 (0.037)	
Worker Age:		
Young	0.1262 (0.037)	0
Old	0.1219 (0.036)	-0.0043*
Job Gender Type:		
Female-dominated	0.1178 (0.038)	-0.0111***
Gender Neutral	0.1289 (0.033)	0
Male-dominated	0.1254 (0.039)	-0.0035
Job Skill Level:		
Entry	0.1151 (0.036)	0
Middle	0.1330 (0.038)	0.0179***
High	0.1250 (0.034)	0.0100***
Job Location:		
Beijing	0.1238 (0.038)	0
Shanghai	0.1240 (0.036)	0.0002
Shenzhen	0.1244 (0.035)	0.0006
Guangzhou	0.1239 (0.037)	0.0001

Notes:

1. Statistics are for all four job boards combined.
2. The difference rate equals the number of gender-specific recommendations divided by the number of total recommendations received by both male and female applicants in the gender pair.
3. Between-Group Differences are relative to the indicated omitted category for each characteristic; significance levels are from t-tests. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
4. Duplicate job recommendations from different rounds are counted only once.

Appendix C: Learning from Words

C1: Multiple Hypothesis Testing

In the paper we leverage the Romano-Wolf (2005) and Anderson (2008) procedures to account for multiple hypothesis testing for over-representation in our list of 172 most common words (listed in Table 2). Specifically, we defined a word as over-represented in jobs shown to men or women if both its Romano-Wolf p value and its Anderson q value were below 0.05. The resulting list of over-represented words is provided below.

Table C1: Multiple Hypothesis Testing on Over-Represented Words

Female Words (36 words)	RW p-value	Anderson q-value	Male Words (22 words)	RW p-value	Anderson q-value
Listen	0.001	0.001	DecisionMaking	0.005	0.001
Speak	0.001	0.001	Planning	0.001	0.001
Writing	0.001	0.001	Engineering	0.001	0.001
Documentation	0.001	0.001	Leadership	0.001	0.001
Data	0.001	0.001	Charge	0.016	0.001
ChatTools	0.001	0.001	Supervise	0.001	0.001
Cooperation	0.001	0.001	NightWork	0.047	0.002
Communication	0.001	0.001	Overtime	0.005	0.001
Assist	0.001	0.001	LongTravel	0.008	0.001
Negotiation	0.001	0.001	Commission	0.005	0.001
Administrative	0.001	0.001	Stock	0.001	0.001
Collect	0.001	0.001	Allowance	0.001	0.001
EightHour	0.001	0.001	Reward	0.042	0.002
Flexible	0.001	0.001	Meal	0.001	0.001
WeeklyBreak	0.001	0.001	Shuttle	0.001	0.001
RegularHour	0.001	0.001	Commute	0.001	0.001
MarriageLeave	0.001	0.001	InjuryIns	0.028	0.001
ParentalLeave	0.001	0.001	Public	0.002	0.001
MaternityLeave	0.001	0.001	ScienceEngineering	0.005	0.001
MedicalIns	0.001	0.001	NoCrime	0.001	0.001
SocialSecurity	0.001	0.001	Entrepreneurial	0.033	0.001
MaternityIns	0.001	0.001	Pressure	0.001	0.001

Training	0.001	0.001	
Atmosphere	0.001	0.001	
Certificate	0.005	0.001	
NewGrad	0.001	0.001	
NonExperience	0.046	0.002	
Careful	0.001	0.001	
Patient	0.001	0.001	
Active	0.001	0.001	
Outgoing	0.001	0.001	
Generous	0.001	0.001	
Punctual	0.001	0.001	
Figure	0.001	0.001	
Temperament	0.001	0.001	
Facial	0.001	0.001	

Note: The Romano-Wolf p-value is calculated through 1000 bootstrap replications.

C2: Most Frequent Words in Job Ads

Figure C1: Word Cloud in Chinese

Note: The figure presents a word cloud generated from the original Chinese job advertisements. It includes English terms that were present in the Chinese ads; for example, "word" refers to Microsoft Word.

Table C2: Word List in Job Ads

The word cloud in Figure 2 shows the 172 most common words in recommended job ads, with larger type indicating more common words. In Table C2, we present this list a different way: organized into six (author-defined) categories and sub-categories.

Standardized PIACC Skills (47 words)	**Literacy:** listening, speaking, reading, writing, language, documentation **Numeracy:** data, accounting, analysis **ICT skills:** programing, Microsoft Office, chat tools **Problem-solving:** learning, comprehension, thinking, logic, decision-making, planning, problem-solving, engineering, independent, insight **Influencing:** leadership, team management, charge, supervise **Cooperation:** cooperation, communication, teamwork, assisting, coordination, organizing, negotiation, public relations, marketing, advertising, sales, client, compliance **Self-organization:** administrative, designing, collecting, reception, driving, execution, testing, task management
Benefits (35 words)	**Compensation:** base pay, commission, stock, allowance, promotion, reward **Leave and Vacation:** vacation, marriage leave, parental leave, maternity leave, sick leave, funeral leave, holiday **Facilities and Transportation:** office supplements, vehicle, meal, housing, shuttle, subway, commute friendly, snacks **Insurance**: Fiveone[3], medical insurance, commercial insurance, social security, housing funds, maternity insurance, unemployment insurance, endowment insurance, injury insurance, disease insurance **Other benefits:** training, staffing, activities, mentor
Work Timing and Location (17 words)	**Schedule:** work shift, night work, morning work, evening work, big and small weeks[4], eight-hour, flexible, attendance, overtime, no overtime **Business travel:** regular travel, short travel, long travel **Breaks:** weekly break, monthly break, noon break, regular working hours

Table C2, continued:

Company and Rank (16 words)	**Rank:** senior, medium, core **Culture:** atmosphere, employee care, career, dream, culture, screening **Company Type:** direct recruiting, public company, top500, startup, flat management, financing, big company
Other Qualifications (16 words)	**Education:** non education, certificate, new grad, Tongzhao[5], tier-one school, fulltime school, top school, nonmajor, major, science&engineering **Experience:** no experience required, experienced, overseas **Other:** no crime history, law abiding, solitary
Personality, Age, and Appearance (41 words)	**Personality:** effective, methodical, 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:** no gender restriction, no age restriction, age below 35, age below 40 **Appearance:** figure, temperament, healthy, facial, clothing, shape, voice

Notes:

1. This table shows the 172 most common words in recommended job ads, according to the authors' categorizations (see Section 3.3 for details).
2. Every listed word includes its variations, such as leadership vs leading, and confidence vs confident.
3. "Fiveone" represents "five social insurance plans plus one housing fund" (五险一金), including endowment insurance, medical insurance, unemployment insurance, employment injury insurance, maternity insurance, and housing fund.
4. "Big and small weeks" describes working schedules in which workers have one rest day in one week and two rest days rest in the next week.
5. "Tongzhao" means that admission to the applicant's university or college requires taking the Gaokao in high school.

C3: Stereotypically Male and Female Words, Source 1: Literature

The following words (from our list of 172 most-common words in ads) were identified as significantly gendered in the following sources: Gaucher et al. (2011); Kuhn et al. (2020); Chaturvedi et al. (2021).

Table C3: Gendered Words from Literature

Female Words	Male Words
speak, documentation, Microsoft Office, cooperation, communication, assist, coordination, administrative, reception, housing, careful, patient, trustworthy, honest, kind, responsive, temperament, facial	read, data, analysis, learning, logic, decision making, problem solving, engineering, independent, leadership, charge, supervise, negotiation, client, driving, work shift, night work, overtime, regular travel, training, law abiding, solitary, energetic, active, self-motivated, hardworking, tenacious, courageous, innovative, punctual, entrepreneurial, devotion, enthusiasm, pressure

C4: Stereotypically Male and Female Words, Source 2: MTurk Survey

To determine the gendered perceptions of words, we 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". This question was asked for each of the 172 words listed in Table C2.

In total, 86 valid surveys were collected from people between the ages of 25 to 55, and 56% of them were men. The table below lists words classified as stereotypically female or male, defined as those significantly associated with seeking women or men, respectively, at a significance level of 5 percent or greater.

Table C4: Gendered Words from Amazon MTurk Survey

Female Words	Male Words
read, write, documentation, learning, assist, compliance, administrative, design, reception, marriage leave, parental leave, maternity leave, sick leave, holiday, maternity insurance, careful, patient, enthusiasm, kind, 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, stock, promotion, vehicle, mentor, startup, science&engineering, experienced, no crime history, effective, practical, responsible, pressure

C5: Stereotypically Male and Female Words, Source 3: Chinese Worker Survey

The Chinese version of our survey on people's perceptions about gendered words in job ads was conducted in Wenjuanxing (问卷星) in September 2021. The survey question is the same as our MTurk survey, 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。This question was asked for each of the 172 words listed in Table C2.

79 valid respondents participated in the survey, 81% of them were between 25 to 55 years old and 73% of them were men. The table below lists words classified as stereotypically female or male, defined as those significantly associated with seeking women or men, respectively, at a significance level of 5 percent or greater.

Table C5: Gendered Words from Chinese Survey

Female Words	Male Words
speak, read, communication, assist, compliance, administrative, design, collect, reception, eight-hour, flexible, marriage leave, parental leave, sick leave, office supplements, 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

C6: Stereotypically Male and Female Words, Source 4: ChatGPT

We tasked ChatGPT version 4.0 with identifying and categorizing the words in Table D1 according to their gender associations. Our prompt was: "We are interested in investigating gendered words in the labor market. Can you categorize each word in the following six categories as neutral, male, or female?"

ChatGPT's response was, "When classifying words in job postings as gender-neutral, male-associated, or female-associated, it's important to note that these classifications are rooted in historical biases and stereotypes that are increasingly being challenged and dismantled in modern workplaces." (ChatGPT 2024). Following this, ChatGPT classified the following words as female- and male-associated.

Table C6: Gendered Words from ChatGPT

Female Words	Male Words
assist, coordination, public relation, reception, vacation, marriage leave, parental leave, maternity leave, sick leave, maternity insurance, unemployment insurance, employee care, patient, kind, figure, temperament, facial, clothing, shape, voice	engineering, leadership, team management, charge, supervise, driving, nightwork, morning work, evening work, overtime, regular travel, long travel, commission, stock, core, full time school, top school, courageous, entrepreneurial

Reference:

OpenAI (2023). ChatGPT (GPT-4, March 14 Version) [Large language model]. Response to query made on 11/05/2023. https://chat.openai.com/chat

C7: Gender Gaps in Stereotypical Ad Content using Different External Word Lists

The stereotype content index in the paper combines data from the four external sources described in Sections C3-C6. To check if this procedure is robust to the sources used, Table C7 replicates Table 4 using stereotype indicators derived from each of the four sources individually.

Table C7: Gender Differences in the Stereotypical Content of Job Ads

Source of Word List	(1) Index of Stereotypically Female Content (S^f) (standardized)	(2) Index of Stereotypically Male Content (S^m) (standardized)
A. Literature		
Male	-0.3443***	0.0581***
	(0.012)	(0.013)
B. MTurk Survey		
Male	-0.5635***	0.1568***
	(0.013)	(0.013)
C. Chinese Survey		
Male	-0.5316***	0.0935***
	(0.013)	(0.013)
D. ChatGPT		
Male	-0.6644***	0.2012***
	(0.014)	(0.014)
N	22,023	22,023

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. For each source, a word was assigned a stereotype score (s^f or s^m) of one if the source classified it as gender-stereotypical or not. As in equation 3, job ads' stereotype scores (S^f or S^m) summed these scores over the words in each ad. For the regressions in Table C7, we then standardized these scores to have a mean of zero and standard deviation of one.
2. The unstandardized means of S^f and S^m were 1.96 and 2.07 in row A, 1.24 and 2.02 in row B, 2.04 and 2.90 in row C, and 0.56 and 1.37 in row D.

Appendix D: Main Results, Separately by Job Boards

To simplify the presentation, all the main results in our paper are based on combined data from all four job boards we audited. Since different job boards may use different job recommendation algorithms, it is important to replicate our main results for each job board separately. Here, we replicate the following results from the paper, by job board:

- Tables 1 and 4, which show gender gaps in the observed characteristics of jobs recommended to men versus women, including the amount of stereotypically male or female content the job ads contain.
- Figures 3 and 4, which show the evolution across experimental rounds in the difference rate, and the evolution of gender gaps in job characteristics (wages, education and experience requirements, and firm size) and in stereotypically male and female job ad content.

Overall, while the levels of differentiation between male and female recommendations vary substantially across the boards --for example, the set difference rates (see below) are 8.07%, 11.56%, 14.31%, and 15.68%-- the following patterns are observed on all four job boards:

- Jobs recommended to men pay more than jobs recommended to women (insignificant in one of four cases).
- Jobs recommended to men request more experience.
- Jobs recommended to men are in larger firms (insignificant in two of four cases).
- Jobs recommended to men contain much less stereotypically female content.
- Jobs recommended to men contain more stereotypically male content.
- All four job boards show increasing difference rates across experimental rounds. In three of four cases, the increase between Rounds 0 and 1.1 is substantially larger than between all other adjacent rounds.
- None of the job boards show trends in types of jobs recommended to men versus women with respect to the following characteristics: posted wage, education requirement, experience requirement, and stereotypically male content. Men's firm size advantage declines across round on one board (Board 4). Consistent with the aggregate result in Figure 4, the gender gap in stereotypically female content increases on two boards (3 and 4).

D1: Job Board 1

Table D1: Gender Differences in Characteristics of Job Recommendations in Job Board 1

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	2,236*	0.0680*	0.1732***	-0.0050	-0.6323***	0.1204***
	(1,327)	(0.039)	(0.048)	(0.013)	(0.034)	(0.032)
N	3,452	3,041	3,598	3,615	3,615	3,615
R^2	0.782	0.551	0.467	0.132	0.373	0.164

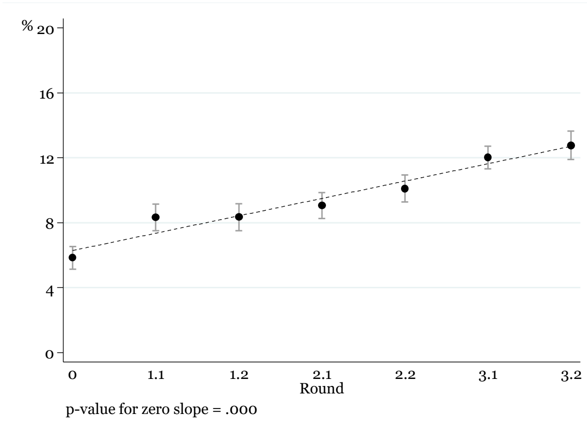
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

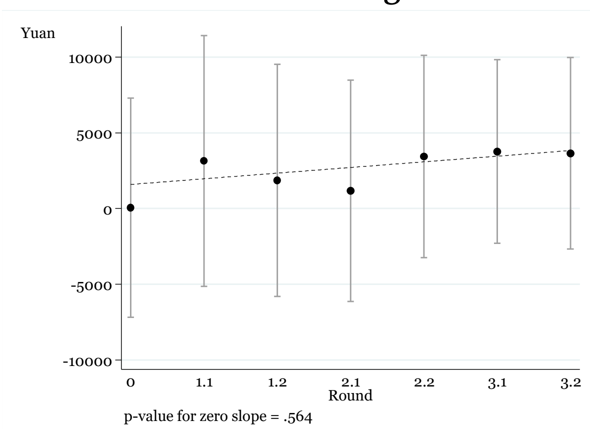
1. This Table replicates Tables 1 and 4 using data from job board 1 only.
2. On job board 1, we collected 44,800 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 8.07%, with 3,615 jobs being exclusively recommended to either male or female applicants only.

Figure D1: Gender Differences in Characteristics of Job Recommendations in Job Board 1, by Rounds

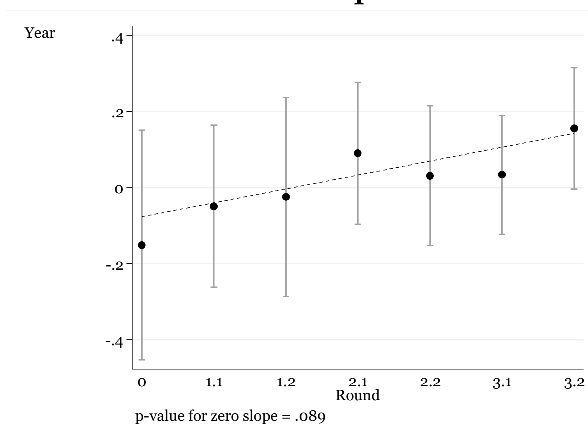
(a) Set Difference Rate



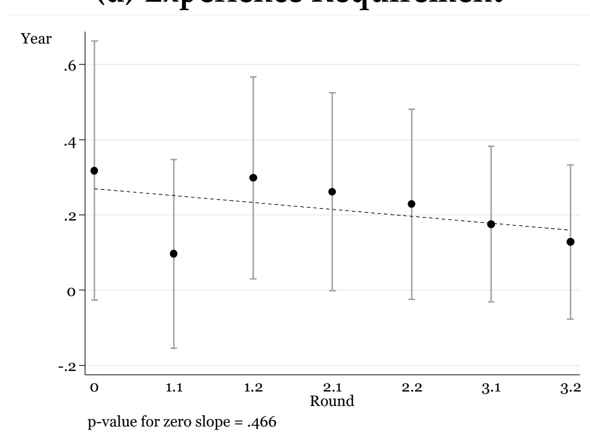
(b) Posted Wage



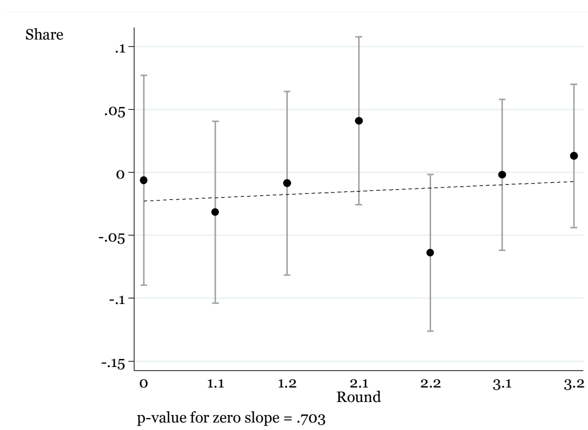
(c) Education Requirement



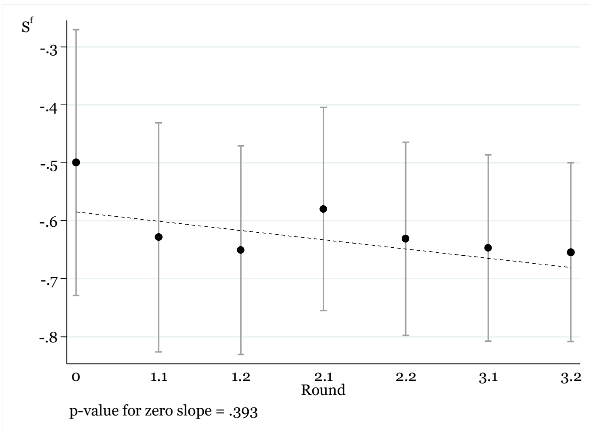
(d) Experience Requirement



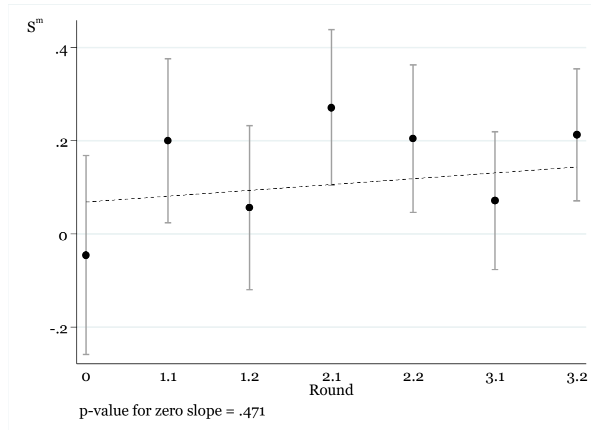
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D1 replicates Figures 3 and 4 using job recommendations from Job Board 1, showing the difference rate, gender differences in posted wage, education, working in experience, female content and male content by each round, respectively.
2. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
3. All panels show regression lines of the gender gap in the outcome on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively). Sample sizes vary between 1,148 and 1,960, depending on the outcome. p-values for a slope of zero are shown below each panel.

D2: Job Board 2

Table D2: Gender Differences in Characteristics of Job Recommendations in Job Board 2

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	2,492	0.0299	0.1607***	0.0355***	-0.4906***	-0.0954***
	(1,530)	(0.019)	(0.044)	(0.013)	(0.026)	(0.026)
N	4,960	4,443	5,071	5,087	5,087	5,087
R^2	0.619	0.426	0.425	0.115	0.299	0.162

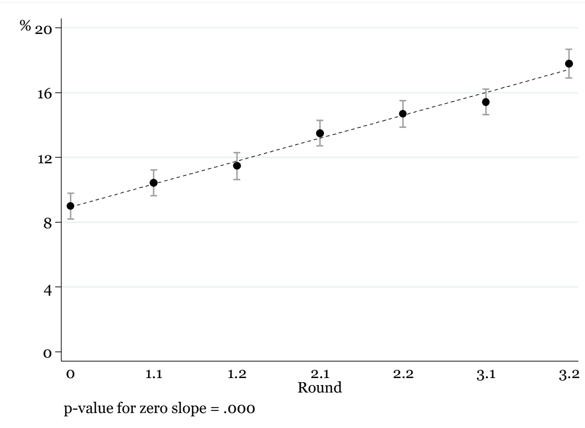
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

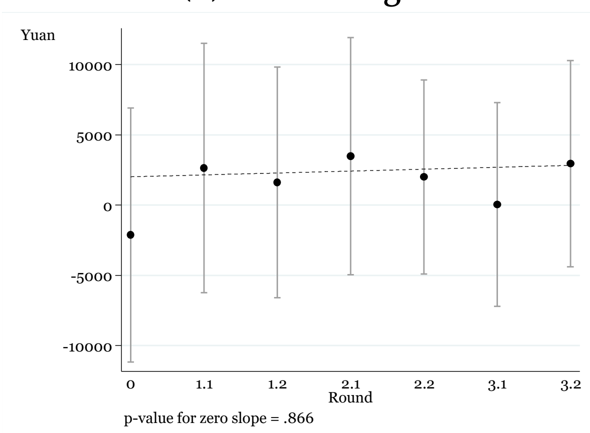
1. This Table replicates Tables 1 and 4 using data from job board 2 only.
2. On job board 2, we collected 43,800 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 11.56%, with 5,087 jobs being exclusively recommended to either male or female applicants only.

Figure D2: Gender Differences in Characteristics of Job Recommendations in Job Board 2, by Rounds

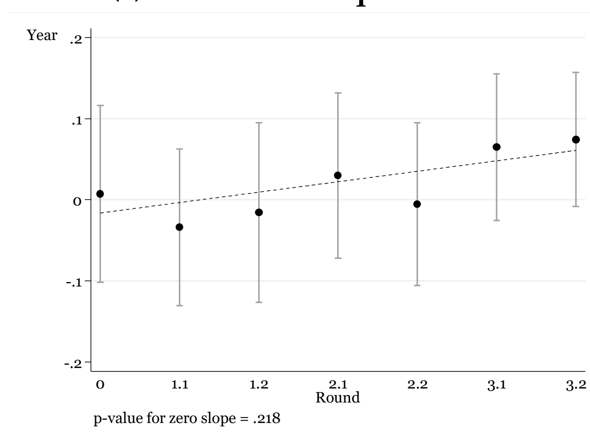
(a) Set Difference Rate



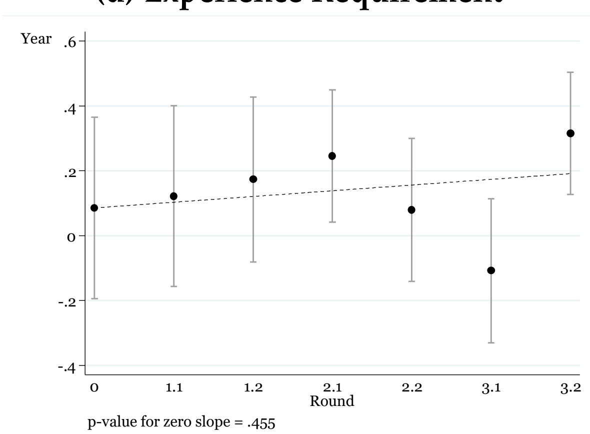
(b) Posted Wage



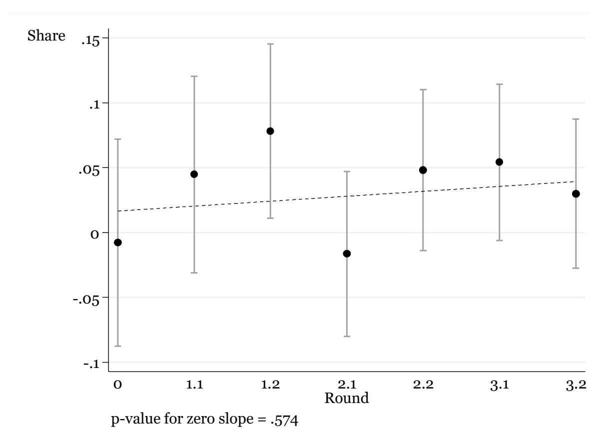
(c) Education Requirement



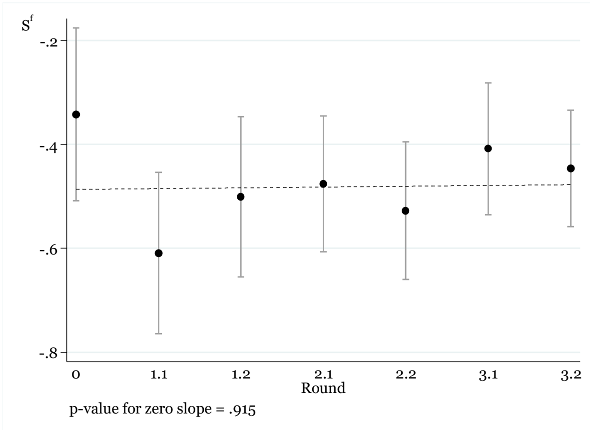
(d) Experience Requirement



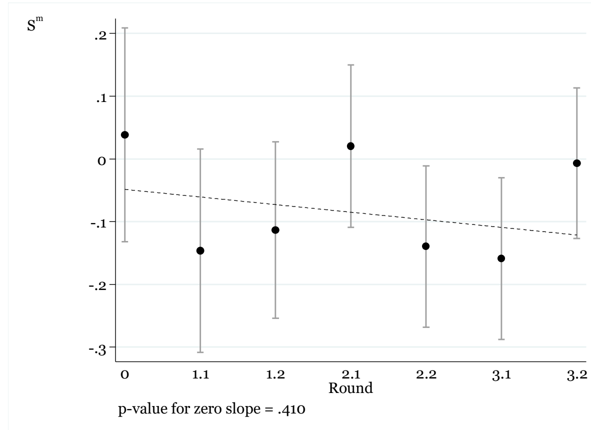
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D2 replicates Figures 3 and 4 using job recommendations from Job Board 2, showing the difference rate, gender differences in posted wage, education, working in experience, female content and male content by each round, respectively.
2. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
3. All panels show regression lines of the gender gap in the outcome on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively). Sample sizes vary between 1,469 and 1,914, depending on the outcome. p-values for a slope of zero are shown below each panel.

D3: Job Board 3

Table D3: Gender Differences in Characteristics of Job Recommendations in Job Board 3

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	3,412*	0.0295**	0.1200**	0.0670***	-0.5036***	0.2630***
	(2,026)	(0.014)	(0.051)	(0.012)	(0.022)	(0.025)
N	5,915	5,949	6,279	6,347	6,347	6,347
R^2	0.584	0.182	0.246	0.080	0.276	0.102

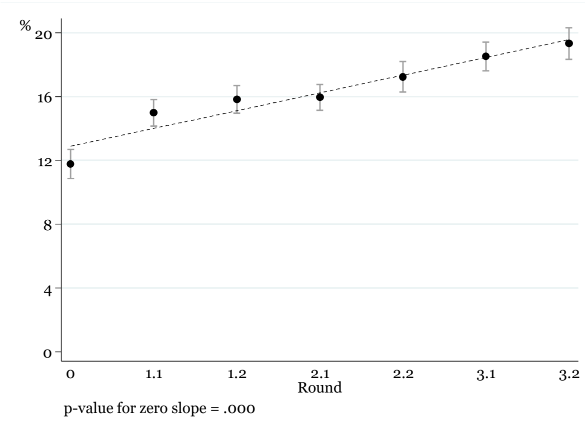
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

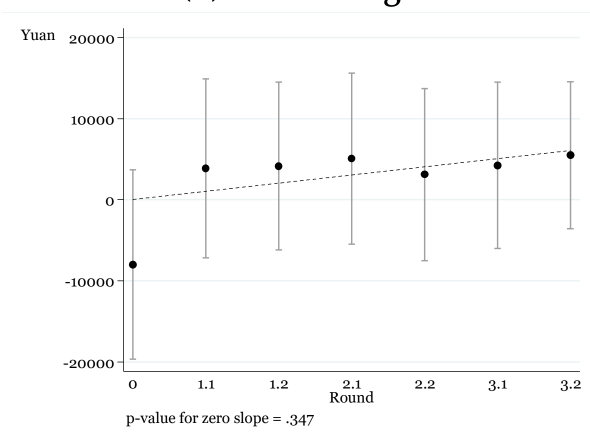
1. This Table replicates Tables 1 and 4 using data from job board 3 only.
2. On job board 3, we collected 44,320 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 14.31%, with 6,347 jobs being exclusively recommended to either male or female applicants only.

Figure D3: Gender Differences in Characteristics of Job Recommendations in Job Board 3, by Rounds

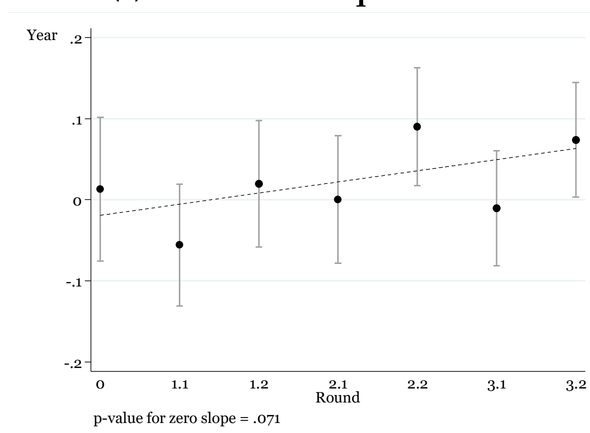
(a) Set Difference Rate



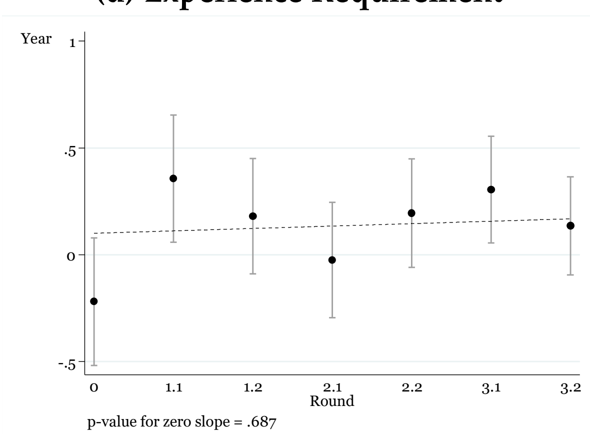
(b) Posted Wage



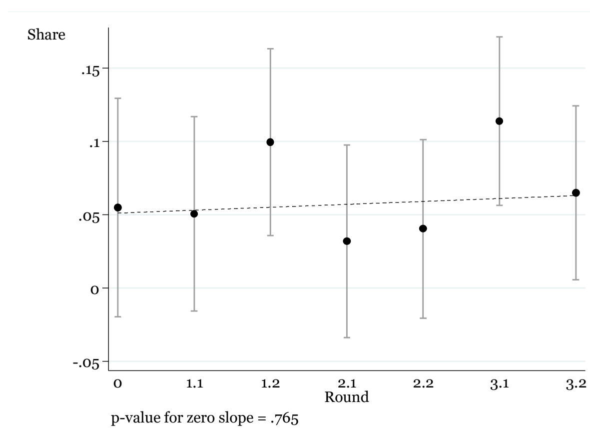
(c) Education Requirement



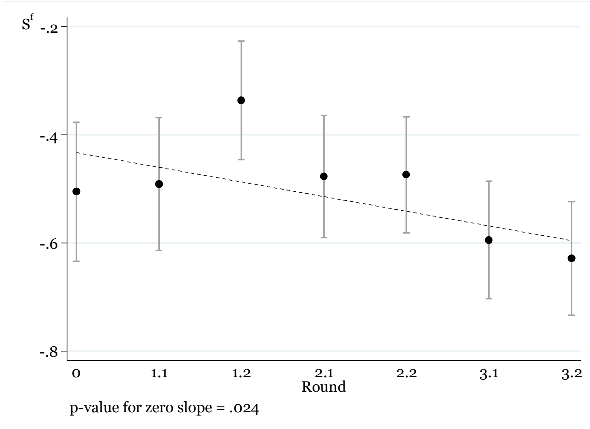
(d) Experience Requirement



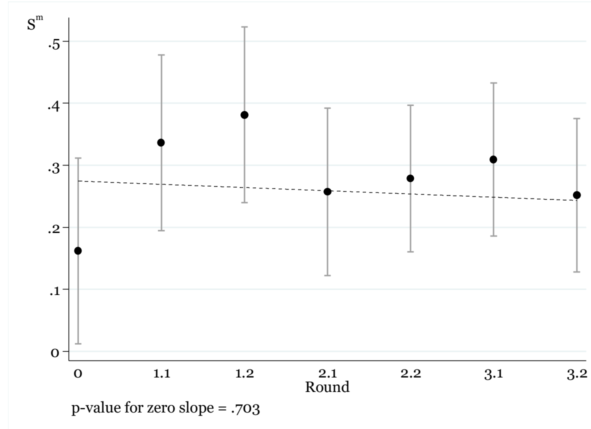
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D3 replicates Figures 3 and 4 using job recommendations from Job Board 3, showing the difference rate, gender differences in posted wage, education, working in experience, female content and male content by each round, respectively.
2. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
3. All panels show regression lines of the gender gap in the outcome on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively). Sample sizes vary between 1,757 and 1,936, depending on the outcome. p-values for a slope of zero are shown below each panel.

D4: Job Board 4

Table D4: Gender Differences in Characteristics of Job Recommendations in Job Board 4

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	3,751*	-0.0254	0.2062***	0.0029	-0.6748***	0.1855***
	(2,267)	(0.022)	(0.033)	(0.011)	(0.022)	(0.024)
N	6,935	6,467	6,974	6,974	6,974	6,974
R^2	0.521	0.285	0.381	0.211	0.257	0.083

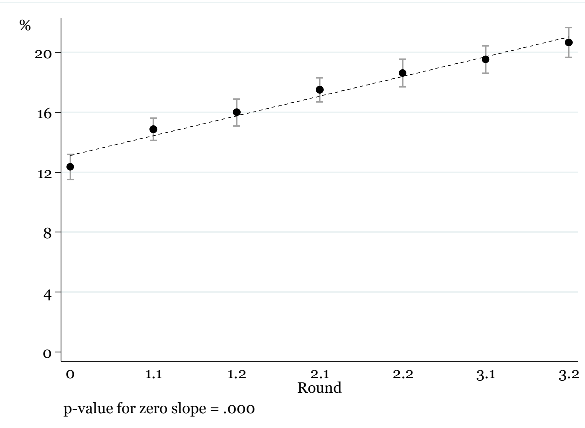
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

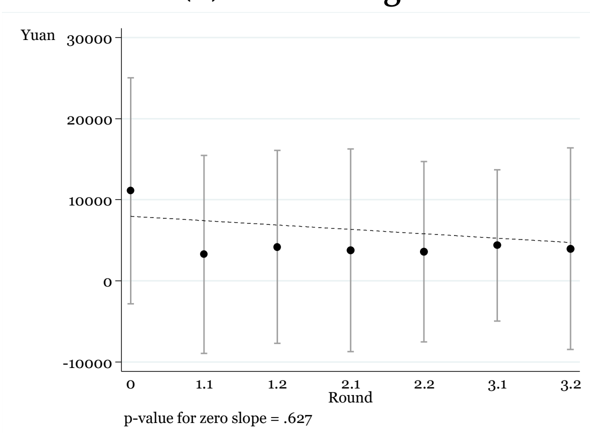
1. This Table replicates Tables 1 and 4 using data from job board 4 only.
2. On job board 4, we collected 44,320 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 15.68%, with 6,974 jobs being exclusively recommended to either male or female applicants only.

Figure D4: Gender Differences in Characteristics of Job Recommendations in Job Board 4, by Rounds

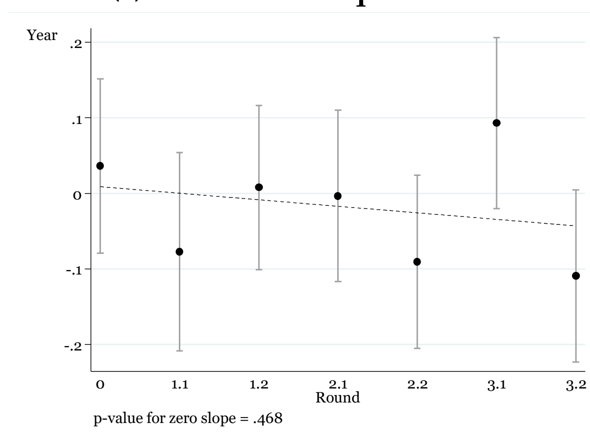
(a) Set Difference Rate



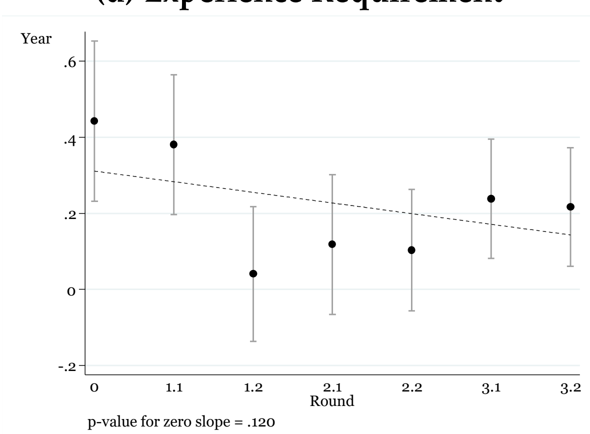
(b) Posted Wage



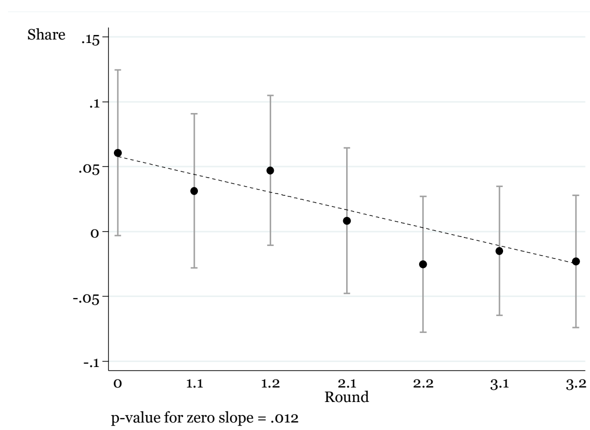
(c) Education Requirement



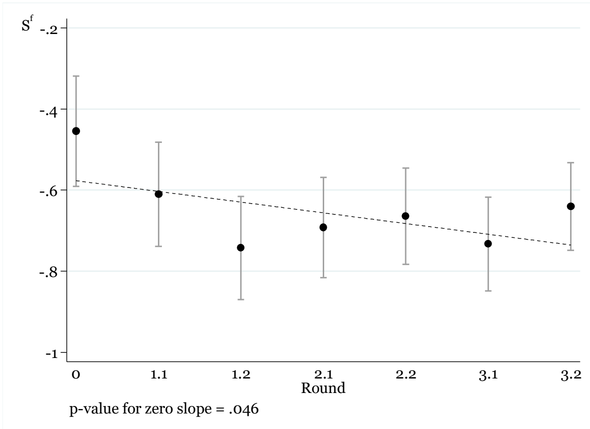
(d) Experience Requirement



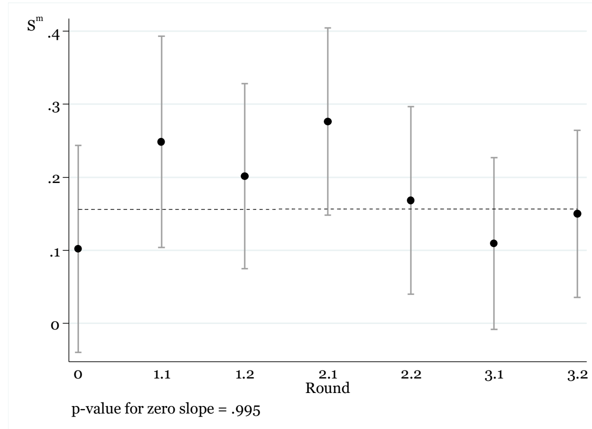
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D4 replicates Figures 3 and 4 using job recommendations from Job Board 4, showing the difference rate, gender differences in posted wage, education, working in experience, female content and male content by each round, respectively.
2. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
3. All panels show regression lines of the gender gap in the outcome on a round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively). Sample sizes vary between 1,780 and 1,936, depending on the outcome. p-values for a slope of zero are shown below each panel.

Appendix E: Heterogeneity and Robustness

E1: Ranking Differences

The set difference rate between jobs recommended to men and women does not consider the ranking of jobs in workers' recommendation lists. To see whether this affects our results, Table E1 replicates Table B4 using the *ranking* difference rate. According to this measure, two job recommendation lists are the same only if the two jobs in the same rank are identical.

The ranking difference rate is defined as:

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

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

Example: 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	

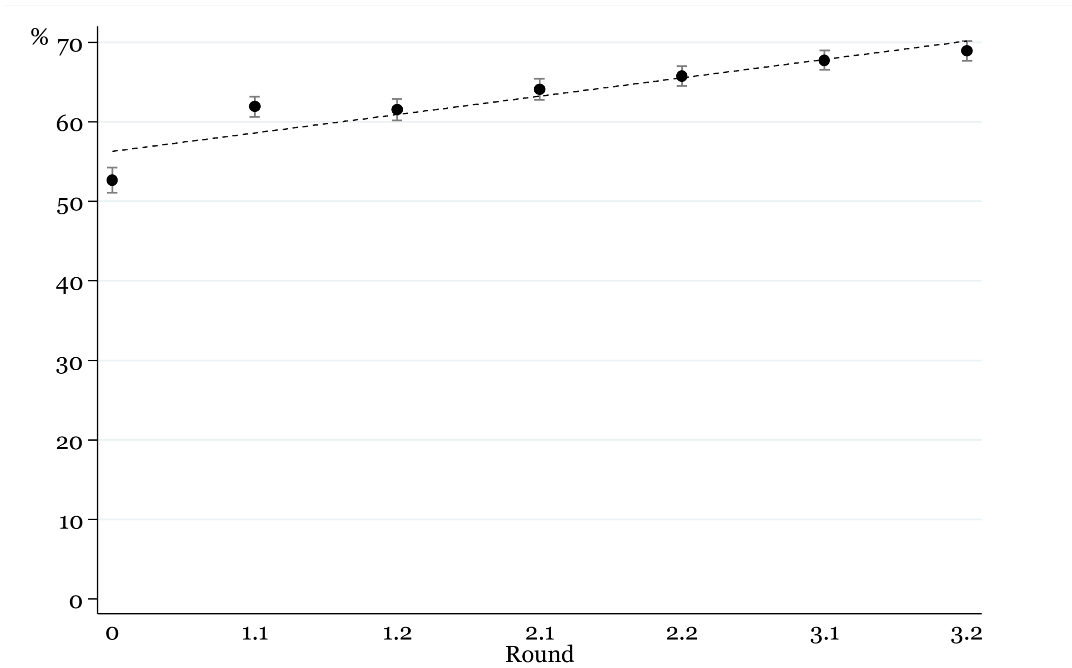
According to Table E1, the overall ranking difference rate is 61.32%, indicating that in a list of 100 recommended jobs, only around 39 jobs are displayed in the same rank to male and female applicants. That said, the cross-sectional patterns in ranking difference rates are very similar to the set difference rate, shown in Table B4.

Table E1: Ranking Difference Rate in Job Recommendations

	Difference Rate (S.D.)	Between-Group Differences
All Recommendations	0.6132 (0.090)	
Worker Age		
Young	0.6162 (0.094)	0
Old	0.6102 (0.095)	-0.0060
Job Gender Type		
Female-dominated	0.5986 (0.096)	-0.0249***
Gender Neutral	0.6234 (0.083)	0
Male-dominated	0.6185 (0.089)	-0.0049
Job Skill Level		
Entry	0.6022 (0.090)	0
Middle	0.6225 (0.091)	0.0202***
High	0.6158 (0.088)	0.0136**
Job Location		
Beijing	0.6035 (0.088)	0
Shanghai	0.6163 (0.086)	0.0128*
Shenzhen	0.6122 (0.089)	0.0087
Guangzhou	0.6208 (0.096)	0.0172**

In Figure E1, we replicate Figure 3 to show the trends across experimental rounds in *rank* (as opposed to set) difference rates. Figure E1 shows a very similar pattern, with the largest increase between Rounds 0 and 1.1, and a highly significant increasing trend overall (from 52.66% in Round 0 to 68.92% in Round 3.2.)

Figure E1: Ranking Difference Rate by Experimental Rounds



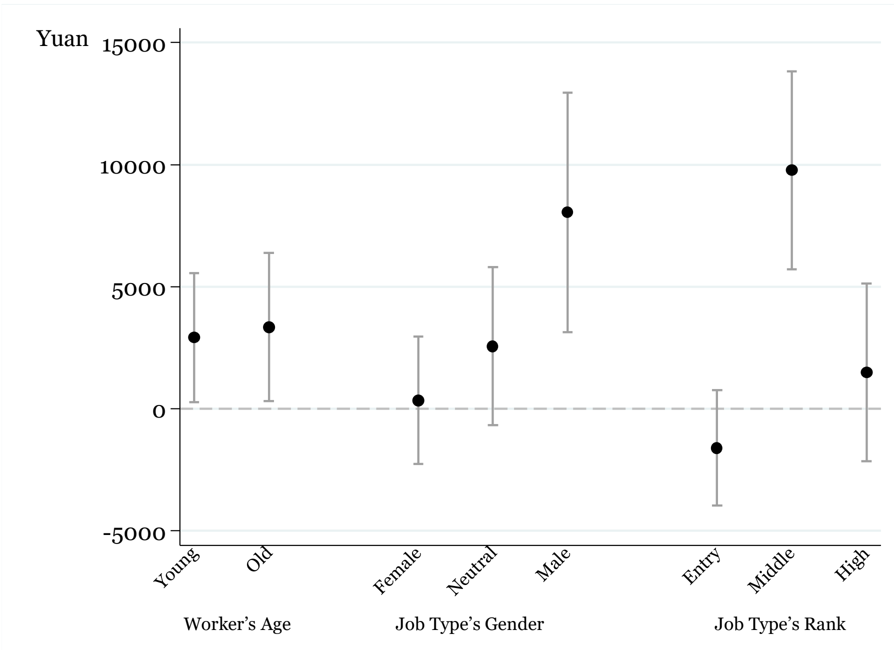
Notes:

1. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
2. Each round displays a 95% confidence interval.
3. A regression of the pair-level difference rate on the round indicator (equal to 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively) yields a coefficient of 2.314 with a standard error of 0.128 (p = 0.000; N = 7,746).

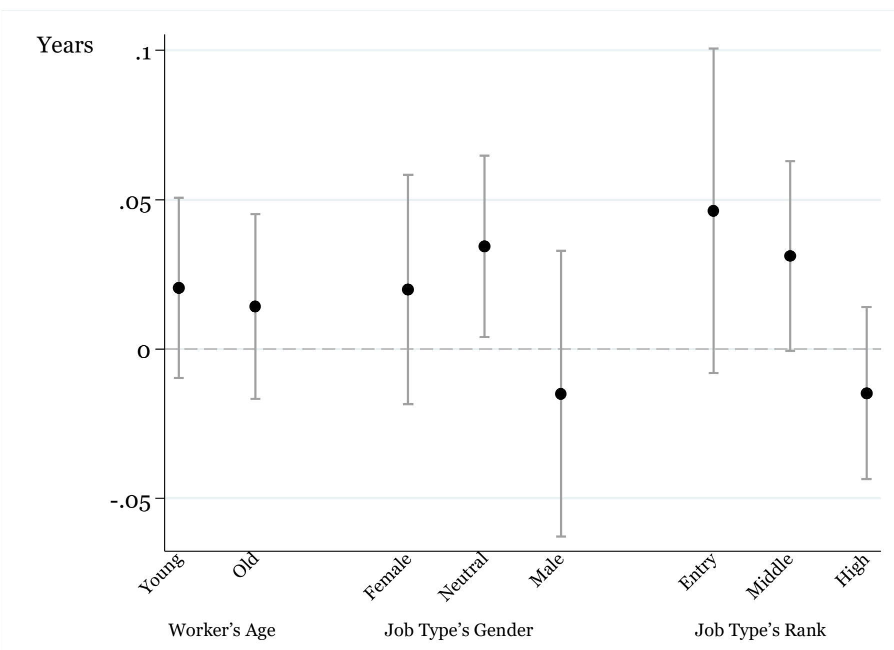
E2: Gender Differences in Recommended Job Characteristics by Applicant Age and Job Types

Figure E2: Heterogeneity in Gender Gaps

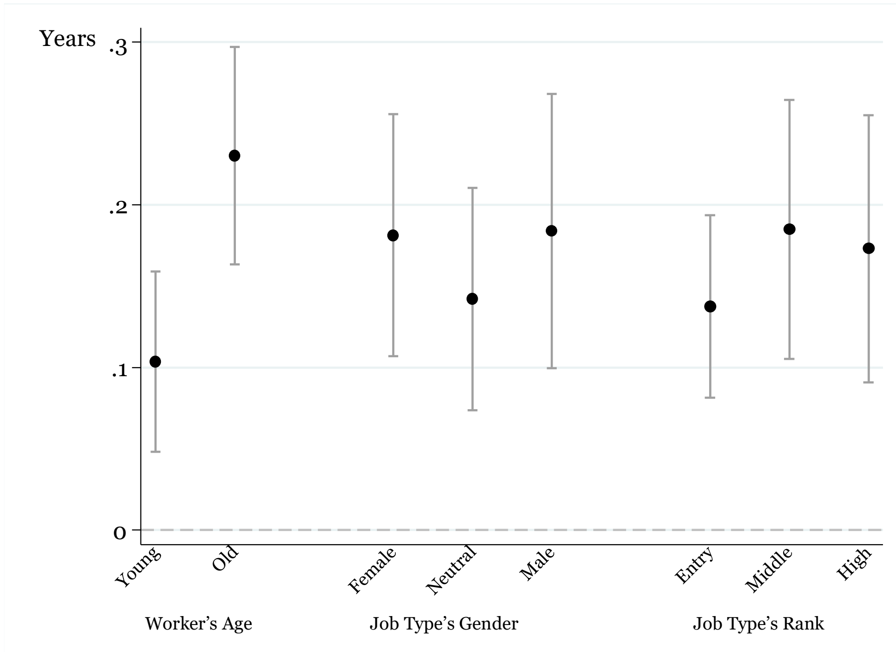
(a) Gender Differences in Posted Wages



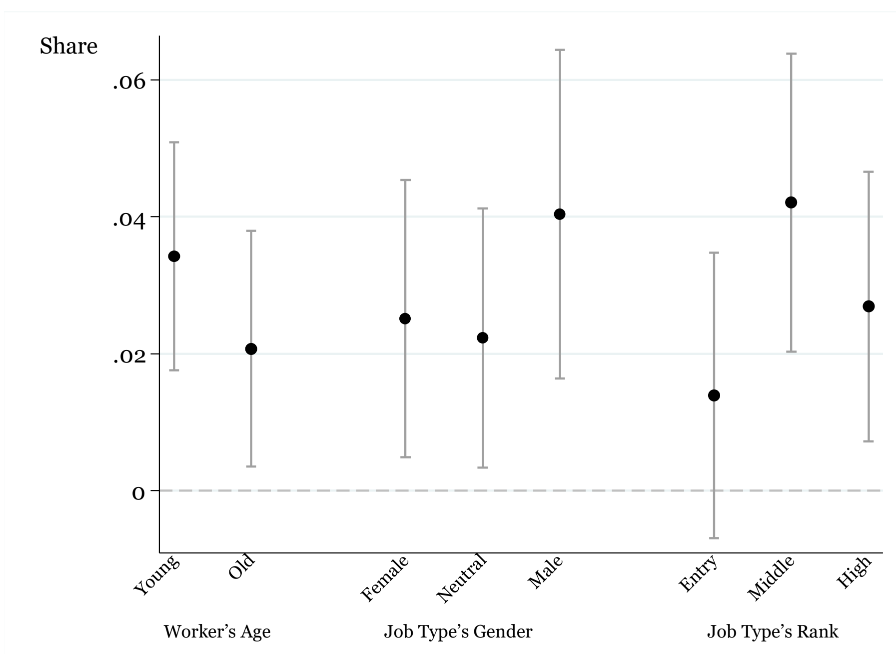
(b) Gender Differences in Requested Education



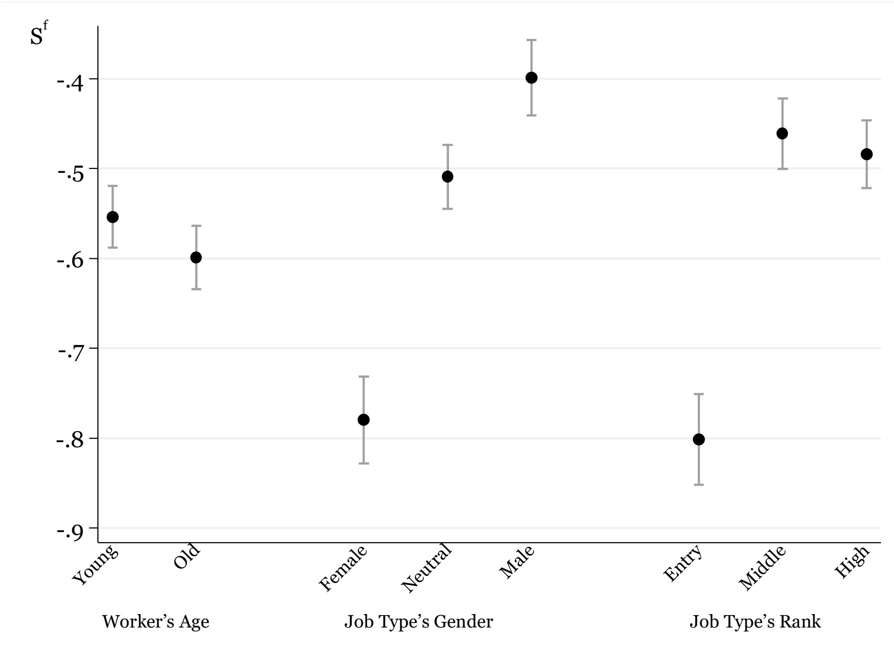
(c) Gender Differences in Requested Experience



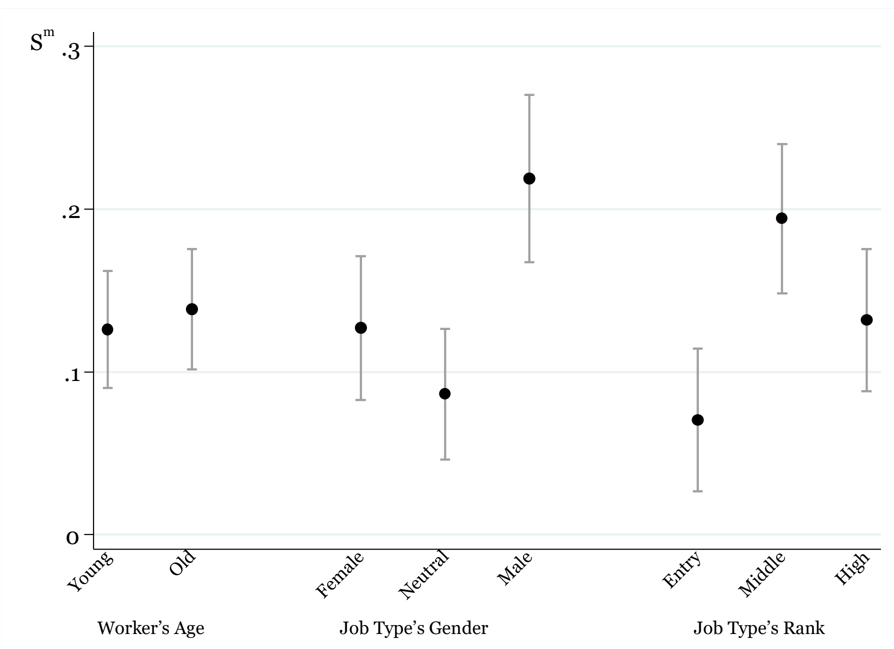
d) Gender Differences in Recommended Firm Size



(e) Gender Differences in Stereotypically Female Content (S^f)



(f) Gender Differences in Stereotypically Male Content (S^m)



Note: Young and Old refer to the age of the worker profile pair; Female, Neutral, and Male denote female-dominated, neutral, and male-dominated job types; Entry, Middle, and High denote job skill levels.

E3: Can Experience and Firm Size Gaps Account for the Gender Wage Gap?

In Table 1 we found that jobs recommended to men paid better, requested more experience, and were in larger firms. To shed some additional light on the gender wage gap in job recommendations, here we ask to what extent it can be attributed to these experience and firm size differentials. To accomplish this, Table E2 estimates the cross-sectional return to experience and firm size in all the jobs that were recommended to our profiles. Table E2 shows a robust and precisely measured positive wage return to experience and working in large firms, controlling for the job type, city, and job board fixed effects.

According to Table 1, job postings that are recommended only to men require 0.1656 additional years of experience and are 0.0276 times more likely to be in firms with 1000 or more employees. Using the coefficients in column 5 of Table E2, the combined effect of experience and firm size differences (and their interaction) accounts for a gender wage gap of 19,278*0.1656 + 27,098*0.0276 + 3,610*0.1656*0.0276 = 3,957 RMB, or 1.95 percent. This exceeds Table 1's actual gender wage gap of 3,118 RMB, or 1.54 percent. Thus, the experience and firm size gaps between the jobs recommended to men and women can (more than) fully account for the wage gap between those jobs.

Table E2: Cross-Sectional Returns to Experience and Firm Size, All Recommended Jobs

	(1) Posted Wage	(2) Posted Wage	(3) Posted Wage	(4) Posted Wage	(5) Posted Wage
Experience (years)	25,886***	22,988***	22,975***	20,548***	19,278***
	(192)	(187)	(186.275)	(182)	(222)
Firm Size (⩾1000)	52,272***	48,629***	48,274***	36,086***	27,098***
	(845)	(800)	(798)	(773)	(1,185)
Experience* Firm Size					3,610***
					(361)
Fixed Effects:					
Job Type (1-35)	Yes	Yes	Yes	Yes	Yes
City (1-4)			Yes	Yes	Yes
Job Board (1-4)				Yes	Yes
N	78,528	78,528	78,528	78,528	78,528
R2	0.223	0.318	0.322	0.383	0.384

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. The sample is all jobs that were ever recommended to our fictitious worker profiles.
2. Outcome is the (midpoint of the) job's posted wage in RMB, and the average is 205,928.
3. Experience is measured in years, and Firm Size (⩾1000) is a dummy variable that indicates whether jobs are posted by firms with over 1000 employees.

E4: Are there Gender Gaps in the 'Freshness' of Recommended Jobs?

On most job boards, recently posted job ads are especially valuable to workers (Albrecht et al. 2023); all the boards we study cater to this desire by including measures of an ad's 'freshness' when it is recommended to workers. Inconveniently, however, these measures differ across our four boards: Board 1 displays the date the job ad was posted, Board 2 displays the date the ad was last refreshed, and Boards 3 and 4 post the last time the recruiter who posted the ad was active.

To measure whether there is a gender gap in recommended ad freshness, we proceeded as follows. For each board, we constructed a continuous measure of elapsed time since the reported event (posting, refreshing, or recruiter activity). We then define a job ad as 'fresh' if this elapsed time is less than the median time on that board. Finally, we replicate the Table 1 regression separately for each of our four boards. The results are displayed in Table E3, where all the relevant coefficients are both small and statistically insignificant. We conclude that there is no gender gap in the freshness of jobs that are recommended to workers.

Reference:

Albrecht, James, Bruno Decreuse, and Susan Vroman 2023 "Directed Search With Phantom Vacancies" *International Economic Review* vol 64 no 2. Pages 837-869

Table E3: Gender Gaps in Job Ad Freshness

	(1) All boards combined	(2) Board 1	(3) Board 2	(4) Board 3	(5) Board 4
Male	-0.0005	0.0046	-0.0048	0.0178	-0.0169
	(0.007)	(0.016)	(0.014)	(0.013)	(0.012)
N	21,743	3,601	5,071	6,303	6,768
R^2	0.067	0.115	0.068	0.042	0.043
Mean of above	0.53	0.51	0.60	0.49	0.51

Notes:

1. This table replicates Table 1 for a different outcome variable—the freshness of the job posting.
2. The outcome variable equals one if the time elapsed since the last recorded recruiter action (posting the ad (Board 1), refreshing the ad (Board 2), or any recruiter action (Boards 3 and 4) is below the median for that job board.
3. Observations are job ads that are seen by a single gender in a profile pair; Male indicates the ad was seen only by the male profile.
4. Job Boards are numbered in the same order as in Tables B1 and B2, and in Appendix D.

E5: Are there Gender Gaps in Recommended Firms' Capital?

On two of our four job boards, job postings provide information on the employer's capital structure. On job board 3, job ads display the firm's registered capital, while job board 4 shows the firm's financing round, ranging from None, Angel, A, B, C, D+, to Public. To investigate a potential gender gap in these job characteristics, here we replicate Table 1 using them as outcomes. On job board 3, our outcome variable is set to one if the registered capital of the firm is above the median (13.36 million RMB). On board 4, the outcome equals one if the firm is listed as a public company or has received five or more rounds of financing (i.e. has attained financing round D+). Table E4 shows no significant gender difference in the financial status of recommended employers.

Table E4: Gender Gaps in Firms' Capital

	(1) Boards 3 and 4, combined	(2) Board 3	(3) Board 4
Male	0.0118	0.0207	0.0044
	(0.009)	(0.014)	(0.011)
N	11,906	5,413	6,493
R^2	0.145	0.059	0.185
Mean of outcome	0.36	0.41	0.32

Notes:

1. This table replicates Table 1 using a different outcome variable—firms' capital, as indicated in the job postings.
2. Observations are job ads that are seen by a single gender in a profile pair; Male indicates the ad was seen only by the male profile.
3. Job Boards are numbered in the same order as in Tables B1 and B2, and in Appendix D.

E6: Stereotypical Ad Content, by Dominant Gender in the Job Type

Table E5: Stereotypical Content in Jobs Recommended (Exclusively) to Men and Women, by Dominant Gender in the Job Type

	In Male-Dominated Job Types	In Female-Dominated Job Types
A. Stereotypically Female Content		
1. In jobs recommended only to men	-0.4192	0.1953
2. In jobs recommended only to women	-0.0212	0.9757
3. Gender gap in stereotypically female content (1-2)	-0.3980***	-0.7803***
	(0.022)	(0.027)
B. Stereotypically Male Content		
1. In jobs recommended only to men	0.1426	0.2356
2. In jobs recommended only to women	-0.0754	0.1061
3. Gender gap in stereotypically male content (1-2)	0.2180***	0.1295***
	(0.027)	(0.023)

Notes: Stereotypically male and female content are measured using the standardized indicators employed in Table 4.

E7: Gender Differences in Characteristics of Job Recommendations with Board x Week Fixed Effects

Table E6: Gender Differences in Characteristics of Job Recommendations with Board x Week Fixed Effects

	(1) Posted Wage (RMB)	(2) Education (years)	(3) Experience (years)	(4) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	3,113***	0.0172	0.1657***	0.0275***	-0.5759***	0.1323***
	(1,022)	(0.011)	(0.022)	(0.006)	(0.012)	(0.013)
N	21,262	19,899	21,922	22,023	22,023	22,023
R^2	0.613	0.454	0.394	0.172	0.302	0.122

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: This Table replicates regressions in Table 1 and 4. In addition to pair fixed effects, we also add the controls for job board × calendar week fixed effects (of the job recommendation).

E8: Gender Gaps in Stereotypical Ad Content using All Frequently Occurring Words

In Section 3.5, we developed an index of a job ad's stereotypical content by starting with the list of words that were statistically over-represented (in either male *or* female-directed) jobs ads on our boards. We then assigned stereotype scores to those words based purely on external sources, including previous literature and our own surveys. These scores were then used to calculate ad-level stereotype scores that yielded the large gender gaps in stereotypical ad content documented in Table 4 of the paper. The motivation for limiting the calculations to over-represented words is for consistency with Section 3.4, which provides a purely inductive description of the words that are over-represented in our data.

As a robustness check, Table E7 replicates the above analysis using the entire list of 172 most common words on our four datasets (without restriction to being over-represented in either only-to-male and only-to-female job ads). While we do not expect this to change our results –because we did not restrict our over-represented word list with respect to the *direction* of over-representation—readers might be concerned that the restriction to statistically over-represented words might skew our results in some way.

Table E7 shows that Table 4's coefficient magnitudes become somewhat smaller when we use the larger word sample to compute stereotype scores, likely because we now include many words that have little connection to gender stereotyping. Both coefficients of interest, however, remain highly statistically significant. The stark difference in the effect sizes of male versus female content also remains.

Table E7: Gender Differences in the Stereotypical Content of Job Ads (all words)

	Table 4 Results (using the 58 over-represented words)		New Results (using all 172 frequently-used words)	
	(1)	(2)	(3)	(4)
	Index of Stereotypically Female Content (S^f) (standardized)	Index of Stereotypically Male Content (S^m) (standardized)	Index of Stereotypically Female Content (S^f) (standardized)	Index of Stereotypically Male Content (S^m) (standardized)
Male	-0.5760***	0.1322***	-0.4081***	0.0898***
	(0.012)	(0.013)	(0.012)	(0.013)
N	22,023	22,023	22,023	22,023
R^2	0.297	0.117	0.255	0.125

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Columns 3 and 4 replicate Table 4 using all 172 "most common" words to compute word- and ad-level stereotype content scores. Sample and regression specification are the same as Table 4: Sample is all only-to-male jobs plus all only-to-female jobs. *Male* indicates the ad was only seen by the male profile in a gender pair. All regressions include pair fixed effects.
2. Our index of stereotypically female *ad* content is calculated as: $S^f = \sum_{w \in \text{ad}} s_w^f$, where s_w^f is the female stereotype score of each word in the ad, defined in Table 3. Stereotypical male ad content, S^m, is constructed analogously; both S^f and S^m are standardized to have a mean of zero and standard deviation of 1. Thus, column 1 indicates that compared to the ads that only the female profile saw the ads displayed only to male profiles contained words that were .576 standard deviations less stereotypically female.
3. Using the S^f and S^m indicators in columns 1 and 2 (derived from over-represented words only), the (unstandardized) means of S^f and S^m were 5.79 and 8.37 respectively across all job ads in our sample, and 6.66 and 9.09 in ads seen only by one member of a gender pair (i.e. the regression sample for all columns of Table E7).
4. Using the S^f and S^m indicators in columns 3 and 4 (derived from all frequently-occurring words), the (unstandardized) means of S^f and S^m 10.89 and 15.54 respectively across all job ads in our sample, and 12.13 and 16.82 in ads seen only by one member of a gender pair (i.e. the regression sample for all columns of Table E7).

E9: Replicating Tables 1, 4 and 5, including the 'Common' Jobs

While Section 3's main analysis compares only-to-male jobs to only-to-female jobs within gender pairs, it may be of some interest to see how these gender-exclusive jobs compare to the jobs that were recommended to both the male and female profile in the pair ('common' jobs).

To address this question, Table E8 replicates Tables 1, 4, and 5 using the full sample of all recommended jobs, using the common jobs as the omitted category and comparing the male- and female-only jobs to them. Combining all experimental rounds (Tables 1 and 4), both male- and female-only jobs pay less than common jobs, mirroring a pattern found for explicit gender requests in Kuhn and Shen (2013) and Helleseter et al. (2020). Female-only jobs require less education and experience than common jobs, while male-only jobs are similar to common jobs. With respect to stereotypically female content, common jobs fall between male-only and female-only jobs. However, common jobs contain less stereotypically male content than both male-only and female-only jobs. While we do not have appealing hypotheses for all the ways in which gender-specific jobs differ from common jobs, we remind readers that these differences do not affect our main estimand --gender gaps in the types of jobs recommended to women versus men—because the common jobs are seen by both genders. Future work may find ways to use the common jobs to learn more about the mechanisms used by job recommender algorithms.

The most striking feature of Table E8, however, is the strong explanatory power of stereotypically female content in explaining which jobs are seen by women, mirroring our main results in Tables 4 and 5. As discussed in the paper, this is consistent with the idea that male characteristics are in a sense the 'default' ones in labor market language, while female characteristics are strong signals of deviations from these defaults.

Table E8: Including 'Common' Jobs in Tables 1, 4, and 5

A. Tables 1 and 4

	(1) Posted Wage	(2) Education	(3) Experience	(4) Firm Size (⩾1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	-4,612***	-0.0154*	0.0239	0.0226***	-0.0645***	0.2295***
	(870)	(0.009)	(0.018)	(0.005)	(0.009)	(0.010)
Female	-9,511***	-0.0385***	-0.1570***	-0.0088*	0.5223***	0.0964***
	(909)	(0.009)	(0.019)	(0.005)	(0.010)	(0.011)
N	78,551	73,924	81,870	83,793	83,793	83,793
R^2	0.599	0.489	0.365	0.123	0.247	0.092

B. Table 5 (Round 0 recommendations only)

	(1) Posted Wage	(2) Education	(3) Experience	(4) Firm Size (⩾1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	-5,259***	-0.0016	-0.0058	0.0311***	-0.0887***	0.1784***
	(1,659)	(0.018)	(0.036)	(0.010)	(0.019)	(0.021)
Female	-9,686***	-0.0148	-0.2126***	0.0013	0.5301***	0.0697***
	(1,925)	(0.021)	(0.042)	(0.011)	(0.022)	(0.024)
N	22,078	20,734	23,067	23,610	23,610	23,610
R^2	0.633	0.513	0.378	0.148	0.266	0.124

Note: This Table replicates Tables 1, 4 and 5, including the jobs seen by both genders as the omitted category.

Appendix F: Hiring Agents' Views and Job Recommendations

In Table 7, we assessed the plausibility of Channels 4 and 7 (which are based on recruiters' views of our profiles' resumes) by measuring the correlation between the number of times a profile was viewed and subsequent gender difference rates and gender gaps in recommended job characteristics. In doing so, our regressor was the total number of profile views received by a gender pair. We found a positive association between profile views and the subsequent difference rate, but essentially no association between profile views and the gender gap in job characteristics, concluding that Channels 4 and 7 directed different jobs to male versus female profiles, but that these Channels did not change the *types* of jobs men versus women were shown.

In Tables F1 and F2, we test the robustness of this result by replicating Table 7 using number of views received by the male versus female profiles in a pair separately. In Table F3, we use the number of views received by the pair during only the first 10 recommendations received in each Round (to avoid any possible influence of the applications our profiles made after we collected those first 10 recommendations.) In all three tables, the resulting patterns are very similar to Table 7.

Table F4 provides background information for our assessment of Channels 4 and 7, showing that –even though our profiles submit their first applications *after* Interval 1— the number of profile views is higher in the two weeks before Interval 1 than any of the subsequent two-week Intervals. This underscores the importance of recruiter-based search of applicants' profiles on these job boards and suggests that recruiters have a strong preference for newly posted resumes. Table F4 also shows that men's profiles are 11.5 percent more likely to be viewed by recruiters than an identical female profile. This difference is highly statistically significant.

Finally, motivated by the gender gap in the number of times our profiles were viewed by HR agents, Table F5 replicates Table 7 using the gender gap (male minus female) in the number of profile views as a regressor. We do not detect any association between the within-pair gender gap in profile views and the gender gap in the *types of jobs* recommended to those profiles. Thus, while our results suggest that overall, human recruiters (and/or the resume search algorithms they use) are biased against women, these biases don't create gender gaps in the boards' job recommendations to workers.

Table F1: Effects of *Male* Profile Views during Intervals 1, 2, and 3 on Subsequent Gender Recommendation Gaps

		Gender Gap (*male – female*) in:					
	(1) Difference Rate	(2) Posted Wage	(3) Requested Education	(4) Requested Experience	(5) Firm Size (≥1000)	(6) Stereotypically Female Content (S^f)	(7) Stereotypically Male Content (S^m)
A. Interval 1							
Male Profile Views	0.1254***	-245	0.0106*	-0.0006	-0.0000	-0.0065	-0.0095
	(0.028)	(364)	(0.005)	(0.009)	(0.003)	(0.006)	(0.006)
N	1,118	1,073	1,016	1,089	1,090	1,090	1,090
R^2	0.317	0.005	0.007	0.010	0.010	0.063	0.041
B. Interval 2							
Male Profile Views	0.0364	215	-0.0008	0.0012	-0.0004	-0.0022	-0.0068
	(0.027)	(343)	(0.005)	(0.008)	(0.002)	(0.005)	(0.005)
N	1,100	1,078	1,049	1,088	1,089	1,089	1,089
R^2	0.338	0.009	0.005	0.006	0.006	0.027	0.032
C. Interval 3							
Male Profile Views	0.1341***	74	-0.0022	0.0045	0.0015	0.0014	0.0033
	(0.026)	(318)	(0.004)	(0.007)	(0.002)	(0.005)	(0.005)
N	1,095	1,082	1,054	1,090	1,090	1,090	1,090
R^2	0.305	0.007	0.006	0.009	0.018	0.057	0.029

Standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes:

1. This Table replicates Table 7, replacing the regressor with *Male Profile Views*, which is the total number of male profile views for the pair during each Interval.
2. Mean views of male profiles are 8.17, 7.65 and 7.17 in Intervals 1-3 respectively.

Table F2: Effects of *Female* Profile Views during Intervals 1, 2, and 3 on Subsequent Gender Recommendation Gaps

		Gender Gap (*male – female*) in:					
	(1) Difference Rate	(2) Posted Wage	(3) Requested Education	(4) Requested Experience	(5) Firm Size (≥1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
A. Interval 1							
Female Profile Views	0.1274***	-340	0.0001	0.0025	-0.0020	-0.0063	-0.0055
	(0.029)	(386)	(0.006)	(0.010)	(0.003)	(0.006)	(0.006)
N	1,118	1,073	1,016	1,089	1,090	1,090	1,090
R^2	0.316	0.005	0.003	0.010	0.010	0.063	0.039
B. Interval 2							
Female Profile Views	0.1506***	804**	0.0032	-0.0071	0.0029	-0.0055	-0.0027
	(0.030)	(390)	(0.006)	(0.009)	(0.003)	(0.006)	(0.006)
N	1,100	1,078	1,049	1,088	1,089	1,089	1,089
R^2	0.352	0.012	0.006	0.006	0.007	0.028	0.031
C. Interval 3							
Female Profile Views	0.0989***	429	-0.0053	0.0048	0.0014	-0.0068	0.0037
	(0.030)	(359)	(0.005)	(0.008)	(0.002)	(0.005)	(0.005)
N	1,095	1,082	1,054	1,090	1,090	1,090	1,090
R^2	0.295	0.008	0.007	0.009	0.017	0.058	0.029

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. This Table replicates Table 7, replacing the regressor with Fe*male Profile Views*, which is the total number of female profile views for the pair during each Interval.
2. Mean views of female profiles are 7.67, 7.01 and 6.21 in Intervals 1-3 respectively.

Table F3: Effects of Profile Views during Rounds 1.1, 2.1, and 3.1 on Subsequent Gender Recommendation Gaps

		Gender Gap (*male – female*) in:					
	(1) Difference Rate	(2) Posted Wage	(3) Requested Education	(4) Requested Experience	(5) Firm Size (≥1000)	(6) Stereotypically Female Content (S^f)	(7) Stereotypically Male Content (S^m)
A. Interval 1							
Views	0.0906***	533	0.0032	0.0060	0.0014	-0.0079	-0.0065
	(0.026)	(325)	(0.004)	(0.008)	(0.002)	(0.005)	(0.005)
N	1,118	894	813	925	928	928	928
R^2	0.164	0.010	0.007	0.023	0.005	0.034	0.023
B. Interval 2							
Views	0.0779***	533*	0.0047	0.0021	0.0029	-0.0042	-0.0065
	(0.027)	(310)	(0.004)	(0.008)	(0.002)	(0.005)	(0.005)
N	1,100	948	869	980	984	984	984
R^2	0.193	0.012	0.004	0.007	0.006	0.015	0.011
C. Interval 3							
Views	0.0748***	154	0.0008	0.0046	0.0033*	-0.0061	0.0002
	(0.027)	(277)	(0.004)	(0.007)	(0.002)	(0.004)	(0.004)
N	1,095	1,023	946	1,055	1,057	1,057	1,057
R^2	0.162	0.007	0.006	0.009	0.017	0.035	0.023

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. This Table replicates Table 7 using data from Rounds 1.1, 2.1 and 3.1 only. Observations are profile pairs, consisting of an identical male and female profile.
2. The regressor, *Views*, is the total number of profile views for the pair (male plus female) during each Round. Dependent variable in column 1 is the number of gender-specific jobs per 100 job recommendations received by the two applicants in each gender pair (difference rate*100), and in column 2 to 6 the outcomes are the gender difference rate in recommendations received immediately after the Interval, or the gender gap (male – female) in those recommended jobs' characteristics.
3. Panel A regresses gender recommendation gaps during Round 1.1, on the number of views the pair received during the preceding two weeks (interval 1).
4. Panels B (C) regress gender gaps during Round 2.1 (3.1) on the number of views the pair received during interval 2 (3).
5. All regressions control for the pair's age, the gender type of the pair's current (and sought) job, and job board fixed effects.
6. Mean profile views are 15.83, 14.66, and 13.38 in Intervals 1-3 respectively.

Table F4: Mean Number of Resume Views per Gender Pair, by Interval

Interval	Views of the Male Profile	Views of the Female Profile	Total Views	Male-Female
1	8.17	7.67	15.83	0.50**
2	7.65	7.01	14.66	0.64***
3	7.17	6.21	13.38	0.96***
4	6.78	5.79	12.57	0.98***
All Intervals	29.76	26.68	56.45	3.08***

Notes:

1. Each cell shows the number of views generated during the two-week-long intervals between our experimental Rounds. For example, during interval 2 (i.e. between Rounds 1 and 2) an average profile was viewed 7.65 times.
2. Means are based on 1,120 male and 1,120 female profiles.
3. The last column shows the difference of views between male and female profiles, *** p<0.01, ** p<0.05, * p<0.1.
4. The gender gap in the number of profile views is (29.76-26.68)/26.68 = 11.5 percent.

Table F5: Effects of the *Gender Gap* in Profile Views during Intervals 1, 2, and 3 on Subsequent Gender Recommendation Gaps

		Gender Gap (*male – female*) in:					
	(1) Difference Rate	(2) Posted Wage	(3) Requested Education	(4) Requested Experience	(5) Firm Size (≥1000)	(6) Stereotypically Female Content (S^f)	(7) Stereotypically Male Content (S^m)
A. Interval 1							
M-F Views	0.0089	43	0.0081*	-0.0022	0.0013	-0.0006	-0.0035
	(0.024)	(319)	(0.005)	(0.008)	(0.002)	(0.005)	(0.005)
N	1,118	1,073	1,016	1,089	1,090	1,090	1,090
R^2	0.304	0.004	0.006	0.010	0.010	0.062	0.039
B. Interval 2							
M-F Views	-0.0548**	-276	-0.0022	0.0046	-0.0018	0.0014	-0.0031
	(0.022)	(283)	(0.004)	(0.007)	(0.002)	(0.004)	(0.004)
N	1,100	1,078	1,049	1,088	1,089	1,089	1,089
R^2	0.340	0.009	0.006	0.006	0.007	0.027	0.031
C. Interval 3							
M-F Views	0.0374*	-172	0.0012	0.0005	0.0002	0.0044	0.0003
	(0.022)	(257)	(0.004)	(0.006)	(0.002)	(0.004)	(0.004)
N	1,095	1,082	1,054	1,090	1,090	1,090	1,090
R^2	0.290	0.007	0.006	0.009	0.017	0.058	0.029

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. This Table replicates Table 7, replacing the regressor with gender gap (male—female) in the number of profile views The means of gender gap in profiles views are 0.50, 0.64 and 0.96 in Intervals 1-3 respectively.