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, Revised November 2025

We thank Amanda Agan, 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), LM2C2Workshop on Gender and Labor Market Mismatches (Santiago, Chile), the 2025 meeting of the Trans-Pacific Labor Seminar (Teikyo University), the 2025 Society of Labor Economists conference (Toronto), and the 2025 KOF/ETH Zurich and IDSC/IZA Workshop: Matching Workers and Jobs Online 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 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, Revised November 2025
JEL No. C99, J16, J71, M50, O33

ABSTRACT

We use an algorithm audit of China's four largest job boards to measure the causal effect of a job seeker's gender on the jobs that are recommended to them, and to identify the algorithmic processes that generate those recommendations. Focusing on identical male and female worker profiles seeking jobs in the same industry-occupation cell, we find precisely estimated but modest amounts of gender bias: Jobs recommended to women pay 0.2 percent less, request 0.9 percent less experience, come from smaller firms, and contain .07 standard deviations more stereotypically female content such as requests for patience, carefulness, and beauty. The dominant driver of these gender gaps is content-based matching between posted job ads and the declared gender in new workers' resumes. 'Action-based' mechanisms – based on a worker's own actions or recruiters' reactions to their resume – contribute relatively little to the gaps we measure.

Shuo Zhang
Northeastern University
s.zhang@northeastern.edu

Peter J. Kuhn
University of California, Santa Barbara
Department of Economics
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

Job boards (like Careerbuilder, Indeed, Naukri, and Zhaopin) now play a central role in the labor market matching process; they also possess a wealth of data on job ads, resumes, and the (current and historical) interactions between them. Many boards are now using this data to suggest matches between firms and workers, with the goal of improving the outcomes of both groups (Bied et al., 2023a). As is well known, however, algorithmically generated recommendations can learn humans' biases and stereotypes from training data (Prates et al., 2020; Miceli et al., 2022). The size, nature and direction of these biases remains unknown, in part due to the relatively recent adoption of active job recommender systems. Since most of the world's largest job boards are proprietary, even less is known about the processes and pieces of information that generate any biases.[1]

In this paper, we use an algorithm audit (Metaxa et al., 2021) to answer the above questions for the job recommender systems used by China's four largest job boards, which together host over 100 million monthly active job seekers and over 25 million paying employers.[2] Specifically, focusing on identical profiles of male and female workers looking for jobs in the same industry, occupation, and city, we measure the causal effect of a job seeker's gender on the jobs that are recommended to them. We also identify the information (such as resume content and workers' past actions) and the algorithmic processes (such as rules-based matching and collaborative filtering) that are used to make recommendations.

In more detail, we created identical worker profiles on each job board, which differ only in their declared gender (male or female) and a gender-matched name (henceforth

[1] Appendix B1 presents the results of an AI-based analysis of job board traffic in 16 countries, which together represent the 10 largest economies and 10 most populous countries. While a publicly operated platform has the most traffic in three of those countries (Germany, France, and Pakistan), six countries have no publicly operated boards in the top ten by traffic. The remaining countries have a single publicly operated board in their top ten, ranked between fourth and ninth in the country. Among the top ten boards in all countries, the population-weighted mean market share of all the public boards is 3.4 percent.

[2] Numbers are for 2025, taken from the boards' own web pages. The boards are ranked in China's top four on most size metrics in industry reports. To protect the individual boards' identities, we cannot provide links to sources that reveal the boards' names.

'gender').[3] 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 our profiles' application decisions and on recruiters' reactions to their resumes, our profiles 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 the job recommender algorithms on all these boards recommend lower-wage, lower-ranked jobs to women than men and that the characteristics of the recommended jobs align with widely held gender stereotypes. While precisely estimated, these gaps are small in magnitude: Jobs recommended to women pay 0.2 percent less, request 0.9 percent less experience (suggesting a lower rank), are 1 percent less likely to come from firms with over 1,000 workers, contain .072 standard deviations more stereotypically female content, and .016 standard deviations less stereotypically male content. For example, 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*. In contrast, *leadership*, *entrepreneurial*, and *work under pressure* appear more often in male-only jobs. Job characteristics like *commissions*, *night work*, *overtime*, and *travel* (male) and *parental leave*, *eight-hour days*, and *flexible hours* (female) also differ.[4] While highly statistically significant, one important reason why our estimated gaps are so small is the large overlap in the individual job ads seen by identical male and female worker profiles: 87.6 percent of those ads *are the same ones*.

To understand how these recommendation gaps are generated (and why they are so small), the second half of our paper attempts to isolate which pieces of information and which common components of recommender algorithms are used by our job boards, and

[3] 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.

[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.

which ones account for the gender gaps in recommended jobs' characteristics. Specifically, we divide the set of possible mechanisms into three main types based on the information they use: Content-based, or C-mechanisms use rules and/or natural language processing (NLP) to identify similarities between the contents of a worker's profile and the contents of current posted vacancies. W-mechanisms use a worker's *actions* (in our case their previous job applications) to identify other jobs that same worker might want, while R-mechanisms use the actions of recruiters who encountered our profiles (such as viewing, clicking, or downloading the resume) to recommend jobs to those profiles.[5]

With respect to mechanisms, our first main finding is that C-Mechanisms are present on our job boards and create gender gaps in recommended ad content. We base this conclusion on the fact that all the gender gaps described above are present at the very start of our experiment, when only C-mechanisms are available: In Round 0, none of our profiles have taken any actions (such as viewing, clicking or applying to jobs), and no recruiters have had a chance to react to our profiles. Our finding that C-mechanisms create gender gaps is consistent with the fact that NLP processes –which are widely used to compute similarities between jobs and resumes– are known to learn human biases (Prates et al., 2020; Miceli et al., 2022). Our second main finding is that these C-mechanisms must be using the worker's declared gender as an input: This follows from the fact that our profiles differ only in their gender. While explicit use of applicants' gender by job recommender algorithms is permitted in China, this illustrates how algorithm audits –as designed in our paper– can be used by outsiders to test for the explicit use of any worker characteristic.

Third, we find suggestive evidence that R-mechanisms are active on our job boards and work to accentuate gender gaps in one highly informative job characteristic– the amount of stereotypically female content in the job ad. (Of all the job characteristics we measure, stereotypical female content is by far the best predictor of which gender sees the ad.)

[5] Importantly, the W- and R-mechanisms identified by our experiment capture the effects of actions taken by our profiles *after they are published*, and of recruiters' reactions to those profiles. Historical actions taken by other workers and recruiters can also affect a board's recommendations if they are used to train content-based matching algorithms. (For example, historical data on applications and call-backs could be used to create a worker-job similarity metric.) When such historical actions are used to train content-matching routines, they are categorized as C-mechanisms in our experiment.

For example, recruiters may react more positively to female profiles in jobs with features like flexible hours, administrative duties, and assisting roles. The fact that R-mechanisms are active highlights an important difference between our platforms and one-sided search platforms, like shopping (Amazon), entertainment (Spotify), and information retrieval (Wikipedia), where the analog of R-mechanisms is mostly irrelevant: Items on Amazon do not care who buys them, and sellers on Amazon rarely care who their customers are. In contrast, the algorithms we study appear to recognize and exploit the two-sided nature of labor markets when recommending matches to workers.

Finally, if job boards use workers' applications during a previous logon session to recommend jobs at the start of the next session, W-mechanisms ('you applied to job x, so you might be interested in job y) should cause the characteristics of recommended jobs to exhibit growing gender gaps across experimental Rounds after our profiles start submitting applications. (This is because gender gaps exist before any applications are sent and our profiles apply to their recommended jobs.) We document such growing gaps, but (again) only for jobs' stereotypically female content. Thus, across-session W-mechanisms appear to be active, and to accentuate gender gaps in stereotypically female content when workers follow the boards' recommendations.[6]

This paper contributes to four literatures, the first of which uses resume audit methods to study employers' responses to applications with different characteristics (Bertrand and Mullainathan, 2004; Kline et al., 2022). We extend the resume audit method by applying it to algorithms rather than people; this is important because proprietary black-box algorithms are increasingly prevalent as economic actors. While our algorithm audit studies a different outcome –who *sees* a job ad?– than resume audits, we note that algorithm audits like ours have some practical advantages, including greater scalability: Algorithm audits can be conducted electronically with sparse worker profiles that do not require the investigator to fabricate detailed personal working histories and statements of purpose,

[6] Our non-experimental interactions with the job boards reveal that all four of them also use W-mechanisms to provide *within-session* job recommendations: Right after workers apply to a job, the board shows them a list of seven to fifteen similar jobs they might be interested in. Three of the four boards also offer a 'batch apply' button that automatically sends applications to all the jobs in this list. Our design cannot capture the effects of following these within-session recommendations.

or to make formatting decisions (font, margins, etc.) that consume investigator resources and introduce noise.[7] Algorithm audits are also harder for employers to detect (Avivi et al., 2021) and (unlike resume audits, which consist of a single action –submitting an application) allow investigators to probe their subject's responses by taking a series of actions, such as viewing, clicking, or applying to jobs. Finally, algorithm audits have an ethical advantage because the inconvenience to human recruiters is negligible: the recommendations we study are made by machines, not people.[8]

A second related literature studies gender and other differences in *job application behavior* –a phase of the job search process that precedes the candidate selection phase studied by resume audits. This literature has studied gender differences in application rates to jobs that are far away (Eriksson and Lagerström, 2012; Le Barbanchon et al., 2021), have flexible hours (Mas and Pallais, 2017), competitive work environments (Flory et al., 2015), ambiguous information about job requirements (Coffman et al., 2023; Abraham et al., 2024; Kline et al., 2022) or affirmative action statements (Ibañez and Riener, 2018). We contribute to the application behavior literature 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? If automated job recommender systems inadvertently channel workers toward jobs that match common gender stereotypes, algorithms can create the appearance that men and women are choosing to apply to stereotypical jobs, when in fact men and women simply see different jobs when they're looking for work.

Third, our work relates to a substantial literature in computer science that measures algorithmic bias and stereotyping in a wide variety of contexts, including Google Image search results (Vlasceanu and Amodio, 2022); AI-generated yes/no decisions for hypothetical security clearances, dating, and employment (Tamkin et al., 2023); and price and steering differences on e-commerce sites Hannak et al. (2014).[9] Computer science studies

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

[8] Our fictitious resumes applied to jobs in Rounds 1-3 of the experiment, and were infrequently contacted by (presumably) 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.

[9] Barocas et al. (2023) provide an overview that defines various fairness concepts used in this literature.

of *job ad delivery* take two main forms. The more common form is 'insider' or 'lab' studies of job boards' recommender systems, where investigators apply known algorithms to generate a set of job recommendations for a population of naturally-occurring resumes, typically drawn from historical data (Rus et al., 2022; Peña et al., 2023). For example, Bied et al. (2023a) evaluated their own job recommender system by generating recommendations for all the workers in a proprietary data set from the French Public Employment Service. They then used regression controls to determine whether equally qualified workers receive different job recommendations.[10]

Less frequently, computer scientists have also audited live platforms as outsiders to measure biases in their recommendations of products, services, and jobs. With respect to job ads, the main studies we know of –Ali et al. (2019) and Imana et al. (2021)– proceed in a similar way as the insider studies: They purchase job ads on general-purpose social networking platforms and investigate who sees them, using aggregate statistics provided by the platform's advertiser interface or by creating their own audience of users at risk of seeing the ads. In this approach, the investigators' ability to control for the qualifications of workers seeing the ads is even more limited than in the insider studies.[11]

Summing up, we contribute to the computer science literature on job ad delivery in two main ways. First, by adapting the resume audit approach to job recommender algorithms, we can observe the job recommender system's responses to identically qualified male and female resumes, with identical job search objectives. This yields 'clean' experimental measures of algorithmic bias and avoids the need for regression controls for skill. It also provides a conclusive test for whether a platform's algorithms are explicitly using the worker's gender (or other protected characteristic) as an input. Second, to our knowledge we are the first outsider audit of job ad delivery on job boards, which are the platforms on which most online job search takes place.[12]

Other examples are Bolukbasi et al. (2016) and Kay et al. (2015).

[10] Some economists have also experimented with job recommender algorithms from an insider perspective. See for example Belot et al. (2019), Hensvik et al. (2025).

[11] For example, Imana et al. (2021) define ad delivery algorithms as non-discriminatory if the female share of job ad recipients is the same for jewelry sales as for car sales (because both require sales skills).

[12] Chen et al. (2018) conduct an outsider study of resume search engines, which serve the employer side of

A final related literature studies bias in commercial and investigator-supplied *prediction algorithms* for outcomes that include parolee recidivism (Dressel and Farid, 2018), mortgage default risk (Fuster et al., 2022), future illness (Obermeyer et al., 2019; Kilby, 2021), and work performance (Hoffman et al., 2018; Li et al., 2020).[13] Prediction algorithms differ from the recommender systems we study because they take as given a low-dimensional, objective measure of decision quality (like re-arrest rates and mortgage default rates).[14] This gives investigators a yardstick against which to measure bias, but creates an analytical challenge because the outcome is typically only measured for a selected sample (such as the offenders who receive parole.) In contrast, the outcomes we study –a set of recommended jobs– are observed for all our resumes, but these outcomes are highly multidimensional. Indeed, we are particularly interested in characterizing gender differences in the unstructured text of recommended job ads, without imposing any prior assumptions about how we might expect that text to differ.

2 Experimental Design

2.1 Platform Environments

The four job boards we study have similar interfaces and functions: Job seekers can register and create a profile for free, while employers are charged for posting 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' 'ap-

the labor market by recommending resumes to recruiters. While resume search may be important in some labor markets, worker-initiated search appears to be much more common. For example, data from one of our 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.

[13] 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 simulate 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.

[14] Like studies of recommender systems, studies of prediction algorithms include both 'insider' studies of known algorithms (Kilby, 2021; Li et al., 2020), and 'outsider' studies of live, commercial products (Dressel and Farid, 2018; Obermeyer et al., 2019).

ply' buttons. Firms' hiring agents can search for workers, view resumes, process applications and contact applicants through each board's recruiter-facing portal. All four of these boards claim to 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 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-industries 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 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' 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 board. 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 real 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 current occupation and industry are the same as the job type's occupation and industry, and all workers are seeking

[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.

jobs in their current city and occupation-industry cell.[19]

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 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

Since many search platforms learn from their users' previous actions –and our platforms' algorithms are not known to us– our experimental design casts a wide net for possible mechanisms using a multi-stage, *sock puppet* design. Sock puppet designs create fictitious online entities – worker profiles in our case– that follow a prescribed sequence of actions that help us infer some aspects of how an algorithm works. These designs contrast with traditional resume audits, where fictitious workers engage in a single action–applying for a job.

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 top 100 job ads shown to each worker, and the workers log off.[20]
- **Round 1**. Two weeks later, the male and female workers simultaneously log into

[19] Appendix A2 provides additional details on this process.

[20] The vast majority of our profiles (97.8 percent) received at least 100 job recommendations in Round 0; all of them received at least 20. For consistency across experimental Rounds, all our main analyses use either the top 10 or 20 Round 0 recommendations, so missing values are not a concern.

their accounts again. We then record the number of times their profile was viewed by HR agents since the worker's account was published.[21] 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.
- **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.[22] 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. That noted, the 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

[21] 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).

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

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 be within a fairly narrow occupation-industry range.[23] Rather than directing, say, women out of highly-male *job types* and into more female ones, we expect most of the gendered steering in our experiment to occur on subtler margins, such as workers' preferences for work hours, competition, and employers' gendered preferences for beauty and personality types.

2.5 Descriptive Statistics

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, our 2,240 fictitious profiles received 177,320 job recommendations from 81,231 individual job ads.[24] Descriptive statistics on our samples of fictitious workers and the jobs recommended to them are provided in Tables B2.1-B2.3. Reflecting the high wage levels of jobs and resumes on these platforms, the average annual wage of our resume sample is 142,507 RMB, or about twice the 2020 average wage in urban China.[25] The workers' desired wages are 26.12% higher than 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.[26]

[23] While workers' current and desired job types are categorized using the same industry-occupation cells used by recruiters to post jobs, as 'outside' auditors we do not observe the exact job type of the ads that are displayed to our fictitious resumes. Thus we cannot measure the exact correspondence between recommended jobs' and workers' desired job types. That said, Appendix B3 compares the content of recommended job titles with the content of workers' desired occupations and shows that the boards do not steer users very far from workers' desired occupations.

[24] These 177,320 recommendations were about 0.5 percent less than our designed number (2,240*80 = 179,200). The main reasons are that some job boards blocked suspicious workers' accounts and some hyperlinks to jobs were blank. The share of missing recommendations is independent of the workers' gender.

[25] 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).

[26] 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 B2.1,

The characteristics of the job ads that were recommended to our fictitious workers are summarized in Table B2.2. Over 95% of recommended jobs posted a wage (or wage range), and one-third of the recommended positions are from companies that have more than 1,000 employees. The average posted wage in recommended jobs was 205,928 RMB; mean requested years of education and experience were 15.42 and 2.44 respectively.[27] Overall, the jobs recommended to our fictitious workers were well matched with those workers, as shown in Table B2.3. 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, 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.

which is calculated from the midpoints of these desired wage ranges.

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

3 Results

Our first set of results combines the data from Rounds 0-3 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. To represent recommendations from across all Rounds equally, these results use only the top 20 recommendations from Round 0, to match the 20 recommendations we collected in each of Rounds 1-3.

3.1 The Set Difference Rate

The most basic measure of the difference between jobs recommended to men versus women is the share of job ads in a pair of top-N recommendation lists that are unique to a gender, 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.[28] 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 show 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.[29]

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%.

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

[29] For example, if the set difference rate begins to increase after employers can start searching workers' resumes, that suggests that *R*-mechanisms are operative on a platform. See Section 4.2.

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.[30] Table B2.4 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 middle 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 E1.1 and Figure E1.1 replicate the preceding analysis using the *ranking* difference rate. While (by construction) 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 estimation 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 esti-

[30] As noted, these numbers are based on 20 recommendations from each of Rounds 1-4. Thus –With the exception of a very small number of worker profiles that did not receive a full set of recommendations– N = 80 in equation 1.

mate 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.[31] 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 and female-only recommendations within gender pairs.

Our baseline estimates of equation 2 are reported in Table 1, which shows that –among jobs that are unique to men or women– 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 gaps 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. Including the 87.6 percent of recommended jobs that are shared by men and women, these gaps shrink to 0.2 percent for wages, 0.9 percent for experience, and 1.0 percent for firm size.[32] Interestingly, using cross-sectional returns calculated from posted job ads, Appendix E3 shows that these differences in education and firm size can fully account for the gender wage gap we estimate.

Figure E2.1 explores heterogeneity in Table 1's gender recommendation gaps by ap-

[31] 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 E7 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. 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.

[32] Another recommended job characteristic that could vary by gender is a job's 'freshness', i.e. the elapsed time since it was posted or last refreshed. Since freshness 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. Finally, to check Table 1 (and Table 4) for robustness, Table E6.1 replicates both of them, adding board-by-week fixed effects to control for time-varying conditions specific to a job board (such as periodic updates to algorithms). The results were very similar.

plicant 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.[33] The gender wage gap is considerably higher 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.

Overall, the above results show a statistically significant but small gender gap in the wages, experience requirements, and firm sizes of jobs recommended to men versus women. 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 with similar characteristics.[34] To see if these small gaps extend to other, harder-to-quantify aspects of recommended jobs, we now study the unstructured text of the words in those ads.

3.3 Gender Differences in Stereotypical Ad Content

To characterize gender differences in the open text of job ads recommended to men versus women in a transparent and reproducible way, we use a five-step procedure. Our approach works directly with the (standardized) words in job ads (in contrast to, for example, their vector embeddings) since our goal is inferential, not predictive: We want to know which natural words are most over-represented in ads seen by men versus women and to measure the extent to which they conform to commonly held stereotypes. Accordingly, we first transform the text of all the recommended ads into standardized chunks and retain only the chunks (henceforth 'words') that appear more than 100 times. This yields a corpus of 172 'most common words' for further analysis. Next, correcting for multiple

[33] 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.

[34] Appendix B3 characterizes the match between workers' requested and recommended occupations, showing an average similarity score of 0.73.

hypothesis testing, we identify which of these words were statistically over-represented in jobs seen only by men (22 words), or in jobs seen only by women (36 words).

Third, we turn to a variety of sources external to our study to determine which of the 172 words in our corpus are widely seen as stereotypically male or female, and assign scores running from 0 to 4 to measure each word's male and female stereotype intensity. Fourth, we calculate the stereotypically male and female content of every recommended job ad by summing over the words it contains. Finally, we standardize these ad-level measures and used them as outcomes in equation 2, thereby quantifying gender differences in the amount of stereotypically male and female content between the jobs recommended to men versus women.

3.3.1 Parsing Job Ads into Their Most Common (Distinct) Words

We began our analysis of the unstructured text of recommended job ads with a database consisting of the 81,231 job ads that were recommended to our profiles. As noted, 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, resulting in a final selection of 172 chunks (henceforth 'words'), each represented by a single word or phrase, such as "listening", "marriage leave", or "regular working hours".

The 172 most common words that emerged from this process are shown in Figure 2, with a larger size representing a higher frequency.[35] 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 discussion of the words in Figure 2 below, we manually assigned them to the following six categories: standardized (PIACC) skills; job benefits; work timing and location; company information; qualifications (other than education

[35] Figure C1.1 shows these words in the original Chinese.

requirements); and personality, age, and appearance. A complete list of all the words, by category, is provided in Table C1.1.

3.3.2 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 C1.1 should appear with roughly equal frequency in the jobs recommended exclusively to the male and female job profiles. To test this hypothesis, we use the sample and regression specification in equation 2, but replace the outcome variable with a dummy for the appearance of each word in the recommended job; the regressor of interest in each of these 172 regressions 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 calculate Romano-Wolf p-values to control for the family wise error rate (Romano and Wolf, 2005*a*,*b*) and Anderson q-values (Anderson, 2008) to control for the false discovery rate (FDR), then we define our list of *over-represented words* as the 58 words whose p- and q-values are *both* below 5 percent.[36]

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, Table 2 uses the word categories we developed in Table C1.1. Starting with the standardized (PIACC) skills (OECD, 2016), 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

[36] 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.

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*. Consistent with evidence on explicit gender requests in Chinese job ads (Kuhn and Shen, 2013), words associated with physical appearance, such as *figure* and *facial* are also more common in only-to-female recommendations.

3.3.3 Relating Over-Represented Words to Gender Stereotypes

While many of the over-represented words identified above feel stereotypical to us, we turned to four external sources to assess to what extent each word represents commonly held stereotypes. The first external source is the union of word lists taken from three published studies: Gaucher et al. (2011)'s list of words that predict Canadian readers' perceptions of gender representation in jobs; Kuhn et al. (2020)'s list of words that predict whether Chinese job ads request men and women; and Chaturvedi et al. (2021)'s list of words that predict whether Indian job ads request men and women. The second and third sources are surveys we conducted ourselves, where we asked subjects whether each of our 172 words was more likely to appear in a job ad seeking men, women or neither. One survey was on MTurk, the other on a Chinese survey platform. Our final source is the results from the following query to ChatGPT 4.0: "Can you categorize each word in the following six categories as neutral, male, or female?" (OpenAI, 2023). With these four word lists in hand, each of our 172 words' stereotypical maleness and femaleness scores (s_w^f and s_w^m) are calculated as the net number of external sources (0-4) that identified the word as stereotypically male (female). Additional details on how these scores were calculated, and evidence on their robustness to changes in our procedure, are provided in Appendix C.

The stereotypical maleness or femaleness scores resulting from this procedure are shown in Table 3, which 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 (a female stereotype score of 4). 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 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 are indeed associated with commonly held gender stereotypes.

3.3.4 Quantifying Gender Differences in Stereotypical Ad Content

To quantify the total stereotype-reinforcing effect of recommendation algorithms, we 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 0.5760 standard deviations more stereotypically female content (S^f) than ads shown exclusively to men. Conversely, only-to-men jobs are only 0.1322 standard deviations more stereotypically male than only-to-women jobs. Rescaled to include the recommendations that are common to the male and female profiles, these gender gaps are 0.072 and 0.016 standard deviations respectively.

To compare Table 4's gaps to the wage, education, experience, and firm size gaps estimated in Table 1, we re-scaled Table 1's outcome variables to have unit standard deviations, yielding effect sizes of 0.0269, 0.0173, 0.0812, and 0.0530. Compared to these estimates and to the .1322 effect for stereotypically male content, the .5760 gender gap in stereotypically female content is much larger. Indeed, expressed in terms of predictive power, Appendix C4 shows that stereotypically female content has an order of magnitude more power to predict which gender sees the ad than any of these five other characteristics, and six times more than our all four 'hard' job characteristics (wage, education, experience, and firm size) combined.[37]

[37] Appendix C4 runs six univariate regressions of the *Male* dummy on each of the six job characteristics in Tables 1 and 4. These regressions show an *R*-squared of 0.093 for stereotypically female content, compared

The outsized role of stereotypically female content in our results is consistent with findings from psychology and linguistics that male job attributes are widely perceived as 'default' attributes that are expected in most jobs, and that humans tend to focus on departures from defaults to classify objects (Smith and Zarate, 1992).[38] A distinct but related possibility is that NLP algorithms on the boards use TF-IDF (Term Frequency–Inverse Document Frequency) methods to match jobs to a worker's gender. For example, words like *maternity leave* may appear relatively infrequently in jobs, while also being strongly associated with women– thus making those terms especially powerful predictors of which gender sees the ad. Male characteristics are less useful because they are more 'generic'– i.e. likely to appear in a wider variety of job ads. The large gender gaps in female-typed content and female content's relatively high predictive power also have implications for our study of mechanisms in Section 4 below. Essentially, female content gives us much more statistical power to distinguish among mechanisms, and we will rely on this to draw some distinctions there.

Finally, Figure E2.1 (parts e and f) explores heterogeneity in male- and female-stereotypical content gaps by applicant age, by femaleness of the occupation-industry cell, and by position level. Most dramatically, we find that gender gaps in recommended jobs' stereotypically female ad content are greatest for workers who are *seeking* female-dominated job types and in entry-level positions. Similarly, the gender gap in stereotypically male content is greatest in male-dominated job types, but the magnitude is much smaller.

to the next largest (0.008, for education requirements). The R-squared for our four 'hard' job characteristics combined is .014.

[38] 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 less informative of which gender is better matched to the job than female words.

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

Having described the overall differences between the jobs that were recommended to women versus men, we now describe how these differences evolved across the seven sub-rounds (0, 1.1, 1.2, 2.1, etc.) of our experiment. Notably, in this section –and whenever we study recommendation trends across experimental sub-rounds (Section 4.2 and 4.3), we make two sample changes: We expand the sample to include the 'overlapping' ads seen by both our male and female profiles, and we use only the first ten jobs recommended to each worker in Round 0. Including common jobs allows changes in the difference rate across rounds to affect the gender gap in recommended jobs' characteristics. Using only the top 10 jobs in Round 0 ensures that our Round 0 observations are strictly comparable to the observations collected in rounds 1.1, 1.2, etc..[39]

We begin with our most basic measure, the set difference rate; Figure 3 shows that it increases sharply between rounds 0 and 1.1 (from 9.74 to 12.15 percent), then steadily but more slowly to a maximum of 17.60 percent in round 3.2.[40] Figures 4(a) to (f) graph the evolution of gender gaps in job characteristics (including stereotypical content) across sub-rounds; the levels of these gaps are lower than in Tables 1 and 4 because we have expanded the sample to include the 'overlapping' ads seen by both our male and female profiles.

In contrast to the strong difference rate trends, five of the six time trends in recommended jobs' characteristics are statistically insignificant. Consistent with the especially large gender gap in stereotypically female content, this characteristic is the exception: Its negative gender gap (male minus female) increases steadily in absolute value throughout the entire experiment, from just over .04 standard deviations to over .10 standard deviations. Also noteworthy in Figure 4 is the fact that the largest increase in the female content gap occurs very early in our experiment: between rounds 0 and 1.1.

[39] When the context is clear, we refer to sub-rounds simply as 'rounds'.

[40] Both the increase between round 0 and 1.1 and a linear trend over the entire period are highly statistically significant ($p<0.01$).

In sum, Figures 3 and 4 show rising trends in the difference rate and in the gender gap in 'female' job ad content throughout the experiment, with an especially large jump between rounds 0 and 1.1. No other time trends are evident. Section 4 exploits these and other facts to explore the likely algorithmic processes that are operative on the boards, and their role in driving the gender gaps we have documented in this Section.

4 Mechanisms

In this Section, we ask which specific processes and pieces of information are responsible for the gender gaps documented in Section 3. Knowing whether algorithms use sensitive or legally protected information can be useful to consumers, legal advocates, and regulators. Isolating which algorithmic processes accentuate or mitigate gender stereotyping could also be useful to regulators and algorithm designers interested in reducing those gaps.

To isolate mechanisms in our 'outsider' context –where we know relatively little about the processes the boards might be using– our design mimics the actions of (a certain kind of) job seeker over several experimental rounds. Specifically, we focus on a worker who –like our profiles– has just created a profile on a job board and consider three broad classes of potential mechanisms:

- content-based mechanisms (C-mechanisms) use the contents of the new worker's resume to recommend jobs to the new worker.
- recruiter-based mechanisms (R-mechanisms) use recruiters' reactions to the new worker's resume to recommend jobs to the new worker.
- worker-based mechanisms (W-mechanisms) use the new worker's own behavior (e.g. applications) to recommend jobs to the new worker.

In Sections 4.1-4.3, we assess these mechanisms in the order in which they become available to recommender systems during the course of our experiment. For each class of mechanisms, we first describe several more detailed forms ('channels') it can take. Then we discuss the evidence indicating (a) whether each type of mechanism is operative, and (b) whether it contributes to the gender-recommendation gaps documented in Section 3. As in Section 3, we group all four of our job boards together for these analyses. Appendix D replicates this section's main results separately for each job board; with one exception described in Section 4.3, our findings are very similar for all four job boards.

Most of the algorithmic processes studied in this Section are well known and have been described in many survey articles (Al-Otaibi and Ykhlef, 2012; Hong et al., 2013; Siting et al., 2012). Because of our experimental design, however, not all these processes are available to job recommender systems in all rounds of our experiment. Our contribution in this Section is to characterize the complete set of mechanisms (including less likely ones) that are available to job recommender systems in different rounds of our experiment, and the likely empirical footprints of each mechanism if it is present. This additional analysis is what allows us to draw conclusions about algorithmic processes from the data we have collected.

4.1 Content-Based Mechanisms (*C*-Mechanisms)

Consider Worker A, who has just published their profile on a job board. An algorithm charged with recommending 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 A-specific information available to the algorithm is the contents of A's resume, an essential first step in all possible processes is to match the content of A's resume with other available content, such as available job ads or other workers' resumes. There are three broad ways the algorithm could proceed, 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), either by rules-based matching of coded content (e.g., the worker's education must meet the job's requirements) or NLP-based similarity scores between the unstructured text of job ads and resumes.[41] Importantly, both these resume-job matching methods can cause gender differences in mean recommended job characteristics in Round 0 *only if* the matching methods use the worker's gender as an input.[42] This is because our male and female profiles are identical except for gender.

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). Even though these Channels (unlike Channel 1) use some information that is not content-based (such as the actions of Worker A's co-applicants) Channels 2 and 3 (like Channel 1) can only yield systematic gender gaps if their initial (worker-worker) content-matching process uses Worker A's gender as an input. In short, combining all three possible Channels, our experimental design allows us to test for the existence of C-mechanisms on a platform, and for whether the worker's gender is explicitly used by those mechanisms.

To conduct this test, Table 5 replicates Tables 1 and 4 using data from Round 0 only. Overall, the Round 0 results are strikingly similar to the full-sample results, given that Round 0 recommendations only comprise about 15 percent of Table 1 and 4's estimation sample.[43] While the gender wage gap in Table 5 becomes statistically insignificant

[41] Worker-job similarity scores can be computed in two main ways: *Unsupervised* learning that looks for similarities in content based on a large corpus of documents, or *supervised* learning where the algorithm searches for worker-job matches that best predict an outcome (such as workers' application probabilities, recruiters' clicks or other measures of engagement) observed in historical interaction data. Our experimental design cannot distinguish between these training choices.

[42] Lack of overlap (i.e. a positive set difference) between jobs recommended to our male and female profiles does not imply that gender is being used by the algorithm. Such differences can be caused by quasi-randomness (e.g. differences in the availability of jobs due to sequential processing of recommendations) and commonly-used decongestion and diversification processes, which 'spread out' good matches across applications and vacancies to improve user satisfaction (Szpektor et al., 2013; Wu et al., 2016; Kunaver and Požrl, 2017).

[43] Using the last column of Tables 4 and 5, the Round 0 share is 3,289/22,023 = 14.9%. While (by design)

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 gender differences in both stereotypically male and stereotypically female ad content. In fact, comparing Table 5's Round 0 gaps to Table 1 and 4's experiment-wide gaps indicates that the Round 0 gaps can account for essentially all of the experiment-wide gaps we measure.[44]

In sum, our analysis of gender gaps in Round 0 yields four 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 C-mechanisms. 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. Third, the C-mechanisms detected in Round 0 can account for essentially all the gender gaps in job content we measure in our experiment. Finally, these content-based processes must be using the worker's declared gender as an input. The latter finding illustrates how our algorithm audit design can reveal whether black-box recommender systems use *any* particular piece of a worker's resume –including protected characteristics– to suggest jobs to the worker. Our design accomplishes this by taking advantage of a well known challenge facing recommender systems –the cold start problem, which essentially starves algorithms of behavior-based information about users– allowing us to cleanly identify which components of the remaining data (i.e. the worker's profile) are being used.

the 20 recommendations collected in Round 0 comprise 25 percent of all recommendations collected, gender-specific recommendations are much less prevalent in Round 0, accounting for this 15 percent share.

[44] Dividing the Round 0 gender gaps from Table 5 by experiment-wide gender gaps from Tables 1 and 4, yields ratios of 0.766, 1.026, 1.178, 0.954 and 0.882 for wages, experience, firm size, stereotypically female content, and male content respectively. We exclude education gaps because they are not precisely estimated in either period.

4.2 *R*-Mechanisms (based on Recruiters' Reactions to Worker A's resume)

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 Round 0 and round 1.1).[45] A direct and natural way for algorithms to exploit recruiters' resume search activity is to encourage workers to apply to recruiters who have already viewed, clicked, or downloaded their resumes. In round 1.1, *R*-Mechanisms can also use job-job similarity scores to direct workers to jobs that *resemble* the jobs that 'found' the worker. Finally, starting in round 2.1 (after our profiles have submitted their first batch of applications) *R*-Mechanisms can also exploit recruiters' reactions to those applications during the preceding two-week Interval (Channel 5).

To assess whether *R*-Mechanisms are *active* on our job boards, we focus on the difference rate between recommendations received by men versus women and first note its sharp increase between Round 0 and round 1.1 in Figure 3: This increase is larger than those between any other sub-rounds of our experiment and highly statistically significant. Since *R*-Mechanisms are the only new mechanism available during this Interval, this suggests that recruiters' reactions to our published resumes (finding, viewing, or clicking some, and not others) are working to reduce the overlap between recommendations received by our male and female profiles. Second, and consistent with a substantial role for *R*-mechanisms during Interval 1, Appendix Table F4 shows that the mean number of pro-

[45] 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. We do, of course, expect round 1.1 recommendations, like Round 0 recommendations, to also use *C*-mechanisms to match the worker's resume to newly-available jobs. But since the resume is a static document and we do not expect the flow of new jobs to be larger or different from the Round 0 jobs, it seems unlikely that *C*-mechanisms can cause a larger gender gap in round 1.1 than Round 0. Accordingly, we attribute any rising gaps between Round 0 and round 1.1 to the newly-available *R*-mechanisms in round 1.1. See Appendix E8 for an expanded discussion of this issue.

file views per pair (15.8) is higher during that Interval than any of the subsequent two-week Intervals: Unsurprisingly, recruiters pay more attention to newly-posted resumes than older ones.

Finally, column 1 of Table 6 pursues the intuition that –if *R*-Mechanisms are active– the recommendations received by our paired male and female resumes should exhibit greater difference rates among pairs whose resumes have been viewed more often. In more detail, the observations in Table 6 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 a two-week Interval (for example the two-week period between rounds 0 and 1.1) and the outcomes are the pair's set difference rate at the start of the next round (round 1.1). According to Table 6, more resume views during Interval 1 are strongly associated with a larger set difference rate when our profiles first log on after that Interval. One more profile view is associated with .0961 more gender-specific jobs per 100 recommendations, an elasticity of about 0.20. Similar correlations are observed after Intervals 2 and 3, with elasticities of 0.09 and 0.10. Together, we view these three pieces of evidence as strongly suggestive that our boards *use* *R*-mechanisms.[46]

While job boards' use of *recruiters'* revealed preferences to recommend jobs to workers may not seem surprising, we note that the preferences of the 'other' side of the market do not play a role in many widely used recommender systems. For example, Amazon's consumer recommendation algorithms do not consider whether a particular good (such as a pair of socks) prefers to be purchased by consumer A or consumer B, or which customers a seller would prefer to serve. The use of *R*-mechanisms thus signals that our job boards

[46] We describe this result as suggestive because the positive association between profile views and difference rates could also result from unobserved aspects of profile pairs that affect both profile views and difference rates. For example, profile pairs in tight labor markets (many vacancies per job seeker) should receive more profile views; we would also expect their male and job recommendations to overlap less since there is a larger pool of good matches to recommend. Appendix Tables F6 and F7 explore this possibility with tighter controls for labor market tightness and by asking whether the total number of 'available' jobs to a resume affects difference rates in the top tier of recommendations. Together these Tables suggest that unobserved variation in market tightness cannot account for all the correlation between resume views and subsequent difference rates. Appendix F also shows that the association between profile views and difference rates is robust to other specification changes, such as disaggregating the resume view effects by profile gender.

have not simply borrowed standard customer-search tools such as item-based collaborative filtering from retail markets (e.g. Hensvik et al. (2025)).

To assess whether *R*-Mechanisms contribute to gender gaps in recommended jobs' *characteristics*, we first refer to Figure 4, which shows no increase in the gender gaps for wages, education, experience, firm size, or stereotypically male content between Round 0 and round 1.1. There is, however, an almost-significant increase in the gender gap in stereotypically female content (p=.105), which is larger than between any other adjacent sub-rounds. Also suggestive of a role for *R*-mechanisms, column 6 of Table 6 shows a significant positive correlation between a pair's number of resume views in all three experimental Intervals and the subsequent gender gap in stereotypically female content. Furthermore, replicating Table 6 for recruiters' views of male and female profiles separately, Tables F1 and F2 show that all of this effect is driven by recruiters' views of *female* profiles. This is consistent with the idea that the gender gaps we observe in later rounds of the experiment result, in part, from the fact that recruiters in 'female-friendly' jobs show more interest in female resumes. For example, recruiters might disproportionately click on female resumes in jobs that have flexible hours or that mention assisting others.[47]

4.3 Worker-Based Mechanisms (*W*-Mechanisms)

A final possible class of mechanisms is based on the worker's own past actions. Since our profiles' first actions (other than posting their resumes) are the applications they submit in round 1.1, our first chance to observe those actions' effects is in round 1.2. One such process (Channel 6, 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.[48]

[47] On some Chinese job boards, this process can be explicitly supported by a feature that lets job ad creators working with multiple recruiters specify a targeted gender for the job that the recruiter sees. This 'guidance' to recruiters remains legal, even while explicit gender requests in posted job ads have been effectively banned (Kuhn and Shen, 2023). Our informal conversations with recruiters suggest this feature is rarely used, however.

[48] 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.

The other main type of W-mechanism (Channel 7) uses data on Worker A's co-applicants at the jobs she has applied to. This process is called Item-Based Collaborative Filtering (IBCF), or "Workers Who Applied to this Job Also Applied to...".[49] For real workers –especially those who found the board's Round 0 suggestions unhelpful – both these channels provide an opportunity to 'teach' the board's algorithms to accommodate their individual preferences. In our experiment –since the profiles in our experiment always follow the board's recommendations– we would expect Channels 6 and 7, if they are operative, to mechanically magnify any gender gaps in recommended jobs' characteristics that were present in preceding experimental rounds.[50]

To assess whether W-mechanisms are affecting the jobs displayed to workers, we return to Figure 4, which shows no significant trends across rounds in gender gaps for five of the six job characteristics we measure after round 1.1. The exception, again, is for stereotypically female content, which grows from .07 standard deviations to over .10 standard deviations (p=.009). This monotonically growing gap is consistent with a process where a combination of C- and R-mechanisms creates a gender gap in stereotypically female content in round 1.1's recommended jobs. Then, starting in round 1.2, W-mechanisms magnify these gaps as our profiles keep applying to the jobs that are recommended to them.[51] Finally, we note from Appendix D that the time trend in the stereotypically-female content gap is driven disproportionately by job boards 3 and 4, which serve a considerably more skilled workforce than boards 1 and 2. This is the only evidence we find that suggests a difference in algorithmic processes between some of our four job boards. Specifically,

[49] 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 of a worker's resume or job ad; it is based purely on jobseekers' behavior, inferring job similarity from co-application patterns (Jannach et al., 2016).

[50] Recall, however, that this test only reveals the presence of W-mechanisms that operate *between* workers' logon sessions (see our results discussion in Section 1). We know that W-mechanisms are present *within* sessions since all the boards suggest similar jobs to workers right after they apply to a job.

[51] Since R-mechanisms could also be contributing to the growing gender gap in stereotypically female job content after round 1.1, Appendix G provides a cleaner test for the presence of W-mechanisms by focusing only on the growth of gender gaps that occur between the rounds of our experiment which are only a few seconds apart, like rounds 1.1 and 1.2. This leaves essentially no time for recruiters to react to the last round's applications. While the difference rate grows significantly over those intervals (p=.000, Table G1), the gender gap in recommended job characteristics (a clearer test for the presence of W-mechanisms) does not grow significantly for any of our six job characteristics in Table G1, though our statistical power is admittedly weak.

it suggests that across-logon W-mechanisms (recommending jobs that are similar to ones you viewed last time you logged on) are only offered on the higher-skill job boards, where workers' job searches may be more likely to continue across several logon sessions.

5 Discussion

Using an algorithm audit, we have measured the size, content, and direction of gender biases in the recommender systems of China's four largest job boards, and identified the main algorithmic mechanisms driving those biases. We find modest amounts of bias, where female profiles are shown jobs that pay less, require less experience, and are in smaller firms. Jobs recommended to women also contain less stereotypically male language and more stereotypically female language than jobs recommended to men. We attribute the modest size of the gender gaps in our study, in part, to the fact that we measure gaps between identically-qualified resumes with identical job search goals; this is a unique feature of our audit design compared to previous audits of job ad placement. Another contributing factor may be our focus on job boards (in contrast to general purpose social networking sites), which specialize in worker-firm matching and have a clear economic incentive to perform this single function well.

We identify the main cause of the above gaps as *content*-based algorithms, which match the content of the worker's resume with available job ads; we also show that these content-based processes must be using the worker's gender as an input. This implies that eliminating gender as an input, or applying de-biasing methods like adversarial learning (e.g. Rus et al. (2022) and Bied et al. (2023b)) to the boards' internal representations of resumes and jobs would likely reduce gender gaps. We also find some evidence suggesting that recruiters' reactions to workers' resumes channel women into stereotypically female jobs, and that algorithms that base recommendations on workers' past applications cause gender gaps to widen over time when workers follow a board's recommendations.

While our analysis has focused entirely on establishing facts about whether and why certain proprietary black-box recommender systems treat male versus female resumes differently, we conclude by raising the normative question of what, if any, is the optimal amount of bias in a job recommender system. To provide context for this question we remind readers that –despite having identical qualifications and stated job preferences– our fictitious male and female worker profiles do not specify any preferences for job characteristics like flexible hours, daytime hours, and job travel demands. In consequence, the recommender systems we study are essentially using the worker's gender to 'guess' the probability the applicant wants those amenities. Making this inference is helpful to women with median (or stereotypical) work schedule preferences because it quickly identifies jobs they are likely to want. At the same time, it violates the principle of equal treatment of identical resumes and makes it harder for workers with atypical preferences to find good matches. One way to escape this dilemma could be for job boards to solicit more information from workers about their job search goals, thereby serving workers with both typical and non-typical preferences better.

References

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

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.

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

Barocas, S., Hardt, M. and Narayanan, A. (2023), *Fairness and machine learning: Limitations and opportunities*, MIT press.

Belot, M., Kircher, P. and Muller, P. (2019), 'Providing advice to jobseekers at low cost: An experimental study on online advice', *The Review of Economic Studies* **86**(4), 1411–1447.

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.

Bertrand, M. and Mullainathan, S. (2004), 'Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination', *American economic review* **94**(4), 991–1013.

Bied, G., Caillou, P., Crépon, B., Gaillac, C., Pérennes, E. and Sebag, M. (2023a), Designing labor market recommender systems: How to improve human-based search, Technical report, Working paper, CREST.

Bied, G., Gaillac, C., Hoffmann, M., Caillou, P., Crépon, B., Nathan, S. and Sebag, M. (2023b), Fairness in job recommendations: estimating, explaining, and reducing gender gaps, *in* 'AEQUITAS 2023-First AEQUITAS Workshop on Fairness and Bias in AI| co-located with ECAI 2023', Vol. 3523, CEUR-WS. org.

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

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

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.

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.

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

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.

Dressel, J. and Farid, H. (2018), 'The accuracy, fairness, and limits of predicting recidivism', *Science advances* **4**(1), eaao5580.

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

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.

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.

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.

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.

Hensvik, L., Barbanchon, T. L. and Rathelot, R. (2025), Recommender systems and the labor market, Technical report, Working paper.

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

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

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

Imana, B., Korolova, A. and Heidemann, J. (2021), Auditing for discrimination in algorithms delivering job ads, *in* 'Proceedings of the web conference 2021', pp. 3767–3778.

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

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.

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.

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.

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

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.

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

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

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., Sandvig, C. et al. (2021), 'Auditing algorithms: Understanding algorithmic systems from the outside in', *Foundations and Trends® in Human–Computer Interaction* **14**(4), 272–344.

Miceli, M., Posada, J. and Yang, T. (2022), 'Studying up machine learning data: Why talk about bias when we mean power?', *Proceedings of the ACM on Human-Computer Interaction* **6**(GROUP), 1–14.

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*

OpenAI (2023), 'ChatGPT (gpt-4, march 14 version)'. Large language model. Response to a user query on 2023-11-05.
URL: *https://chat.openai.com/*

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

Peña, A., Serna, I., Morales, A., Fierrez, J., Ortega, A., Herrarte, A., Alcantara, M. and Ortega-Garcia, J. (2023), 'Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment', *SN Computer Science* **4**(5), 434.

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

Prates, M. O., Avelar, P. H. and Lamb, L. C. (2020), 'Assessing gender bias in machine translation: a case study with google translate', *Neural Computing and Applications* **32**(10), 6363–6381.

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.

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.

Rus, C., Luppes, J., Oosterhuis, H. and Schoenmacker, G. H. (2022), 'Closing the gender wage gap: Adversarial fairness in job recommendation', *arXiv preprint arXiv:2209.09592* .

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.

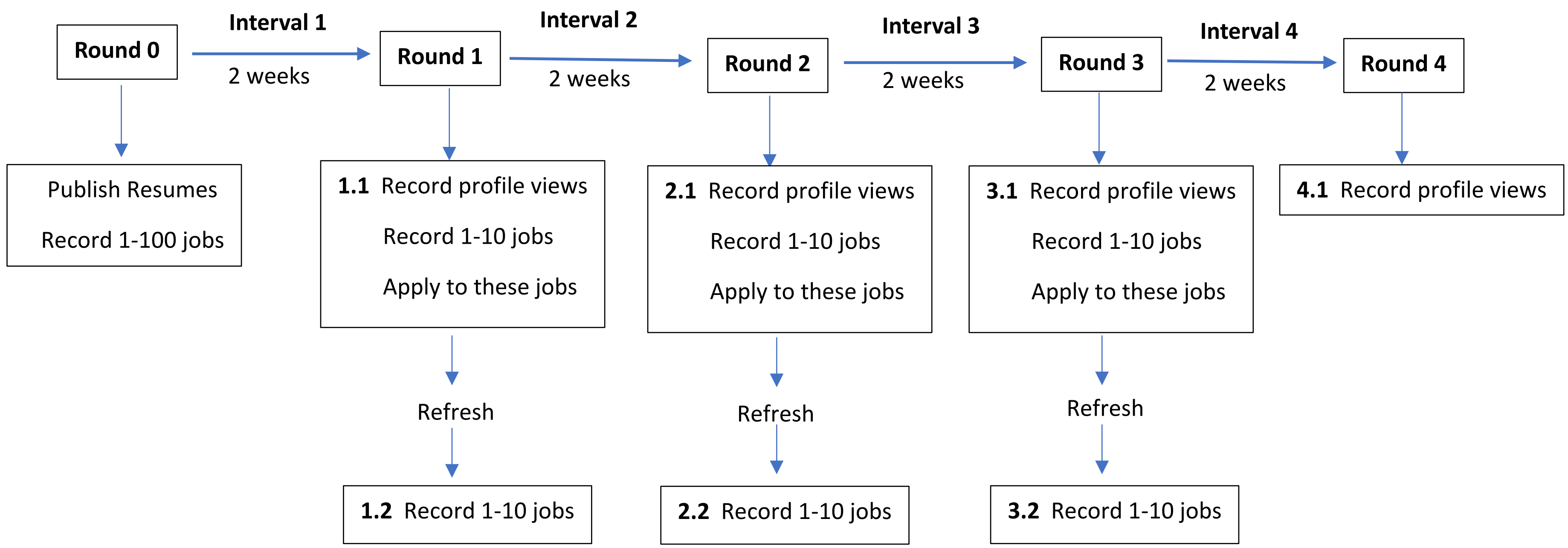
Szpektor, I., Maarek, Y. and Pelleg, D. (2013), When relevance is not enough: Promoting diversity and freshness in personalized question recommendation, *in* 'Proceedings of the 22nd international conference on World Wide Web', pp. 1249–1260.

Tamkin, A., Askell, A., Lovitt, L., Durmus, E., Joseph, N., Kravec, S., Nguyen, K., Kaplan, J. and Ganguli, D. (2023), 'Evaluating and mitigating discrimination in language model decisions', *arXiv preprint arXiv:2312.03689* .

Vlasceanu, M. and Amodio, D. M. (2022), 'Propagation of societal gender inequality by internet search algorithms', *Proceedings of the National Academy of Sciences* **119**(29), e2204529119.

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



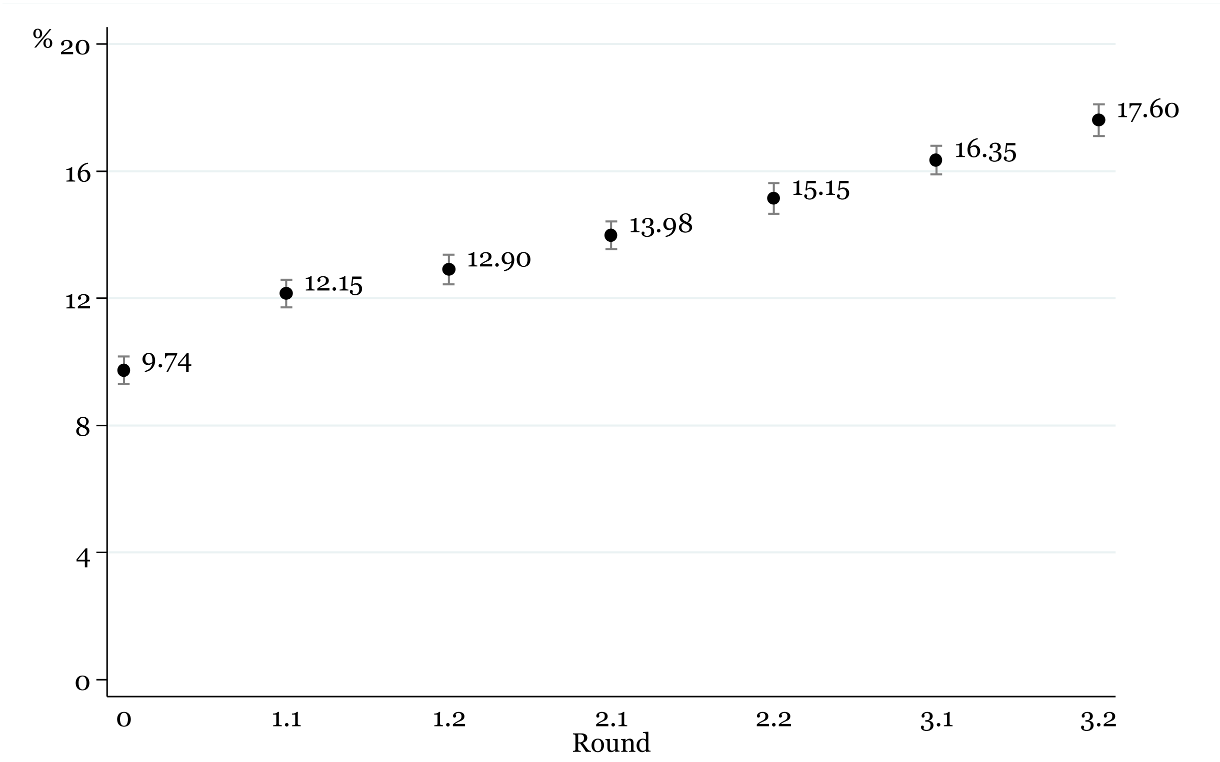
Notes:

1. The two profiles in each gender pair follow the same timeline.
2. In sub-rounds 1.1, 2.1, and 3.1, fictitious workers apply to the top 10 jobs in their customized list of job recommendations.
3. When the context is clear, we refer to sub-rounds 1.1, 1.2, etc. simply as 'rounds' in the paper.
4. When generating results to represent the entire experiment (e.g. Sections 3.1-3.3) we use only the top 20 recommendations in Round 0.
5. When characterizing trends across sub-rounds of the experiment (e.g. Section 3.4), we use only the top 10 recommendations in Round 0.

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 C Figure C1.1.

Figure 3: Set Difference Rate by Experimental Rounds



p-value for zero slope (Rounds 0 – 3.2) = 0.000, with coefficient on round indicator = 1.222, N = 7,746.

p-value for zero slope (Rounds 1.1 – 3.2) = 0.000, with coefficient on round indicator = 1.107, N = 6,626.

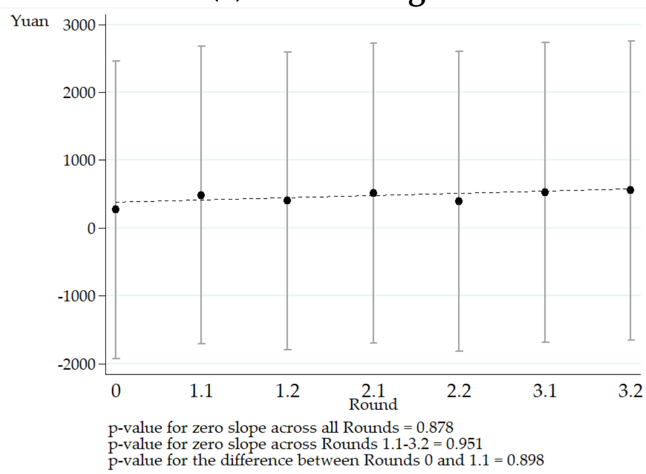
p-value for Round 0 = Round 1.1 = 0.000, with coefficient on round indicator = 2.410, N = 2,238.

Notes:

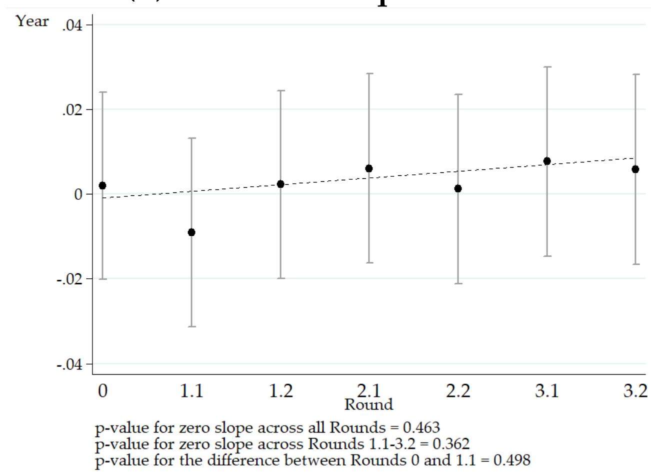
1. The Round indicator used in the slope regressions equals 0,1,2,3,4,5 and 6 for rounds 0, 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 respectively.
2. To ensure comparability across Rounds, the sample for Round 0 includes only the first 10 ads in each worker's recommendation list. 10 recommendations are collected in Rounds 1.1, 1.2, etc.
3. Whiskers show 95% confidence intervals.

Figure 4: Gender Differences in Job Characteristics by Experimental Rounds

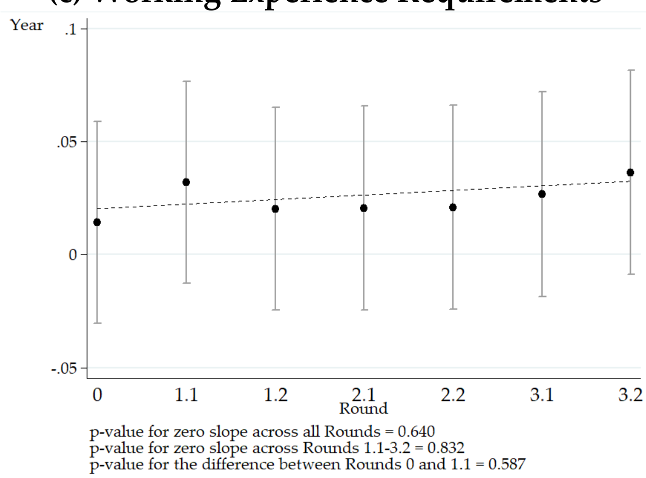
(a) Posted Wage



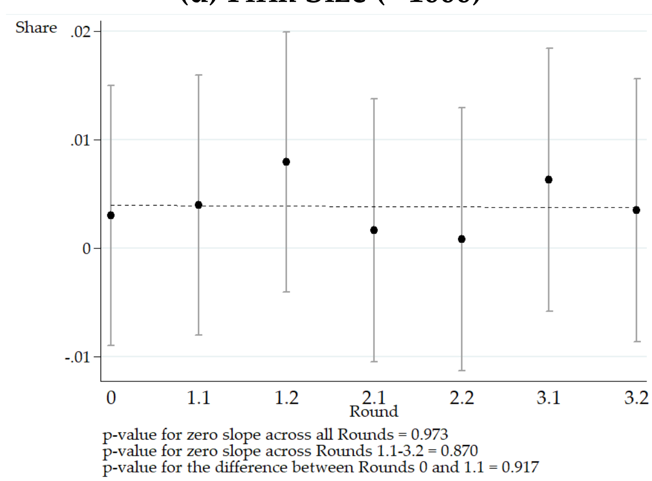
(b) Education Requirements



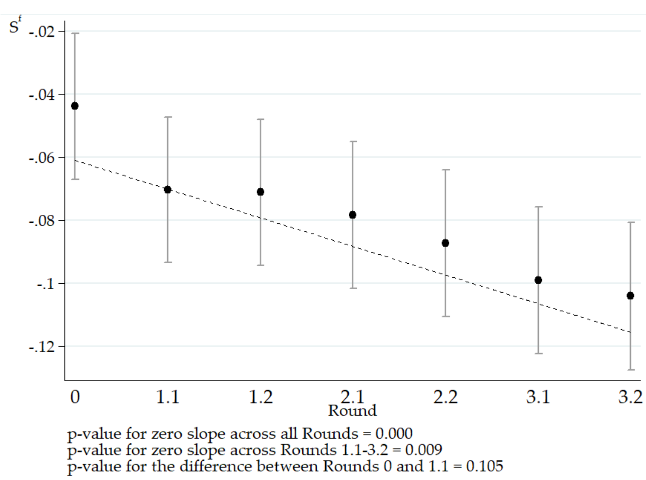
(c) Working Experience Requirements



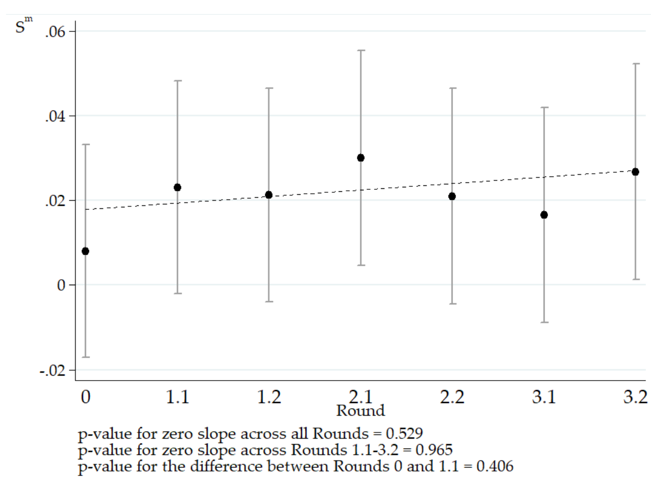
(d) Firm Size (≥1000)



(e) Stereotypically Female Content (S^f)



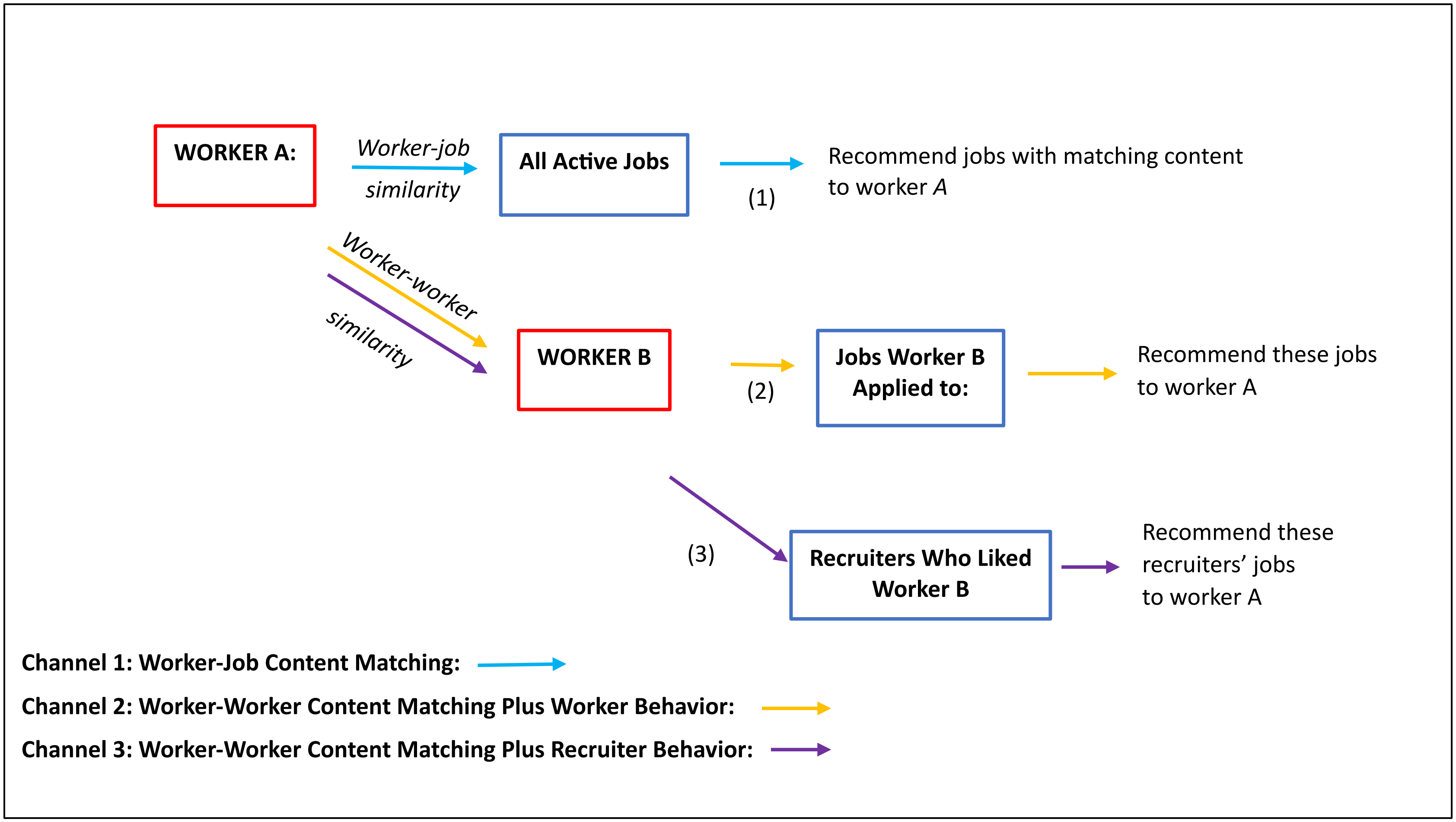
(f) Stereotypically Male Content (S^m)



Notes:

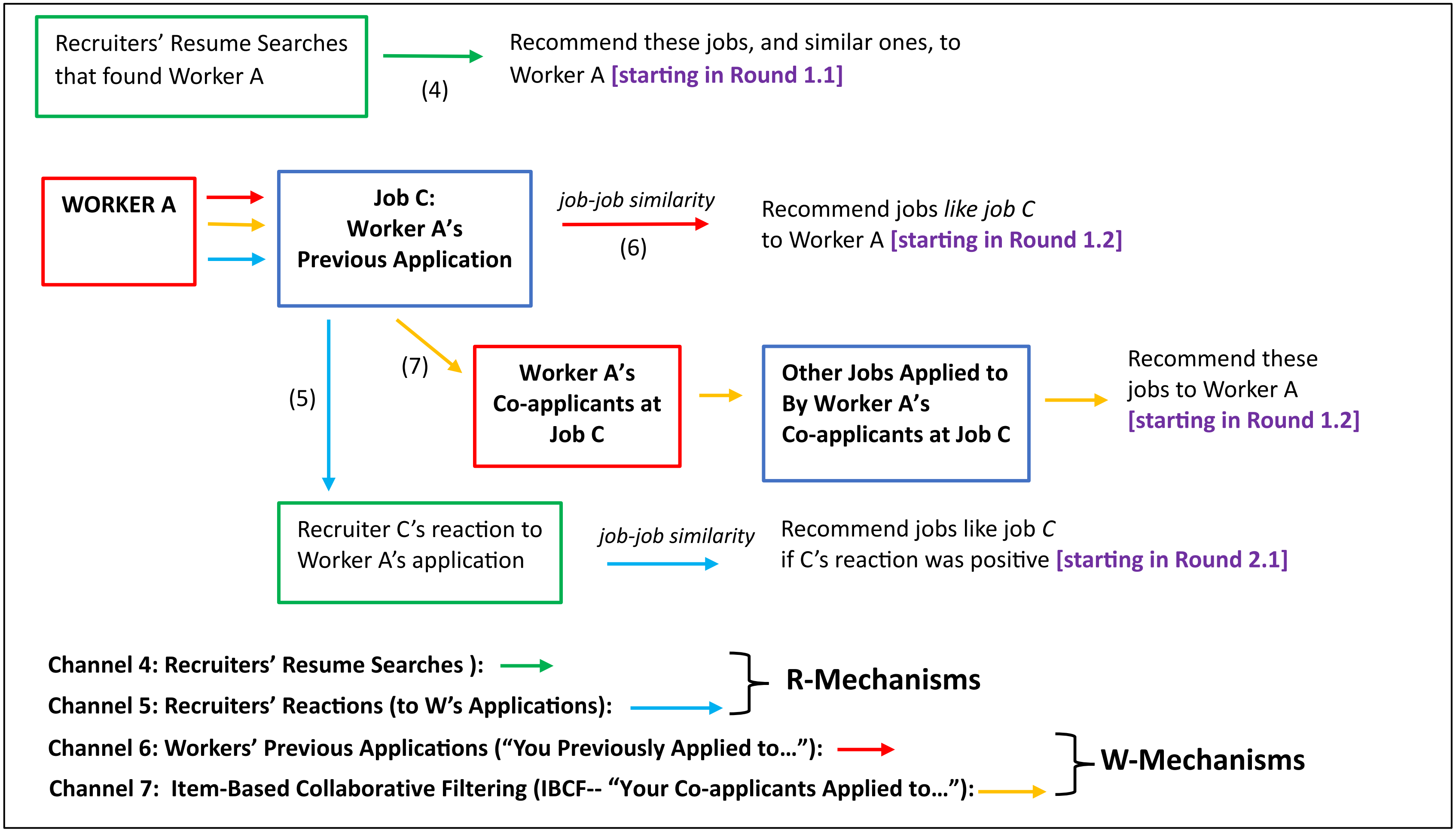
1. For consistency across Rounds, the sample for Figure 4 uses the top 10 recommendations in all Rounds of the experiment: 0, 1.1, 1.2,…, 3.2. Also, to allow changes in the share of overlapping jobs to affect gender gaps in their characteristics, we expand our sample to include all recommended jobs, not just the non-overlapping ones.
2. Figure 4 shows the differences in job characteristics (and their 95% confidence intervals) between the top ten jobs recommended to men and the top ten jobs recommended to women in each Round of the experiment. These estimates come from a regression of a job's characteristic (e.g. wage) on Round fixed effects, a dummy for Male profiles, and its interaction with Round fixed effects.
3. To calculate the trend line in each graph, we replaced the Round fixed effects with a linear Round term (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). Each graph reports the *p*-value for the interaction between Male and this linear Round variable.
4. The graphs also report *p*-values for a zero slope excluding Round 0, and for a zero difference between Rounds 0 and 1.1.

Figure 5: Content-Based Recommendation Mechanisms (C-Mechanisms)



C-Mechanisms are available to job boards throughout the experiment; they are the only mechanisms available in Round 0. 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: Action-Based Recommendation Mechanisms (*R*- and *W*-Mechanisms)



R- and C-Mechanisms become available after Round 0 (as indicated). They use past actions of the focal worker (W) and the recruiters he or she encounters (R) to recommend jobs to W. 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 6 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***	0.0256***
	(1,023)	(0.011)	(0.022)	(0.006)
N	21,262	19,900	21,922	21,949
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 Rounds 1-3 plus the top 20 recommendations from Round 0. 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. Regressions 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, 37.05% of jobs in this regression sample were in firms with 1000 or more workers.
4. All regressions control for profile pair fixed effects.
5. Relative to sample means, the wage, experience, and firm size effects indicate that male-only jobs pay (3,118/202,453=) 1.54 percent higher wages, require 7.19 percent more experience, and are 7.16 percent more likely to be in firms with 1000 or more employees. Including the 87.6 percent of recommended jobs that are shared by men and women, these gaps shrink to 0.2 percent for wages, 0.9 percent for experience, and 1.0 percent for firm size.

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. Column 2 indicates that male-only ads contained .1322 standard deviations more stereotypically male content than female-only ads
3. Including the 87.6 percent of recommended jobs that are shared by men and women, these gaps shrink to 0.071 standard deviations for female content, and .016 standard deviations for male content.

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***	0.0303*	-0.5493***	0.1166***
	(2,502)	(0.028)	(0.058)	(0.016)	(0.032)	(0.034)
N	3,177	2,934	3,278	3,278	3,289	3,289
R^2	0.741	0.681	0.559	0.406	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, 39.48% of jobs in this regression sample were in firms with 1000 or more workers.
4. All columns control for pair fixed effects.

Table 6: 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 (S^f)	(7) Stereotypically Male Content (S^m)
A. Interval 1							
Views	0.0961***	-28	0.0007*	0.0005	-0.0001	-0.0009**	-0.0007
	(0.017)	(32)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
N	1,118	1,118	1,118	1,118	1,118	1,118	1,118
R^2	0.323	0.003	0.010	0.011	0.013	0.075	0.058
B. Interval 2							
Views	0.0741***	90**	0.0000	-0.0003	0.0002	-0.0008*	-0.0004
	(0.018)	(37)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
N	1,100	1,100	1,100	1,100	1,100	1,100	1,100
R^2	0.346	0.014	0.005	0.003	0.008	0.057	0.041
C. Interval 3							
Views	0.1037***	42	-0.0003	0.0009	0.0003	-0.0010*	0.0008
	(0.019)	(36)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.308	0.007	0.003	0.013	0.021	0.079	0.037

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. While we have 1,120 gender pairs in the experiment, the gender gap in outcomes is missing in some pairs and periods.
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.

Appendix to:

Measuring Bias in Job Recommender Systems:

Auditing the Algorithms

Shuo Zhang, Northeastern University

Peter Kuhn, UC Santa Barbara

For online publication only

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 Clerks	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-A1.4, 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] 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

B1: Largest Job Boards in 16 Countries*

In this Section we list and describe the largest job boards in 16 countries, ranked by estimated web traffic. Our country list is the union of the countries with the ten largest economies (highest nominal gdp) and the ten largest populations. Our source for the country list is Wikipedia searches conducted on September 30, 2025. Our main goal is to get a rough estimate of the relative importance of private versus publicly operated job boards in these labor markets.

Our sources for job boards' web traffic and rankings are ChatGPT's responses to the following prompt:

> What are the ten largest job search platforms in [country], ranked by web traffic? Please include government-run platforms in your search if they are large enough, and provide a rough estimate of average monthly traffic for each one. Identify each platform you list as either private or public.
>
> For classified ad sites, please include only traffic related to job search. Estimate this if necessary.
>
> For multi-purpose sites like LinkedIn and Glassdoor, please include only traffic related to job search. Estimate this if necessary.
>
> For international sites like Indeed and Monster.com, please include only traffic to the country-specific portal. Estimate this if necessary.
>
> Make your response very predictable, focused, and factual by setting the temperature to 0.2.

In some cases, ChatGPT's first response to the above queries used crude estimates of the 'jobs' share of classified ad sites like 58.com in China and the job search share of multi-purpose sites like LinkedIn (e.g. just using total traffic). In those cases, we asked for additional research to estimate these shares and the assumptions used. Conservative assumptions were used where possible. These searches were performed between October 21 and 22, 2025. The complete search results (including ChatGPT's text responses and ranking tables for all countries) are accessible at the link provided below. Given the assumptions required to produce these estimates and the challenges in replicating current AI results, we view them as providing only a rough impression of the relative importance of public versus private boards.

*This appendix was prepared with the assistance of Xinlu Zhao, UC Santa Barbara.

Key features of our results are summarized in Table B1.1, which shows that these sixteen countries fall into three distinct groups. The first group consists of three countries (Germany, France, and Pakistan) whose largest job board is a publicly operated one, attracting between 33 and 36 percent of the top ten boards' total traffic. At the other extreme, six countries have no publicly operated boards in the top ten by traffic. This occurs both in developed economies (Japan and Italy) but is more common in populous, less developed nations (India, Indonesia, Nigeria, and Bangladesh). In some of these cases, the largest board is a country-specific private provider, such as Naukri, JobStreet, and BDjobs in India, Indonesia, and Bangladesh; in other cases like Japan and Italy, Indeed is the largest provider.

Finally, a third, intermediate group of six countries have a single publicly operated board in their top ten, ranked between fourth and ninth in the country. In all these cases the public board has less than 8 percent of the total traffic on its country's top ten boards. Combining all 16 countries, only 13 out of 160 top-ten job boards are publicly operated; a simple mean of their country-level market shares is 9.6 percent. The population-weighted mean is 3.4 percent, reflecting the fact that more populous countries tend to rely less on public job boards. Overall, we conclude that a large majority of the world's online jobseekers are served by privately operated job boards.

Link for complete search results:

https://drive.google.com/drive/folders/1b1BIl1S4TwI25EOzPdwXwneSEZCYxL-L?usp=sharing

Table B1.1: Market Shares of Each Country's Top Ten Job Boards

A. Ten Largest Countries by GDP:

Country	Top-ranked board and market share (among the top ten)	Ranks and market share of public boards
USA	Indeed (58.9%)	6 (USAjobs) (1.3%)
China	Zhipin (41.1%)	9 (MOHRSS) (1.3%)
Germany	Jobbörse (35.9%)	1 (Jobbörse) (36%)
India	Naukri (34.0%)	-
Japan	Indeed (22.6%)	-
UK	Indeed (52.2%)	4 (jobs.nhs.uk) (7.0%)
France	Francetravail (35.1%)	1 (francetravail.fr) (35.1%) 4 (urssa.fr) (11.4%) 6 (moncompteformation) (4.1%)
Italy	Indeed (36.1%)	-
Canada	Indeed (52.7%)	4 (Job Bank) (6.1%)
Brazil	Gupy (30.1%)	7 (EmpregaBrasil) (5.2%)

B. Additional Countries in the Ten Most Populous:

Country	Top-ranked board and market share (among the top ten)	Rank, name, and market share of public boards
Indonesia	JobStreet Indonesia (44.7%)	-
Pakistan	PPSC (33.9%)*	1 (PPSC) (33.9%) 7 Punjab Job Portal (3.9%)
Nigeria	LinkedIn (20.8%)	-
Bangladesh	BDjobs (37.3%)	-
Russia	Headhunter (45.4%)	4 (TrudVsem) (6.1%)
Mexico	Indeed (43.1%)	6 (Portal del Empleo / SNE) (2.0%)

*Punjab Public Service Commission

B2: Sample Means and Difference Rates

Table B2.1: Sample Means, Worker Profiles

	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)	(44,275)	(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 B2.2: Sample Means, Recommended Jobs

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	174,319	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	15.6675
	(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)	0.3571	0.2369	0.3178	0.4326	0.4331
	(0.479)	(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)	0.3574	0.3529	0.3575	0.3836
	(0.479)	(0.478)	(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 B2.3: 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. This table 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 B2.4: Heterogeneity in the 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)	0.0099***
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.

B3: Occupation Matches of Recommended Jobs

On all four job platforms, workers select from a set of detailed occupation and industry categories to define the jobs they are seeking, and recruiting agents use these same categories when creating a job posting they wish to fill. Unfortunately, however, the occupation/industry categories selected by recruiters are not displayed in the job ads that workers see. Thus, we cannot compute the exact match between a worker's desired industry or occupation and the job recommendations they receive. The only occupation-related signal visible to job seekers is the job title, which is entered as free text by recruiters and usually contains occupation information. We therefore use similarity scores between the text of the recommended job's title and the worker's desired occupation to measure the amount of occupational matching.

To generate matching rates that are representative of all our four job boards, we create a harmonized set of 40 occupation categories across all the boards. In many cases this was straightforward because the occupations had identical or very similar names across the four boards. For the remainder, we merged the categories manually based on occupation name similarity. The precise crosswalk is provided in Table B3.1. Finally, we quantify how much information a recommended job title conveys about a worker's desired occupation using four increasingly inclusive matching rules and one measure of semantic similarity, as follows:

1. Exact Match:

- The job title must exactly match the worker's occupation in text.
- Examples: "sales manager" == "sales manager"
  "software engineer" == "software engineer"

2. Inclusion Match:

- The job title must contain the full term of the worker's occupation (occupation ⊂ title)
- Examples: "software engineer" == "senior software engineer"
  "product manager" == "e-commerce product manager"

3. Any Information Match:

The job title must contain at least one keyword from the worker's occupation. The occupation is split into multiple keywords (or "information") using word segmentation, with manual adjustments for accuracy. A match is counted if the title contains at least one keyword from the occupation.

- Examples: "sales manager" → "sales", "manager"
  "human resources specialist/assistant" → "human resources", "specialist", "assistant"

4. Most Information Match:

- The job title must contain at least 50% of the total number of keywords from the worker's occupation.
- Example: "software engineer" == "senior software engineer" (2/3 keywords match)
  "sales manager" == "sales representative" (1/2 keywords match)

5. Word similarity:

We compute semantic similarity between the job title and the worker's occupation using the spaCy NLP library. spaCy returns a similarity score from 0 to 100 derived from distributional word embeddings.

Match rates and semantic similarity scores between workers' desired occupations and the titles of the jobs recommended to them are provided in Table B3.2, separately for each desired occupation and for all desired occupations combined. Average match rates using the strictest criteria (exact and inclusion matches) are 15.8% and 32.8%, but the more flexible measures (inclusion and any information) show high rates of correspondence (81.6% and 70.4%) between the top titles recommended to workers and the workers' desired occupations. The average similarity score between the titles of recommended jobs and workers' desired occupations is 70.8.

To assess whether this level of similarity exceeds what would occur with random recommendations, we created a counterfactual as follows. For each of the 177,320 recommended job titles, we randomly assigned it to one of the 39 other desired occupations (uniformly, with probability 1/39 each). The mean occupation-title similarity in this sample of 177,320*39= 6,915,480 random matches was 38.82. A *t*-test of the difference between these two means yielded a *t*-statistic of 694.053, with $p = 0.000$.

- Within occupation sample: N_1= 177,320 , Average Similarity_1 = 70.781, SE_1 = 0.0549
- Across occupation sample: N_2 = 6,915,480, Average Similarity_2 = 38.818, SE_2 = 0.0072
- Difference (Similarity_1 - Similarity_2) = 31.96244, t-stats = 694.053, p-value = 0.000

As an alternative measure of the correspondence between requested occupations and recommended job titles, for each worker occupation we take every job title recommended to it and compute that title's similarity to each of the 40 occupations, including the original worker occupation that generated the set of recommended job titles. Figure B3.1 averages these scores by title (title, occupation) cell, yielding a 40×40 matrix heatmap. This lets us compare the similarity of the job titles recommended to workers in occupation i to the similarity of those job titles to all other occupations, $j \neq i$. Figure B3.1 shows clear diagonal dominance: titles aligned with their own occupation exceed *any* cross-occupation pairing. Put another way, own-occupation matches are not merely higher on average; they attain the highest rank occupation-by-occupation.

Summing up, Table 3.1 shows high keyword-based match rates and semantic similarity scores between the job titles recommended to workers and their desired occupations; these observed similarity scores are much greater than one would expect by chance ($t =$ 694.053). Furthermore, Table B3.1 shows that occupation-title similarity is greater in the jobs actually recommended to occupation i than in all other occupation-title pairings. Based on these findings, we conclude that the job titles recommended to workers on our job boards correspond closely to workers' desired occupations (and more closely than they do to any other occupation).

Table B3.1: Merging Occupation Categories Across Four Job Boards

Worker's Desired Occupation	Job Board 1	Job Board 2	Job Board 3	Job Board 4
Software Engineer	Software Engineer	Software Engineer Mobile Development Engineer	Software Engineer	
Senior Software Engineer	Senior Software Engineer		Senior Software Engineer	
Front-end/Full Stack Engineer			WEB Front-end Developer	WEB Front-end Developer Full Stack Engineer
Algorithm Engineer		Algorithm Engineer		
Machine Learning/Recognition				Pattern Recognition Machine Learning
Test Engineer/Manager				Test Engineer Test Manager
Operations Assistant/Specialist	Operations Specialist	Operations Specialist	Operations Specialist	Operations Assistant Operations Specialist
Operations Manager/Director	Operations Manager/Supervisor	Operations Manager/Supervisor	Operations Manager/Supervisor	Operation and Maintenance Director Operations Manager/Supervisor
Operations Engineer				Operation and Maintenance Engineer
General Worker /Operator	General Worker /Operator			

Mechanical Engineer			Mechanical Engineer	
Architect			Architect	
Legal manager/Supervisor			Legal manager/Supervisor	
Strategy Consultant				Strategy Consultant
Project Assistant				Product Assistant
Project Manager		Real Estate Project Management Project Manager/Supervisor	Project Manager/Supervisor	Project Manager
Product Assistant				Project Assistant
Product Manager/Supervisor	Product Manager/Supervisor	Product Manager/Supervisor	Product Manager/Supervisor	Product Manager
Sales Representative	Sales Representative Store Clerk	Sales Representative	Sales Representative	Sales Representative
Sales/Marketing Manager/Supervisor	Sales Manager Store Manager	Sales Manager	Sales Manager/Supervisor Marketing Manager/Supervisor	Sales Manager/Supervisor
Sales Director	Sales Director	Sales Director	Sales Director	Sales Director
Design Assistant				Design Assistant
Designer				Designer
Design Manager				Design Manager
Public Relations/Media Specialist		Public Relations Specialist/Assistant		Media Specialist
Public Relations/Media Manager		Public Relations Manager/Supervisor		Media Manager
Data Analyst	Data Analyst	Data Analyst		Data Analyst

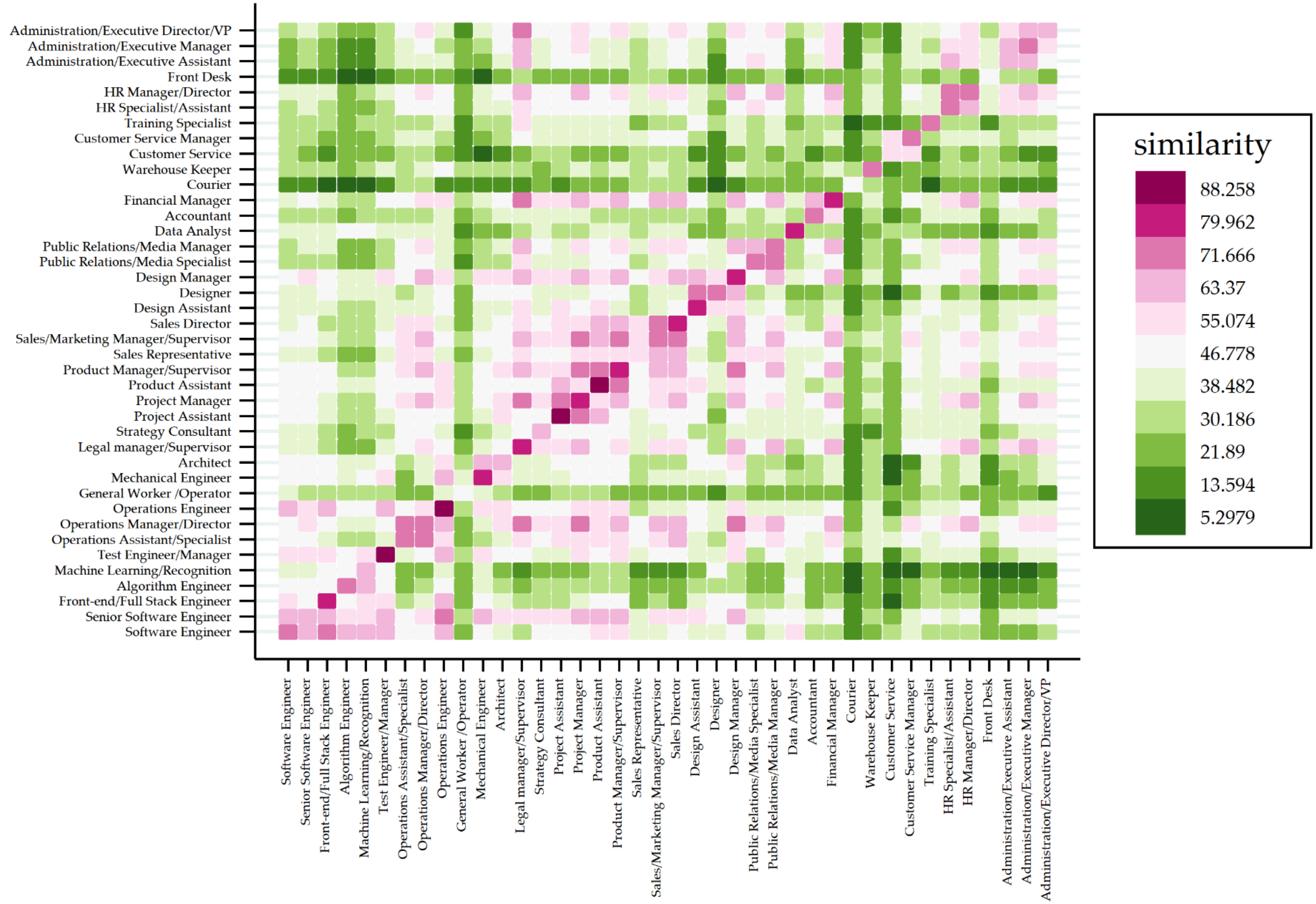
Accountant	Accountant	Accountant	Accountant	Accountant
Financial Manager			Financial Manager	
Courier	Courier			
Warehouse Keeper	Warehouse Keeper			
Customer Service	Customer Service	Web Customer Service Telephone Customer Service		Customer Service
Customer Service Manager	Customer Service Manager	Customer Service Manager		Customer Service Manager
Training Specialist				Training Specialist
HR Specialist/Assistant	HR Specialist/Assistant	HR Specialist/Assistant	Human Resources Specialist/Assistant	Human Resources Specialist/Assistant
HR Manager/Director	Human Resources Manager	Human Resources Manager/Supervisor	Human Resources Manager/Supervisor Human Resources Director	Human Resources Manager/Supervisor Human Resources Director
Front Desk	Front Desk			
Administration/Executive Assistant	Executive Assistant	Executive Assistant/Secretary Administration Specialist/Assistant	Administration Specialist/Assistant Executive Assistant/Secretary	Administration Specialist/Assistant
Administration/Executive Manager	Executive Manager	Administration Manager/Supervisor	Administration Manager/Supervisor	Administration Manager/Supervisor
Administration/Executive Director/VP			Administration Vice President	Administration Director

Table B3.2: Correspondence Between Workers' Desired Occupations and Recommended Job Titles in All Job Boards

Worker's Desired Occupation	Exact Match (%)	Inclusion Match (%)	Any Information Match (%)	Most Information Match (%)	Similarity Score
Software Engineer	2.3	9.9	70.8	61.3	68.7
Senior Software Engineer	1.7	3.2	66.8	37.4	66.0
Front-end/Full Stack Engineer	26.6	42.7	97.7	95.3	78.5
Algorithm Engineer	15.5	65.1	93.5	93.5	73.2
Machine Learning/Recognition	7.5	34.7	54.0	54.0	62.8
Test Engineer/Manager	36.2	62.3	98.5	98.5	86.0
Operations Assistant/Specialist	12.7	54.8	94.2	84.8	75.4
Operations Manager/Director	5.9	13.8	93.1	76.9	73.1
Operations Engineer	50.6	99.8	100.0	100.0	92.4
General Worker /Operator	3.4	7.5	47.8	47.8	42.9
Mechanical Engineer	20.2	40.1	89.2	89.2	78.5
Architect	3.0	4.7	44.6	44.6	61.4
Legal manager/Supervisor	3.8	4.8	91.2	76.8	81.5
Strategy Consultant	3.3	10.3	78.4	78.4	63.6
Project Assistant	53.9	81.2	97.7	97.7	90.4
Project Manager	24.2	46.7	84.0	72.6	78.1
Product Assistant	46.6	73.4	99.3	99.3	89.9
Product Manager/Supervisor	24.8	63.2	95.9	94.4	79.5
Sales Representative	10.4	26.2	72.5	68.6	57.3
Sales/Marketing Manager/Supervisor	13.1	27.4	85.2	66.0	72.7
Sales Director	26.2	48.4	81.7	81.7	78.2
Design Assistant	35.0	60.8	100.0	100.0	78.6

Designer	34.1	80.6	98.8	98.8	72.8
Design Manager	10.2	11.6	97.2	81.1	80.1
Public Relations/Media Specialist	31.0	49.1	87.8	87.8	72.3
Public Relations/Media Manager	20.7	30.0	93.8	76.8	72.4
Data Analyst	32.8	55.9	90.5	90.5	80.6
Accountant	18.1	48.7	61.3	61.3	71.8
Financial Manager	27.3	45.5	80.5	80.5	77.8
Courier	4.9	37.0	44.8	44.8	44.2
Warehouse Keeper	45.3	57.2	73.6	73.6	75.5
Customer Service	15.7	26.9	92.3	54.1	57.5
Customer Service Manager	19.3	42.5	90.8	90.8	72.2
Training Specialist	13.1	59.7	77.6	77.6	69.7
HR Specialist/Assistant	11.5	17.8	78.5	60.3	72.1
HR Manager/Director	9.1	11.9	74.8	56.5	73.3
Front Desk	0.5	0.6	59.0	12.3	43.2
Administration/Executive Assistant	7.2	10.4	79.7	49.7	63.4
Administration/Executive Manager	3.3	8.0	84.1	51.1	70.6
Administration/Executive Director/VP	2.5	5.3	69.6	69.6	61.8
All Occupations (mean)	**15.8**	**32.8**	**81.6**	**70.4**	**70.8**

Figure B3.1: Cross-Occupation Correspondence Heatmap in All Job Boards



Appendix C: Learning from Words

This Appendix provides additional details on the methods used and the robustness of results in Section 3.3 of the paper (Gender Differences in Stereotypical Ad Content).

C1: Most Frequent Words in Job Ads

Figure 2 of the paper presented a translated version of the 172 most common words in job ads. Figure C1.1 presents the original Chinese version.

For ease of discussion, the 172 'most common' words in job ads were manually allocated to six categories, as follows.

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). 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.

Table C1.1 provides a complete list of all 172 'most common' words, by category.

Figure C1.1: 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 C1.1: Word List in Job Ads

The word cloud in Figure 2 shows the 172 most common words in recommended job ads. In Table C1.1, 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 C1.1, 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.

C2: Identifying Over-Represented Words

In the paper we leverage the Romano-Wolf (2005a,b) 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 word list and those values are provided below.

Table C2.1: Multiple Hypothesis Testing for 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.

C3: Relating Over-Represented Words to Gender Stereotypes

Here we provide additional detail on the four external sources used to define stereotypically male and female words.

Published lists

Our first external data source comprises three published papers that have identified gender-stereotypical words in job ads.[33] 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, predict the effect of observing a particular word on the probability the ad explicitly requests only male or female applicants. 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. [2] The resulting words are listed in Table C3.1.

Table C3.1: Gendered Words from Published Lists

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

[2] In more detail, Kuhn et al. (2020) apply the naïve Bayesian classifier to identify the likelihood of an explicit gender request for words in job titles in a Chinese job board. Chaturvedi et al. (2021) use detailed job descriptions in India to 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.

MTurk Survey

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 was: "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 C1.1.

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 C3.2: 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

Chinese 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 C1.1.

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 C3.3: 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

ChatGPT Query

We ask ChatGPT version 4.0 to identify and categorize the words in Table C1.1 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, 2023). Following this, ChatGPT classified the following words as female- and male-associated.

Table C3.4: 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

C4: Which Job Characteristics are Most Predictive of Whether Men or Women See the Ad?

To summarize the relative importance of different job characteristics in determining which gender sees a job ad, columns 1-6 of Table C4.1 run six univariate regressions of the *Male* dummy on different job characteristics, using the samples in Tables 1 and 4. Focusing on the R^2 values in those regressions, these regressions show that a job's stereotypically female content has an order of magnitude more predictive power than any other characteristic: its R^2 is 0.093, compared to the next largest (0.008, for education requirements). Furthermore, column 7 shows that the predictive power of stereotypically female content is more than six times as large as all four of our 'hard' job characteristics (wage, education, experience, and firm size) combined, which yield an R^2 of 0.014.

We conclude that stereotypically female content is especially influential in driving gender differences in job recommendations, in part because female content is relatively rare and highly concentrated in female-intensive jobs. It also supports our interpretation that patterns in female content are critical in identifying which algorithmic recommendations are used on the job boards, because the predictive power of other job characteristics is much weaker.

Table C4.1: Importance of Job Characteristics in Predicting Male Jobs

	(1) MaleJ	(2) MaleJ	(3) MaleJ	(4) MaleJ	(5) MaleJ	(6) MaleJ	(7) MaleJ
Posted Wage	0.0001***						0.0001
(1,000 RMB)	(0.000)						(0.000)
Education		0.0077					0.0036
		(0.005)					(0.005)
Experience			0.0163***				0.0172***
			(0.002)				(0.002)
Firm Size				0.0328***			0.0352***
				(0.008)			(0.008)
Female Content S^f					-0.1602***		
					(0.003)		
Male Content S^m						0.0365***	
						(0.004)	
N	21,262	19,900	21,922	21,949	22,023	22,023	19,227
R^2	0.004	0.008	0.004	0.002	0.093	0.006	0.014

Notes:

1. This table regresses the dummy variable for male-only jobs on different job characteristics. The sample includes only male-only and female-only jobs, consistent with Tables 1 and 4.
2. The posted wage is rescaled to units of 1,000 RMB to ensure that the coefficients are not expressed in very small decimals (e.g., 0.000**).
3. All regressions include resume pair fixed effects.

C5: Gender Differences in Stereotypical Ad Content--Robustness

This Section demonstrates the robustness of our findings in Section 3.3.4 of the paper to two main design choices. The first design choice was our use of only the words that were statistically over-represented (in either male *or* female-directed) job ads to calculate words' stereotype scores based on external sources. 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 3.3.2, which provides a purely inductive description of the words that are over-represented in our data.

As a robustness check, Table C5.1 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 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 C5.1 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 C5.1: 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 C5.1).
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 C5.1).

A second key design choice underlying Table 4 was to combine all four the external sources described in Appendix C3 to calculate an overall stereotype index for each word, assigning equal weight to all four sources. To see if this affects our results, Table C5.2 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.

Table C5.2: 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***	0.0451***
	(0.012)	(0.013)
B. MTurk Survey		
Male	-0.5635***	0.1568***
	(0.013)	(0.013)
C. Chinese Survey		
Male	-0.5336***	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 C5.2, 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 2.11 in row A, 1.24 and 2.02 in row B, 2.07 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 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, firm size, and stereotypically male content. Consistent with the aggregate result in Figure 4, the gender gap in stereotypically female content increases significantly on two boards (3 and 4). There is also a marginally significant increase in Board 1 (p=0.096).

D1: Job Board 1

Table D1.1: 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***	-0.0063	-0.6323***	0.1204***
	(1,327)	(0.039)	(0.048)	(0.013)	(0.034)	(0.032)
N	3,452	3,041	3,598	3,601	3,615	3,615
R^2	0.782	0.551	0.467	0.134	0.373	0.164

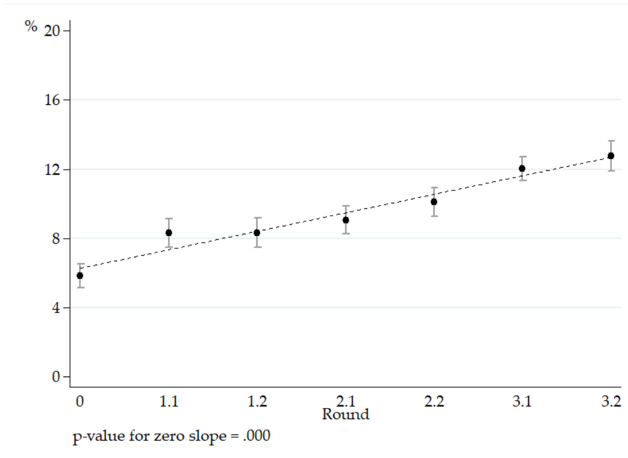
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

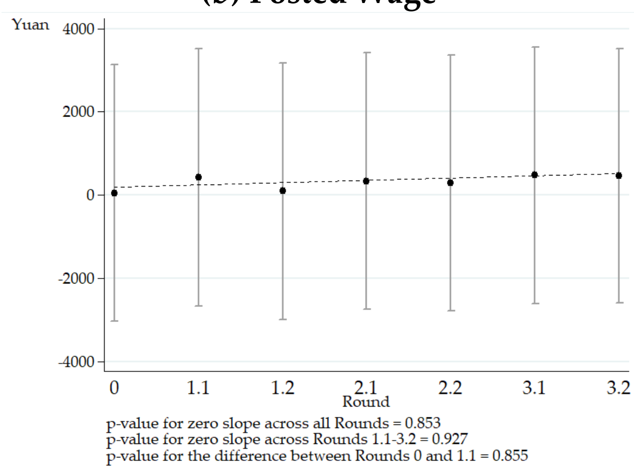
1. This Table replicates Tables 1 and 4 using data from job board 1 only.
2. On job board 1, we collected 44,800 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 8.07%, with 3,615 jobs being exclusively recommended to either male or female applicants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

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

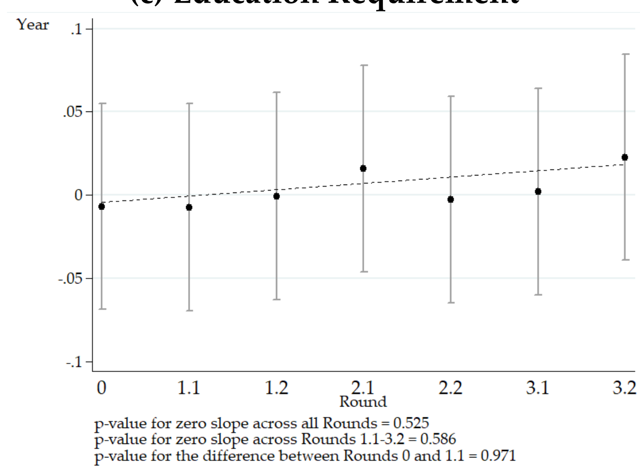
(a) Set Difference Rate



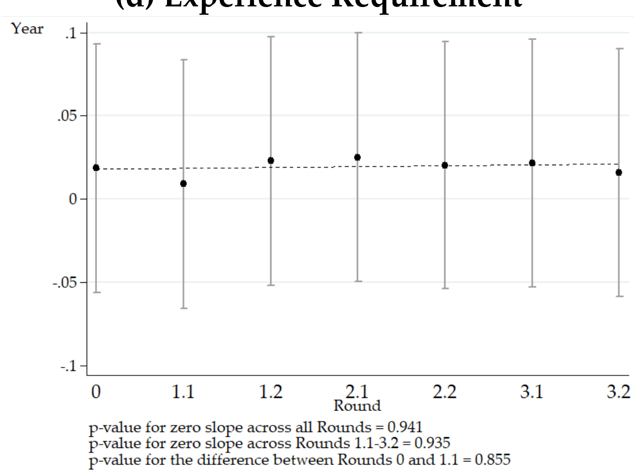
(b) Posted Wage



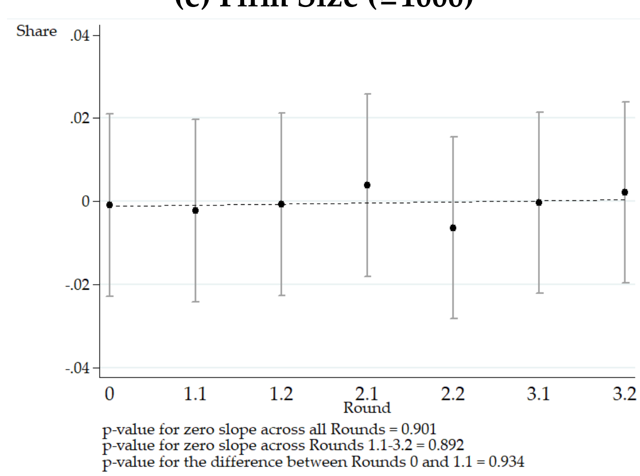
(c) Education Requirement



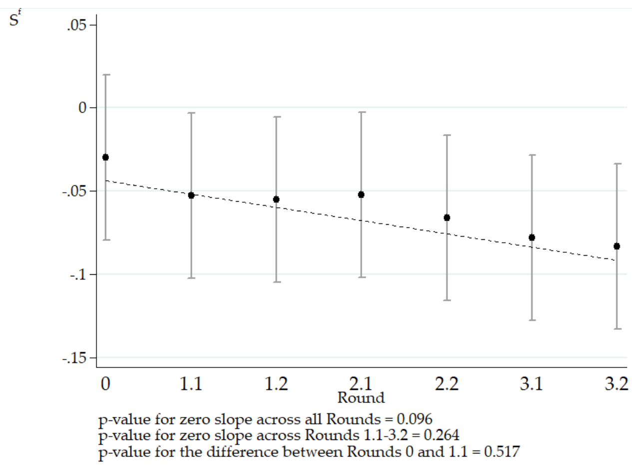
(d) Experience Requirement



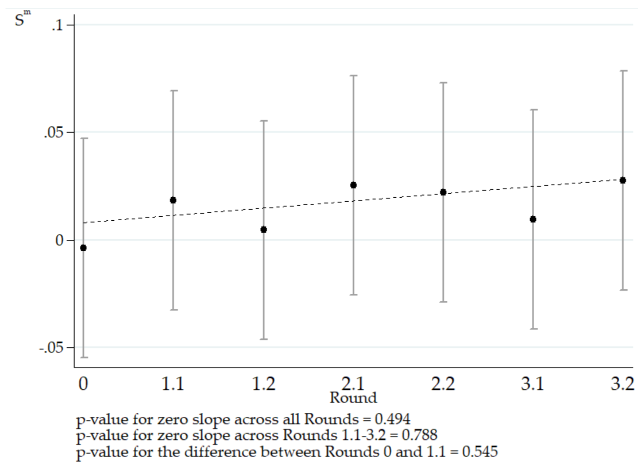
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D1.1 replicates Figures 3 and 4 using job recommendations from Job Board 1, showing the difference rate and differences in job characteristics (and their 95% confidence intervals) between the top ten jobs recommended to men and the top ten jobs recommended to women in each Round of the experiment. These estimates come from a regression of a job's characteristic (e.g. wage) on Round fixed effects, a dummy for Male profiles, and its interaction with Round fixed effects.
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. Also, to allow changes in the share of overlapping jobs to affect gender gaps in their characteristics, we expand our sample to include all recommended jobs, not just the non-overlapping ones.
3. To calculate the trend line in each graph, we replaced the Round fixed effects with a linear Round term (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). Each graph reports the p-value for the interaction between Male and this linear Round variable.
4. The graphs also report p-values for a zero slope excluding Round 0, and for a zero difference between Rounds 0 and 1.1.

D2: Job Board 2

Table D2.1: 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***	0.0350***	-0.4906***	-0.0954***
	(1,530)	(0.019)	(0.044)	(0.013)	(0.026)	(0.026)
N	4,960	4,443	5,071	5,071	5,087	5,087
R^2	0.619	0.426	0.425	0.114	0.299	0.162

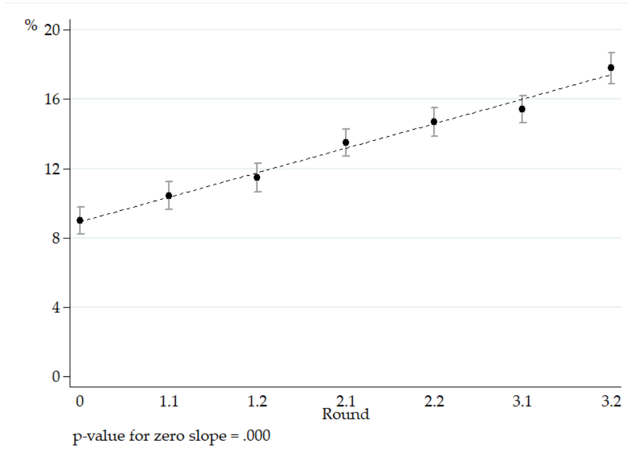
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

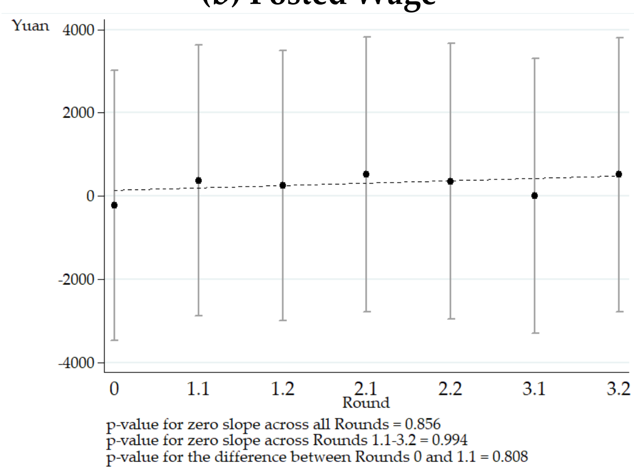
1. This Table replicates Tables 1 and 4 using data from job board 2 only.
2. On job board 2, we collected 43,800 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 11.56%, with 5,087 jobs being exclusively recommended to either male or female applicants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

Figure D2.1: Gender Differences in Characteristics of Job Recommendations in Job Board 2, by Rounds

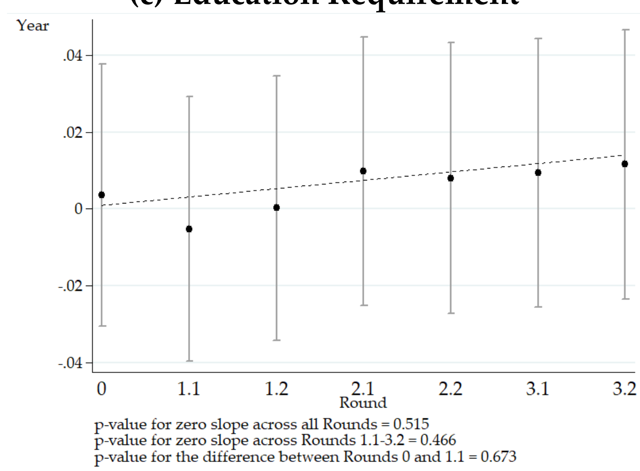
(a) Set Difference Rate



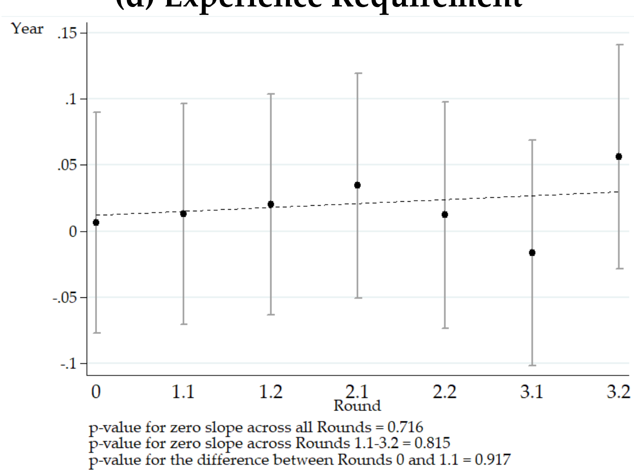
(b) Posted Wage



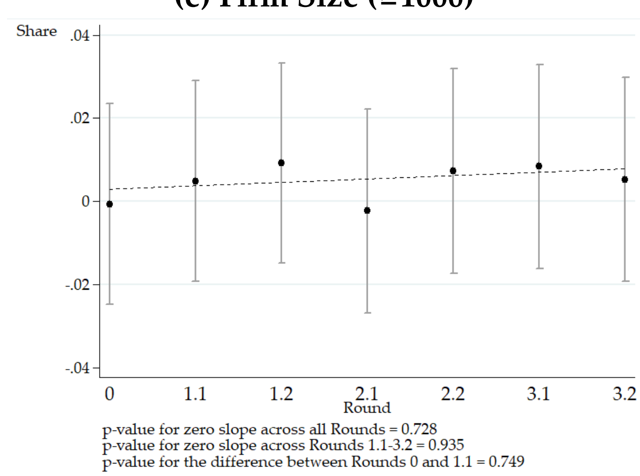
(c) Education Requirement



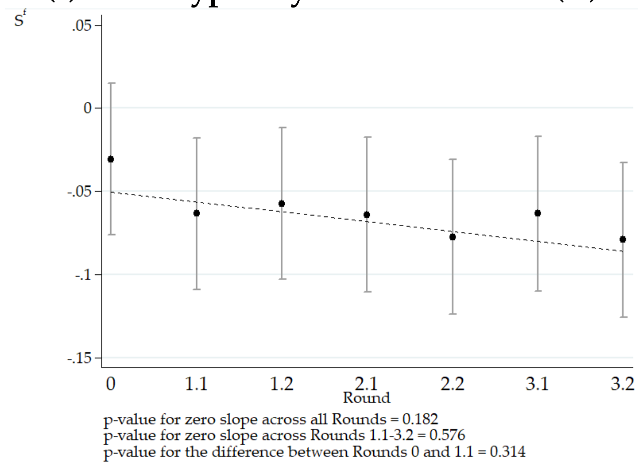
(d) Experience Requirement



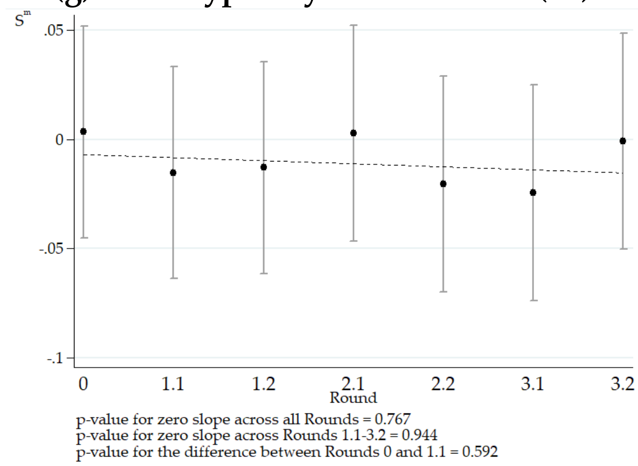
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D2.1 replicates Figures 3 and 4 using job recommendations from Job Board 2, showing the difference rate and differences in job characteristics (and their 95% confidence intervals) between the top ten jobs recommended to men and the top ten jobs recommended to women in each Round of the experiment. These estimates come from a regression of a job's characteristic (e.g. wage) on Round fixed effects, a dummy for Male profiles, and its interaction with Round fixed effects.
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. Also, to allow changes in the share of overlapping jobs to affect gender gaps in their characteristics, we expand our sample to include all recommended jobs, not just the non-overlapping ones.
3. To calculate the trend line in each graph, we replaced the Round fixed effects with a linear Round term (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). Each graph reports the p-value for the interaction between Male and this linear Round variable.
4. The graphs also report p-values for a zero slope excluding Round 0, and for a zero difference between Rounds 0 and 1.1.

D3: Job Board 3

Table D3.1: 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**	0.0615***	-0.5036***	0.2630***
	(2,026)	(0.014)	(0.051)	(0.012)	(0.022)	(0.025)
N	5,915	5,949	6,279	6,303	6,347	6,347
R^2	0.584	0.182	0.246	0.080	0.276	0.102

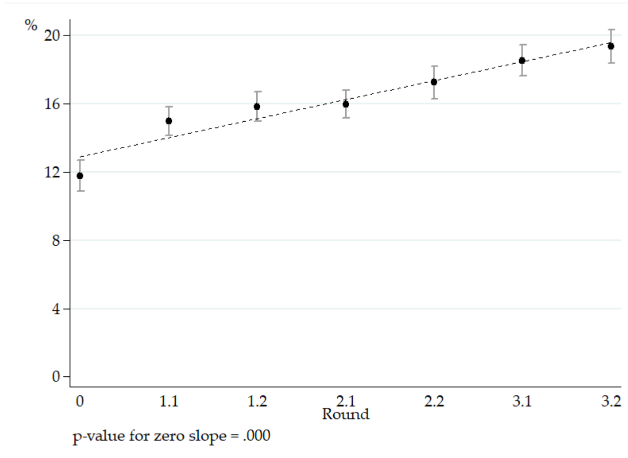
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

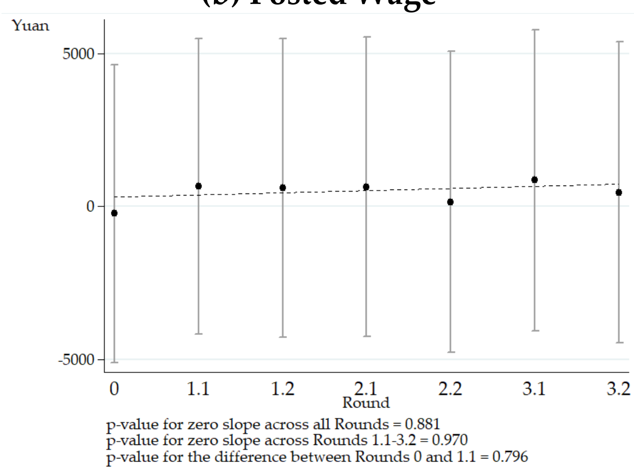
1. This Table replicates Tables 1 and 4 using data from job board 3 only.
2. On job board 3, we collected 44,320 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 14.31%, with 6,347 jobs being exclusively recommended to either male or female applicants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

Figure D3.1: Gender Differences in Characteristics of Job Recommendations in Job Board 3, by Rounds

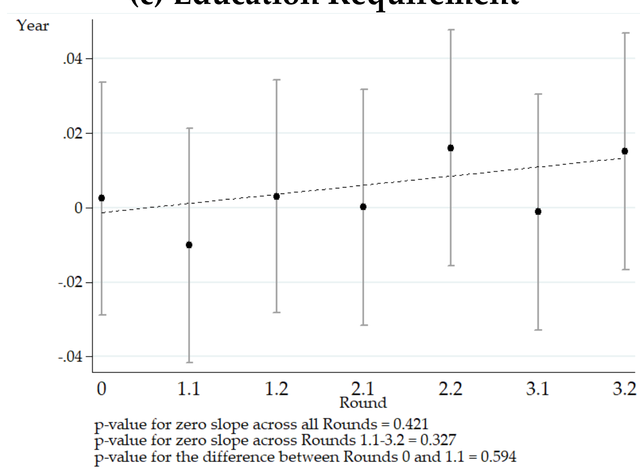
(a) Set Difference Rate



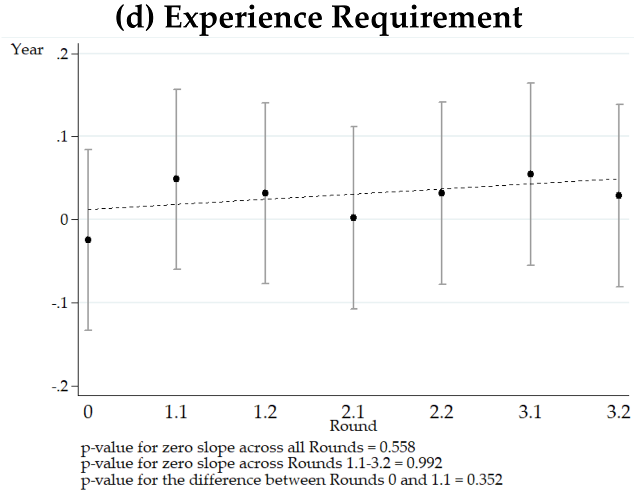
(b) Posted Wage



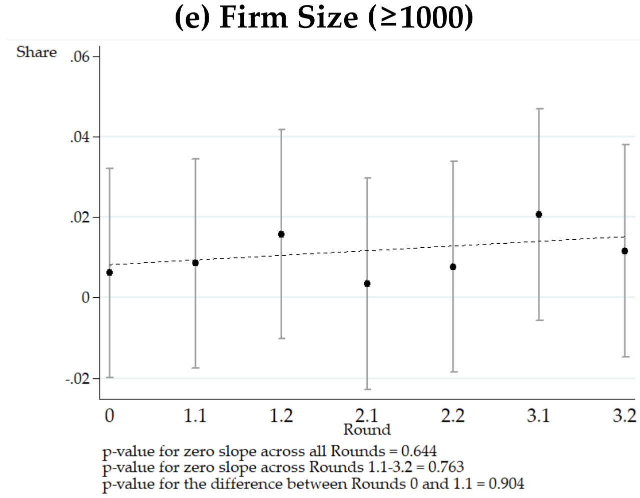
(c) Education Requirement



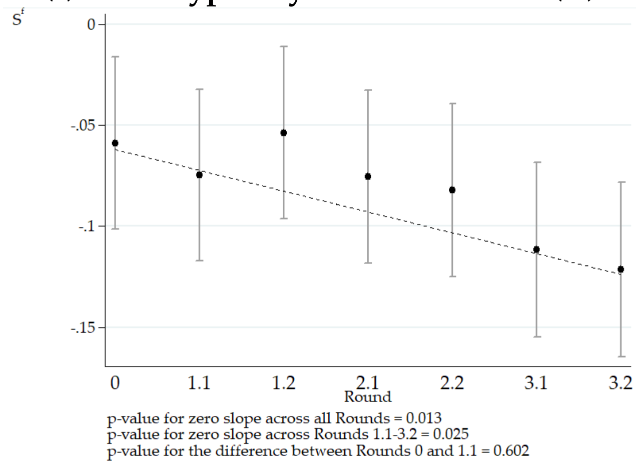
(d) Experience Requirement



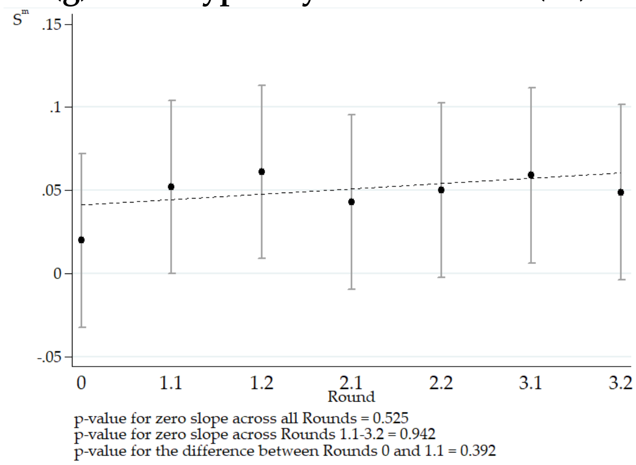
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D3.1 replicates Figures 3 and 4 using job recommendations from Job Board 3, showing the difference rate and differences in job characteristics (and their 95% confidence intervals) between the top ten jobs recommended to men and the top ten jobs recommended to women in each Round of the experiment. These estimates come from a regression of a job's characteristic (e.g. wage) on Round fixed effects, a dummy for Male profiles, and its interaction with Round fixed effects.
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. Also, to allow changes in the share of overlapping jobs to affect gender gaps in their characteristics, we expand our sample to include all recommended jobs, not just the non-overlapping ones.
3. To calculate the trend line in each graph, we replaced the Round fixed effects with a linear Round term (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). Each graph reports the p-value for the interaction between Male and this linear Round variable.
4. The graphs also report p-values for a zero slope excluding Round 0, and for a zero difference between Rounds 0 and 1.1.

D4: Job Board 4

Table D4.1: 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

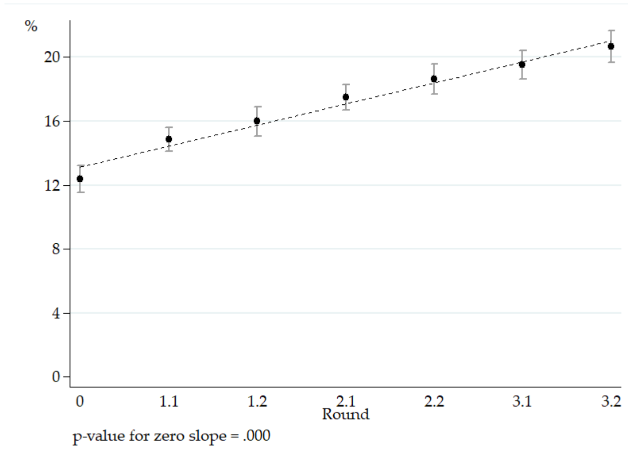
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Notes:

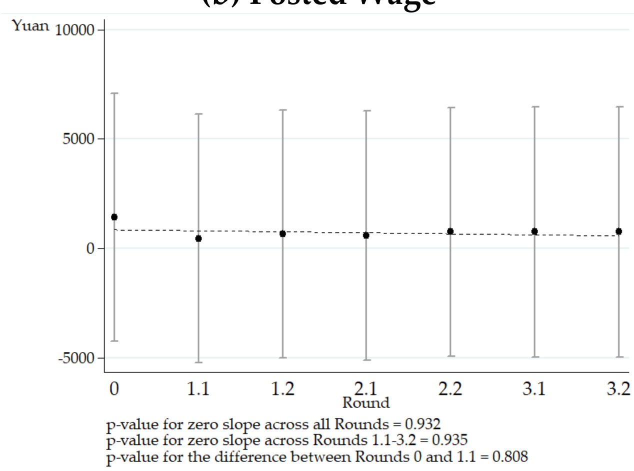
1. This Table replicates Tables 1 and 4 using data from job board 4 only.
2. On job board 4, we collected 44,320 job recommendations for 560 fictitious profiles. The set difference rate in these job recommendations is 15.68%, with 6,974 jobs being exclusively recommended to either male or female applicants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

Figure D4.1: Gender Differences in Characteristics of Job Recommendations in Job Board 4, by Rounds

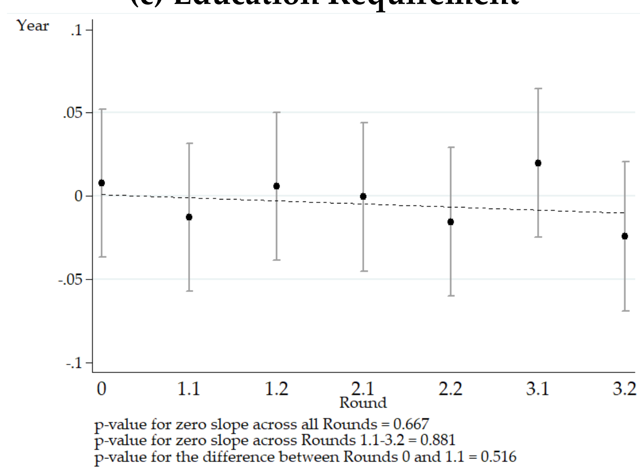
(a) Set Difference Rate



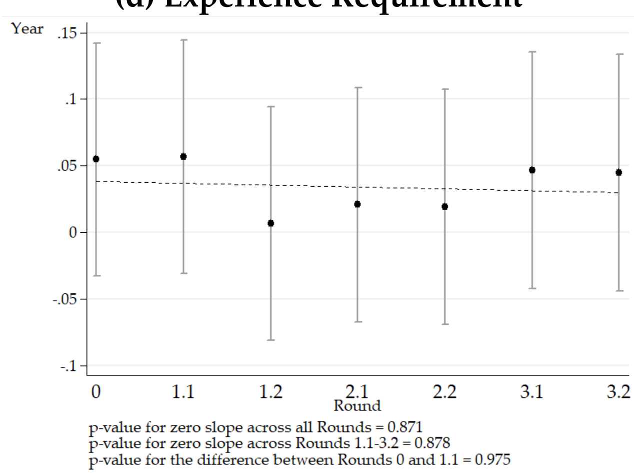
(b) Posted Wage



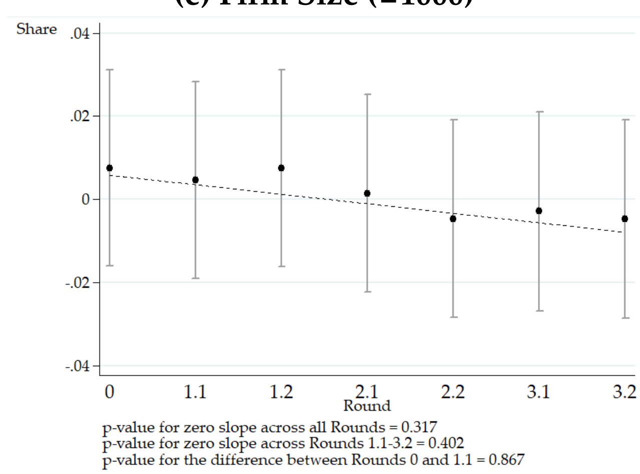
(c) Education Requirement



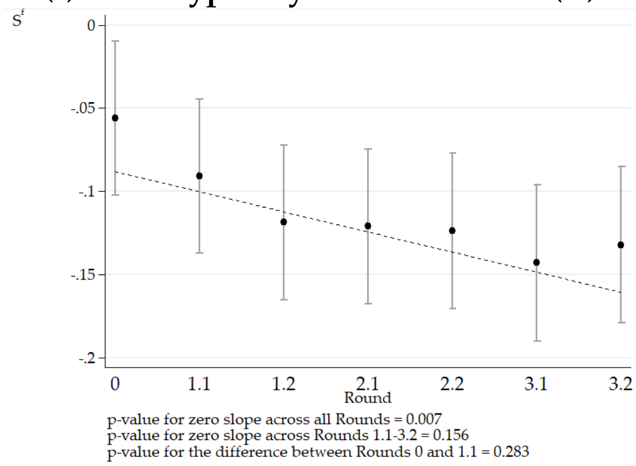
(d) Experience Requirement



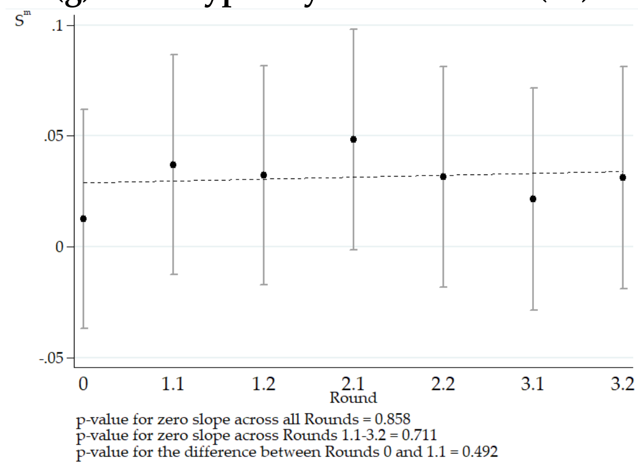
(e) Firm Size (≥1000)



(f) Stereotypically Female Content (S^f)



(g) Stereotypically Male Content (S^m)



Notes:

1. Figure D4.1 replicates Figures 3 and 4 using job recommendations from Job Board 4, showing the difference rate and differences in job characteristics (and their 95% confidence intervals) between the top ten jobs recommended to men and the top ten jobs recommended to women in each Round of the experiment. These estimates come from a regression of a job's characteristic (e.g. wage) on Round fixed effects, a dummy for Male profiles, and its interaction with Round fixed effects.
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. Also, to allow changes in the share of overlapping jobs to affect gender gaps in their characteristics, we expand our sample to include all recommended jobs, not just the non-overlapping ones.
3. To calculate the trend line in each graph, we replaced the Round fixed effects with a linear Round term (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). Each graph reports the p-value for the interaction between Male and this linear Round variable.
4. The graphs also report p-values for a zero slope excluding Round 0, and for a zero difference between Rounds 0 and 1.1.

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 E1.1 replicates Table B2.4 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 E1.1, 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 B2.4.

Table E1.1: Ranking Difference Rate in Job Recommendations

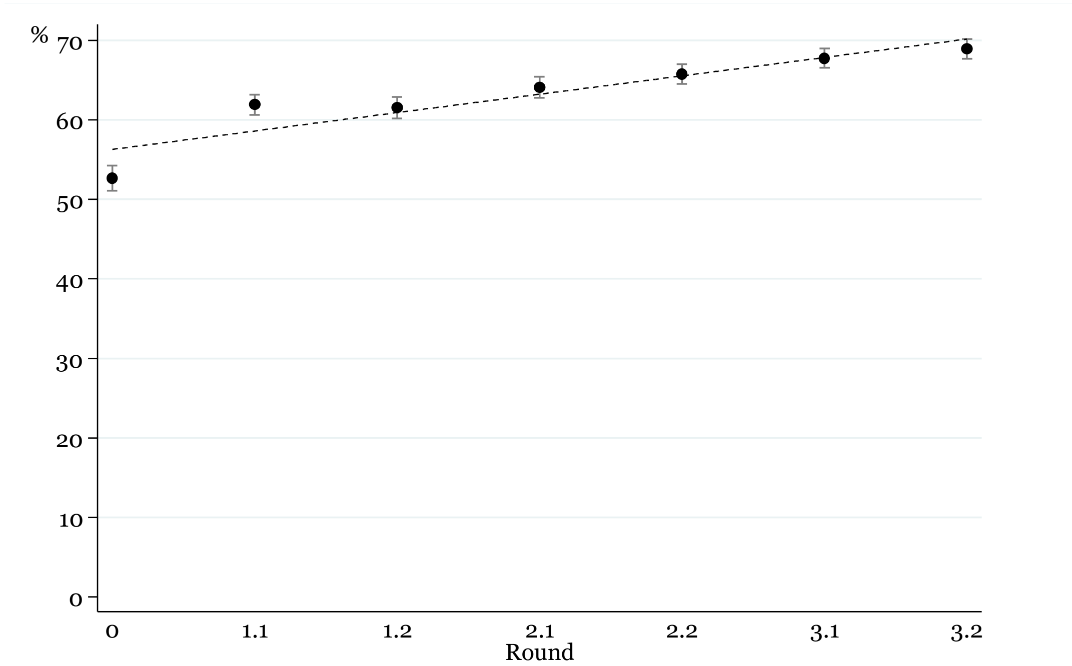
	Difference Rate (S.D.)	Between-Group Differences
All Recommendations	0.6132 (0.090)	
Worker Age		
Young	0.6162 (0.091)	0
Old	0.6102 (0.090)	-0.0060
Job Gender Type		
Female-dominated	0.5986 (0.096)	-0.0249***
Gender Neutral	0.6234 (0.083)	0
Male-dominated	0.6185 (0.089)	-0.0049
Job Skill Level		
Entry	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**

Notes:

1. This Table replicates Table B2.4 with rank difference rate. Statistics are for all four job boards combined.
2. 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.

In Figure E1.1, we replicate Figure 3 to show the trends across experimental rounds in *rank* (as opposed to set) difference rates. Figure E1.1 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 E1.1: Ranking Difference Rate by Experimental Rounds



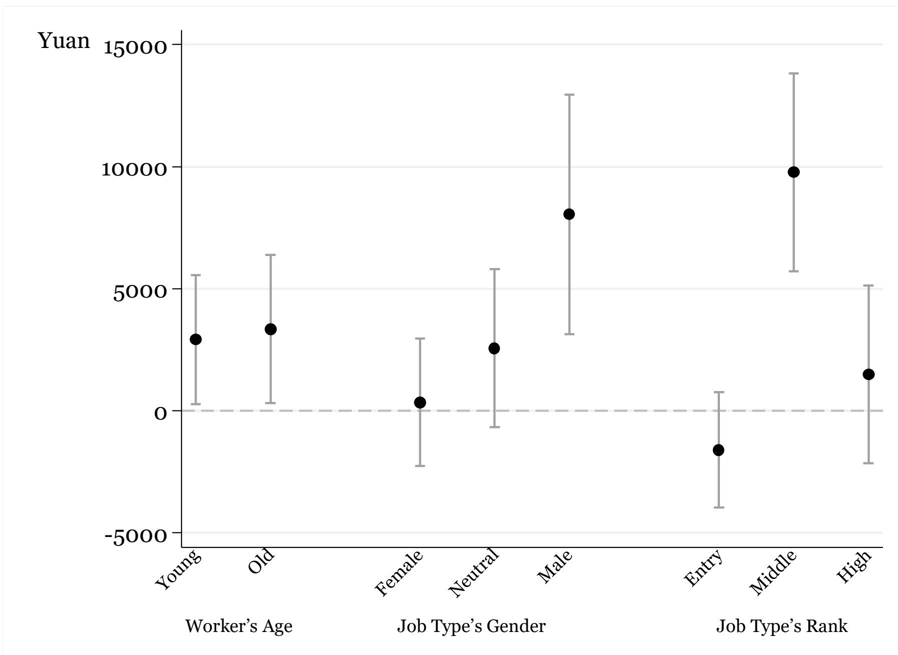
Notes:

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).

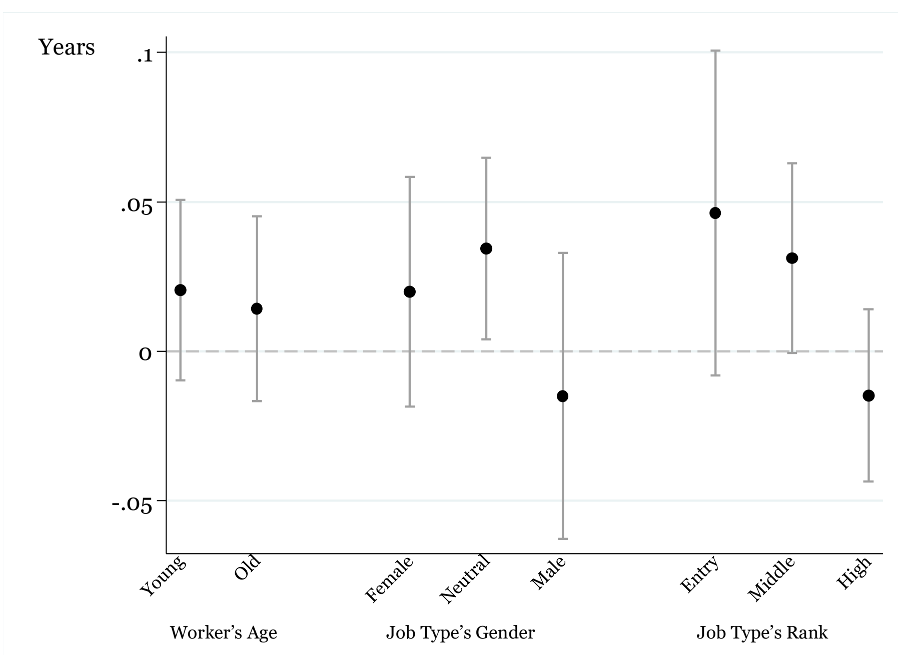
E2: Gender Differences in Recommended Job Characteristics by Applicant Age and Job Types

Figure E2.1: Heterogeneity in Gender Gaps

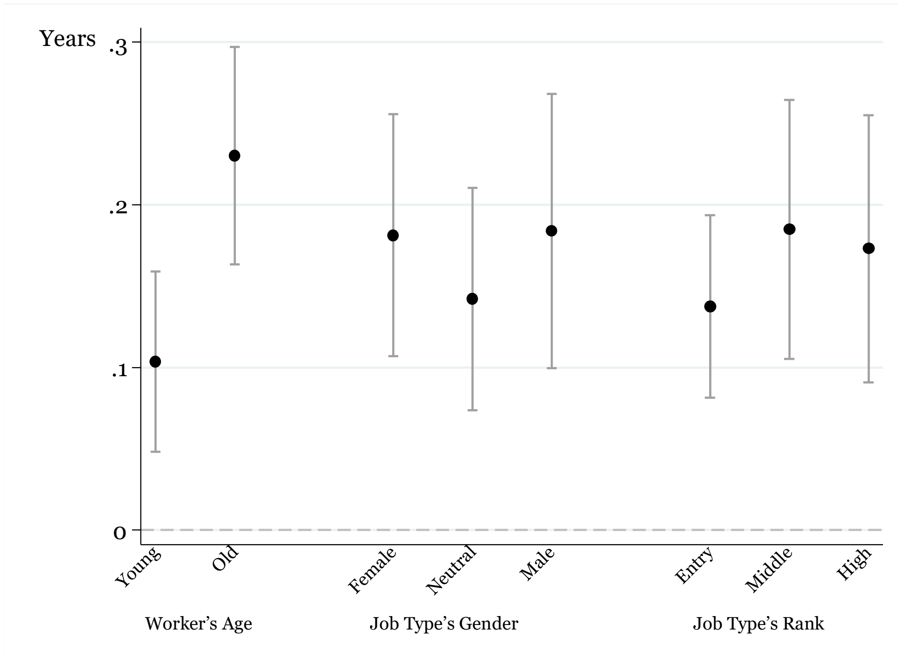
(a) Gender Differences in Posted Wages



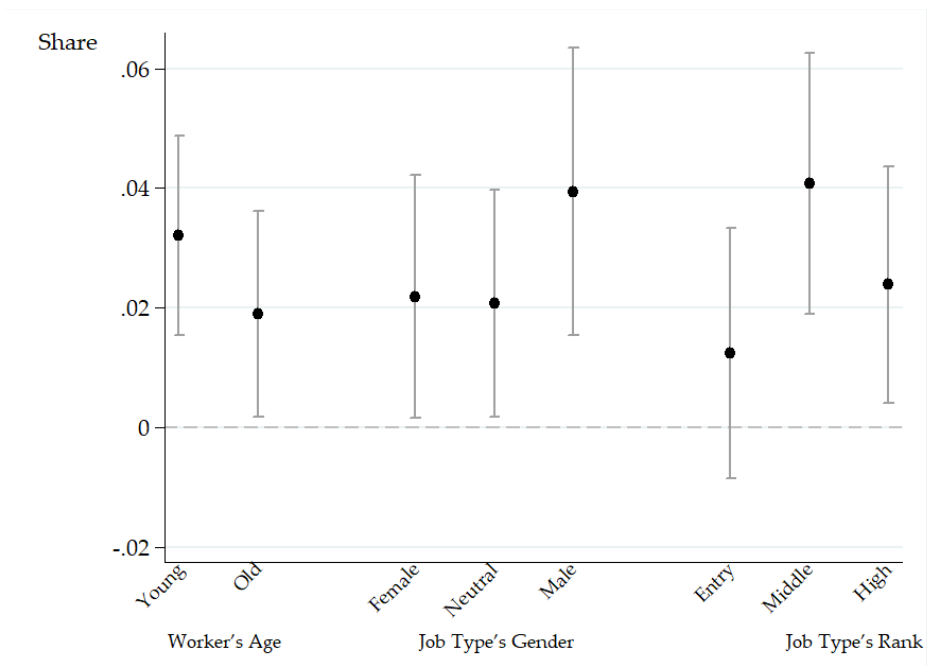
(b) Gender Differences in Requested Education



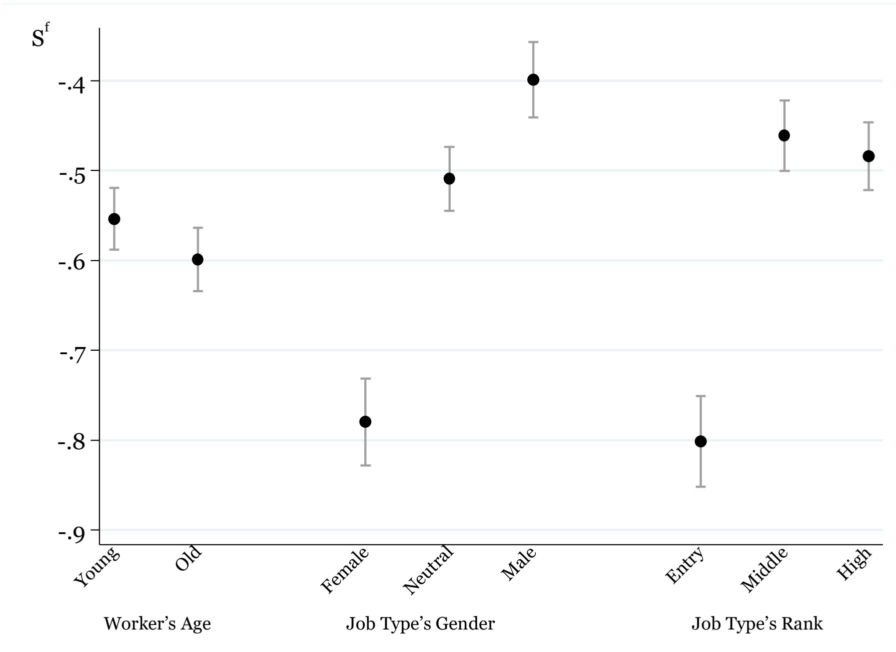
(c) Gender Differences in Requested Experience



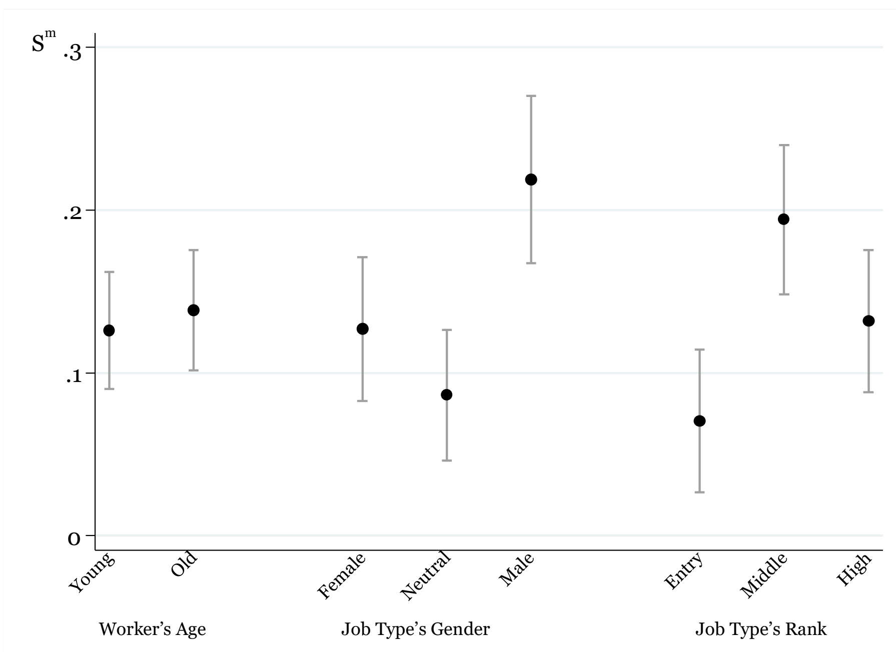
d) Gender Differences in Recommended Firm Size



(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 E3.1 estimates the cross-sectional return to experience and firm size in all the jobs that were recommended to our profiles. Table E3.1 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 0.0256 times more likely to be in firms with 1000 or more employees. Using the coefficients in column 5 of Table E3.1, 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*0.0256 + 3,610*0.1656*0.0256 = 3,901 RMB, or 1.89 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 E3.1: 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 E4.1, 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 E4.1: 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

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

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 B2.1 and B2.2, 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 E5.1 shows no significant gender difference in the financial status of recommended employers.

Table E5.1: 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

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

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 B2.1 and B2.2, and in Appendix D.

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

Table E6.1: 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***	0.0256***	-0.5759***	0.1323***
	(1,022)	(0.011)	(0.022)	(0.006)	(0.012)	(0.013)
N	21,262	19,899	21,922	21,949	22,023	22,023
R^2	0.613	0.454	0.394	0.173	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).

E7: 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 E7.1 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 E7.1, 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 E7.1: Including 'Common' Jobs in Tables 1, 4, and 5

A. Tables 1 and 4

	(1) Posted Wage	(2) Education	(3) Experience	(4) Firm Size ($\geqslant$1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	-4,612***	-0.0154*	0.0239	0.0136***	-0.0645***	0.2295***
	(870)	(0.009)	(0.018)	(0.005)	(0.009)	(0.010)
Female	-9,511***	-0.0385***	-0.1570***	-0.0160***	0.5223***	0.0964***
	(909)	(0.009)	(0.019)	(0.005)	(0.010)	(0.011)
N	78,551	73,924	81,870	81,894	83,793	83,793
R^2	0.599	0.489	0.365	0.125	0.247	0.092

B. Table 5 (Round 0 recommendations only)

	(1) Posted Wage	(2) Education	(3) Experience	(4) Firm Size ($\geqslant$1000)	(5) Stereotypically Female Content (S^f)	(6) Stereotypically Male Content (S^m)
Male	-5,259***	-0.0016	-0.0058	0.0227**	-0.0887***	0.1784***
	(1,659)	(0.018)	(0.036)	(0.010)	(0.019)	(0.021)
Female	-9,686***	-0.0148	-0.2126***	-0.0054	0.5301***	0.0697***
	(1,925)	(0.021)	(0.042)	(0.011)	(0.022)	(0.024)
N	22,078	20,734	23,067	23,067	23,610	23,610
R^2	0.633	0.513	0.378	0.150	0.266	0.124

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

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

E8: The Role of *C*-Mechanisms in Rounds 1.1-3.2

While *C*-mechanisms are available to job boards in all the Rounds of our experiment, we think it is highly unlikely that they can cause either difference rates or gender gaps in recommended ad content to grow across the Rounds of our experiment, and we use this assumption (of weakly declining effects) to identify the effects of the newly-available *R*- and *C*-mechanisms.

We believe that weak decline in the role of *C*-mechanisms is very likely for the following reasons.

1. The resume is a static document. Essentially, *C*-mechanisms look for matches between a static document --the worker's resume-- and the population of vacant resumes on a job board. Unless this population changes systematically during the eight-week period in which a profile is searching, content-based matching should not cause changes in the types of jobs recommended to men versus women over time. We confirm this intuition in Appendix Table E6.1, which replicates our main results while controlling for job board × calendar week fixed effects, showing almost no change.

2. Stock versus flows of vacancies. In Round 0, *C*-mechanisms can draw from the entire stock of ads on the boards to make recommendations; in Rounds 1-3 only the newly-posted ads are likely to be highly relevant to workers (Gregg and Petrongolo, 2005). This should reduce the influence of *C*-mechanisms on the outcomes we measure across Rounds of the experiment.

3. *C*-mechanisms do not learn from the user's or recruiters' behavior. As the worker applies to more and more jobs, and her resume is seen by more recruiters, *R*- and *W*-mechanisms can learn from these actions to make their recommendations more relevant to the individual user. This reduces the relative quality of *C*-derived recommendations, which should reduce boards' reliance on them over time.

For all these reasons, we think it is reasonable to interpret increasing gender recommendation gaps across experimental Rounds to *R*- and *W*-mechanisms, and not to a growing influence of *C*-mechanisms.

Reference:

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

Appendix F: Hiring Agents' Views and Job Recommendations

In Table 6, we assessed the plausibility of Channels 4 and 5 (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 5 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 6 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 6.

Table F4 provides background information for our assessment of Channels 4 and 5, 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.

Motivated by the gender gap in the number of times our profiles were viewed by HR agents, Table F5 replicates Table 6 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.

Tables F6 and F7 explore the possibility that uncontrolled differences in labor market tightness across resume pairs account for Table 6's positive associations between resume views and the set difference rate. Our first approach, in Table F6, is to add controls for cross-market differences in market tightness. To do so, we pool the regressions in Table 6 across our three experimental Intervals (to maximize statistical power), then add 140 fixed effects for our 35 industry-occupation cells ("job types") interacted with four job board fixed effects. We expect market tightness to vary substantially across these different types of work; these fixed effects will control for that variation. While the magnitude of the association between profile views and difference rates is reduced (suggesting some role for unobserved factors that vary across labor markets), the coefficient remains positive and highly statistically significant.[3]

In Table F7 we exploit the fact that in Round 0 (and only in Round 0) we collected up to 100 recommendations for every profile pair, which generates variation in the total number of recommendations received by the pair.[4] We use this variation to ask whether high difference rates in the top 10 recommended jobs (our outcome variables in Table 6) are associated with a larger pool of possible recommendations to choose from. Without job board fixed effects, the table shows a positive association, but this association disappears when we add job board fixed effects. Thus, the total number of jobs 'available' to recommend does not affect the 'top-10' difference rate measures which are our main outcomes in Table 6. This also argues against the idea that the associations in Table 6 are spurious consequences of differences in labor market tightness.

[3] Panel A of Table F6 just pools the data across Intervals, adding Interval fixed effects; the results are very similar to Table 6. Panel B adds the job type x board fixed effects.

[4] We do not use recommendations ranked 21-100 in our main analysis to maintain consistency across experimental Rounds. Adding these recommendations does not change the main results.

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***	-27	0.0014**	0.0004	-0.0002	-0.0011	-0.0011
	(0.028)	(51)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,118	1,118	1,118	1,118	1,118	1,118	1,118
R^2	0.317	0.003	0.012	0.011	0.013	0.074	0.058
B. Interval 2							
Male Profile Views	0.0364	75	-0.0001	0.0002	-0.0001	-0.0004	-0.0008
	(0.027)	(53)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,100	1,100	1,100	1,100	1,100	1,100	1,100
R^2	0.338	0.010	0.005	0.003	0.007	0.055	0.041
C. Interval 3							
Male Profile Views	0.1341***	16	-0.0001	0.0009	0.0004	-0.0007	0.0011
	(0.026)	(52)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.305	0.006	0.003	0.012	0.021	0.077	0.037

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

Notes:

1. This Table replicates Table 6, 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***	-49	0.0004	0.0011	-0.0002	-0.0013*	-0.0007
	(0.029)	(54)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,118	1,118	1,118	1,118	1,118	1,118	1,118
R^2	0.316	0.003	0.008	0.012	0.013	0.074	0.057
B. Interval 2							
Female Profile Views	0.1506***	144**	0.0002	-0.0009	0.0005	-0.0016**	-0.0001
	(0.030)	(60)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,100	1,100	1,100	1,100	1,100	1,100	1,100
R^2	0.352	0.013	0.005	0.003	0.009	0.058	0.040
C. Interval 3							
Female Profile Views	0.0989***	88	-0.0006	0.0011	0.0003	-0.0017**	0.0007
	(0.030)	(59)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.295	0.008	0.004	0.012	0.020	0.080	0.036

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

Notes:

1. This Table replicates Table , replacing the regressor with F*emale 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 Gender Recommendation Gaps in Rounds 1.1, 2.1 and 3.1 only

	(1) Difference Rate	Gender Gap (*male* – *female*) in: (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***	31	0.0002	0.0015	0.0002	-0.0014**	-0.0009
	(0.026)	(43)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,118	1,118	1,116	1,118	1,118	1,118	1,118
R^2	0.164	0.003	0.005	0.022	0.006	0.034	0.028
B. Interval 2							
Views	0.0779***	114**	0.0006	0.0011	0.0006*	-0.0017**	-0.0005
	(0.027)	(47)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,100	1,100	1,099	1,100	1,100	1,100	1,100
R^2	0.193	0.013	0.005	0.005	0.005	0.037	0.016
C. Interval 3							
Views	0.0748***	33	-0.0006	0.0002	0.0007**	-0.0011	0.0005
	(0.027)	(50)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.162	0.005	0.006	0.013	0.024	0.056	0.030

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

Notes:

1. This Table replicates Table 6 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	13	0.0008	-0.0004	0.0000	0.0000	-0.0004
	(0.024)	(45)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,118	1,118	1,118	1,118	1,118	1,118	1,118
R^2	0.304	0.003	0.010	0.011	0.013	0.072	0.057
B. Interval 2							
M-F Views	-0.0548**	-26	-0.0002	0.0006	-0.0003	0.0005	-0.0005
	(0.022)	(44)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,100	1,100	1,100	1,100	1,100	1,100	1,100
R^2	0.340	0.009	0.005	0.003	0.009	0.055	0.041
C. Interval 3							
M-F Views	0.0374*	-35	0.0003	0.0001	0.0001	0.0004	0.0003
	(0.022)	(42)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
N	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.290	0.007	0.003	0.012	0.020	0.077	0.036

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

Notes:

1. This Table replicates Table 6, replacing the regressor with the gender gap (male - female) in the number of profile views
2. The means of gender gap in profiles views are 0.50, 0.64 and 0.96 in Intervals 1-3 respectively.

Table F6: Effects of Profile Views on Gender Recommendation Gaps, adding Job Board × Job Type Fixed Effects

	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 (S^f)	(7) Stereotypically Male Content (S^m)
A. Pooling the 3 Intervals with Interval Fixed Effects							
Views	0.0812***	29	0.0002	0.0003	0.0001	-0.0008***	-0.0000
	(0.010)	(20)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3,313	3,313	3,313	3,313	3,313	3,313	3,313
R^2	0.379	0.005	0.004	0.006	0.011	0.073	0.038
B. Adding Job Type × Job Board Fixed effects							
Views	0.0488***	18	0.0003	0.0002	0.0001	-0.0006**	-0.0001
	(0.010)	(20)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3,313	3,313	3,313	3,313	3,313	3,313	3,313
R^2	0.436	0.076	0.055	0.058	0.061	0.173	0.094

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

Notes:

1. Panel A replicates Table 6, pooling the regressions across the three Intervals and adding Interval fixed effects.
2. Panel B adds job type × job board (140) fixed effects to the specification in Panel A.

Table F7: Round 0 Regressions-- Relation Between a Pair's Difference Rate in the Top 10 or 20 Recommended Jobs and the Number of Additional Recommendations Received

	(1)	(2)	(3)	(4)
Difference Rate in 1-10	0.3587***	0.0077		
	(0.086)	(0.065)		
Difference Rate in 1-20			1.0004***	0.0897
			(0.150)	(0.122)
Age & Gender type FE	Yes	Yes	Yes	Yes
Job board FE		Yes		Yes
N	1,120	1,120	1,120	1,120
R^2	0.020	0.522	0.044	0.522

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

Notes:

1. This table shows the relationship between a pair's set difference rate in Round 0's top 10 or 20 jobs (the outcome variables in the paper), and the total number of additional job recommendations received by the pair in Round 0 (which can be as high as 160).
2. Columns 1 and 3 control for the predominant gender in the job sought (female, neutral, or male) and profile's age (young versus older). Columns 2 and 4 add Job Board fixed effects.
3. Duplicate jobs received by the same profile are removed from the calculations. The outcome variable is the total number of unique jobs received by the two workers in each pair. 23% of pairs received 160 unique job recommendations, and the average is 135.

Appendix G: Workers' Previous Applications and Job Recommendations

Here we provide a 'clean' test for the presence of *W*-mechanisms by focusing on three specific points in our experiment where *only* those mechanisms should be available: *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 6 and 7 ("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 G1 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. While this is consistent with a causal effect *W*-mechanisms on recommendations, but --as discussed in Section 3.1—several other processes including quasi-randomness and dispersion/decongestion processes could also account for these changes. That noted, Table G1 shows that none of the gender gaps in recommended jobs' *characteristics* grow between these sub-rounds.[5] Contrary to our expectations, there is no evidence that workers' recent applications affect the recommendations they receive when they log on to their profiles.

[5] While the standard errors for some of these comparisons are large, Table G1 rules out even fairly small effects relative to the baseline gender gaps in experience and stereotypical ad content.

Table G1: Gender Gaps in Job Recommendations *Within* Rounds 1-3 (including the common jobs)

		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	509	0.0014	0.0263**	0.0040	-0.0825***	0.0232***
	(0.146)	(655)	(0.007)	(0.013)	(0.004)	(0.007)	(0.007)
Last: 11-20 (β_2)	15.06	447	0.0030	0.0258*	0.0041	-0.0875***	0.0230***
	(0.164)	(655)	(0.007)	(0.013)	(0.004)	(0.007)	(0.007)
Difference ($\beta_2 - \beta_1$)	**1.09*** (p = 0.000)**	**-62 (p = 0.947)**	**0.0017 (p = 0.856)**	**-0.0005 (p = 0.977)**	**0.0001 (p = 0.978)**	**-0.0049 (p = 0.612)**	**-0.0002 (p = 0.984)**
N	2,236	123,771	116,925	129,476	129,503	132,520	132,520

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

Notes:

1. Table G1 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. In column 1, the difference rate is calculated at the pair level, and the p-value is derived from a t-test comparing the first 10 jobs with the last jobs.
3. In column 2 to 7, 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).
4. Row 3 shows the increase in the gender gap ($\beta_2 - \beta_1$) and the p value from the F test of $\beta_2 - \beta_1 = 0$.