Gender Bias in the Job Market: A Longitudinal Analysis

SHILIANG TANG, UC Santa Barbara
XINYI ZHANG, UC Santa Barbara
JENNA CRYAN, University of Chicago
MIRIAM J. METZGER, UC Santa Barbara
HAITAO ZHENG, University of Chicago
BEN Y. ZHAO, University of Chicago

For millions of workers, online job listings provide the first point of contact to potential employers. As a result, job listings and their word choices can significantly affect the makeup of the responding applicant pool. Here, we study the effects of potentially gender-biased terminology in job listings, and their impact on job applicants, using a large historical corpus of 17 million listings on LinkedIn spanning 10 years. We develop algorithms to detect and quantify gender bias, validate them using external tools, and use them to quantify job listing bias over time. We then perform a user survey over two user populations (N_1=469, N_2=273) to validate our findings and to quantify the end-to-end impact of such bias on applicant decisions. Our findings show gender-bias has decreased significantly over the last 10 years. More surprisingly, we find that impact of gender bias in listings is dwarfed by our respondents' inherent bias towards specific job types.

CCS Concepts: • **Human-centered computing** → **Computer supported cooperative work**; • **Social and professional topics** → *Employment issues*; • **Applied computing** → Sociology;

Additional Key Words and Phrases: Gender Bias, Online Job Advertisements, Quantitative Analysis, User Studies

1 INTRODUCTION

Despite recent strides made by women in the workplace, workplace inequality persists [9, 20]. In addition to the widely distributed reports of wage inequality across genders [7], it is also well known that there are significantly fewer women in male dominated positions, across both industry sectors (*e.g.,* technology [29]), and job types (*e.g.,* corporate executives [41]).

A number of contemporary theories have hypothesized the source of these gender imbalances, whether they come from different educational paths and biases introduced early on [12, 27, 37],

or whether they are the result of workplace attrition [13, 22, 44]. The job search and hiring process might be another contributor [6, 15, 26, 30, 34, 35, 43]. Hidden biases often creep into job listings [2, 5, 17, 21] , and can either actively or passively discourage certain applicants from applying.

The goal of this work is to understand the role that job postings play in introducing or exacerbating gender imbalance in the workplace. More specifically, we are interested in two general questions. First, how significant is gender bias in today's job listings? How much does gender bias vary over different industry sectors, and how has it changed over time? We hope to track levels of gender bias in job postings through time, to see if any changes are reflected as a result of society's growing awareness of gender bias. Second, we look to measure the end-to-end impact of gender bias in job posts on the actual decisions of potential job applicants. Are applicants aware of such bias in job posts, and do these biases play a role in their decisions to apply for jobs?

To answer these questions, we perform a study with two components, an empirical, data-driven component that quantifies the presence and magnitude of gender bias in job postings over the last 10 years, and a qualitative user-study that seeks to understand the end-to-end impact of biases on whether applicants apply to a posted position.

On the empirical side, we analyze 17 million job posts collected over the last 10 years (2005–2016). We obtain this dataset of job posts through a near-complete download of job posts maintained by LinkedIn, the largest professional networking site online with 500 million users. To quantify gender bias over large datasets, we develop two scalable algorithms that match the same metrics as online services that evaluate job postings for gender bias, Textio and Unitive. We tune our algorithms and show they can approximate Unitive and Textio in bias classification, generating a raw score and a normalized score of gender bias between feminine and masculine. We use a test sample of key words/phrases to validate our approaches against Unitive and Textio. We then apply our algorithm to the LinkedIn dataset to quantify gender bias in the whole market, specific sectors, and its changes over time.

In our user study, we augment our data analysis with a user survey that captures how gender bias in wording actually affects job applicants. We ask detailed questions to 2 user populations, 469 Amazon Turk workers, and 273 undergraduate college students, to understand their perceived levels of gender bias in our job posts, its correlation to specific gendered keywords or phrases, and the ultimate effect they have on the job application decision.

Our analysis generated several key findings.

(1) There is significant gender bias in job listings, but bias has been dropping significantly over the last decade, led by specific job sectors which now trend feminine.
(2) Changes in bias levels vary significantly over different sectors, driven by significant changes in usage of a small number of heavily gendered keywords.
(3) Our user study shows that users do indeed detect gender bias in job postings, consistent with bias detected by our algorithms.
(4) Observed bias still had low levels of impact on user decisions to apply or not apply for a specific position, and there was more sensitivity to bias by men than women.
(5) Surprisingly, we observed that users had strong internal biases which played significant roles in their decision on whether they would apply. These biases played a much bigger role than any gender bias language we observed.

To the best of our knowledge, our study is the first large-scale study to look at longitudinal shifts in gender bias in job postings. While our study has clear limitations (lack of historical advertisements from non-LinkedIn job sites, and potential sampling bias in our user study), we

believe our results shed light on an important component of the debate on gender equality in the workplace.

2 BACKGROUND

Understanding the evolution of characteristic gender differences in job listings helps to mitigate the effects of masculine and feminine stereotypes, thus reducing gender bias. Here, we provide background on gender disparity in the workplace, and previous efforts to detect gender bias in language.

2.1 Gender Equality in Job Market

Historically, certain industries have been dominated by one gender over the other. Approximately one-third of men and women work in occupations with a workforce comprised of at least 80% males or females, respectively. In the past, men tended to dominate engineering and construction occupations while women consistently dominated clerical assistant and teaching occupations [32, 33]. Although census data shows an increase in overall participation of women in the workforce throughout recent decades, the disparity among genders across particular industries remains over time [16].

One reason for such disparity lies in people's stereotypes of genders and occupations. Research shows that people are most likely to seek out occupations that are compatible with one's sense of self [19], suggesting that people are less likely to seek out occupations in industries dominated by the opposite gender. Moreover, a study of perceptions across 80 occupations found that people assume stereotypes associated with the gender of a worker must correlate with the requirements of their occupation. In particular, both genders perceived that masculine physical and agentic qualities were associated with more prestige and earnings [11]. Across industries, managerial positions have historically been perceived as requiring masculine traits. Men even view women negatively for displaying masculine traits in a management role because it is regarded as inconsistent with female role expectations [3]. In the field of Information Technology (IT) and Information Systems (IS), general stereotypes skew towards masculine traits, due to men consistently dominating the field [24, 45].

Another source of gender disparity is institutional discrimination in job markets, which has been observed in both traditional [6, 15, 30, 34] and online settings [21]. In a lab experiment that simulated a hiring decision process, male participants displayed a strong tendency to choose male candidates, even if a female candidate appears as a slightly better performer [15]. In another field study, comparably matched men and women are sent to apply for jobs in restaurants, and the study found female applicants were significantly less likely to get an offer from high-end restaurants [30]. Later, a similar study was conducted in a male-dominated occupation (engineer) and a female dominated occupation (secretary). Results show statistical significant discrimination against women in the male-dominated occupation and against men in the female-dominated occupation [34].

Another line of research shows that significant improvement in gender equality has been made over of last two or three decades [7, 10, 28]. The overall wage gap between two genders has declined considerably [7], and no institutional discrimination can now be observed in academia [10, 46].

2.2 Detecting Gender Bias by Word Analysis

Gender stereotypes are embedded in language use, *i.e.,* different word usage patterns when writing about males or females. Significant prior research used text analysis to examine gender differences in a number of contexts. Some examined how men and women use language differently in text and conversation [31]. Other work studied how text analysis algorithms express unintentional bias, and detect occupational stereotypes in text, *i.e.,* suggesting sexist analogies such as men

are analogous to computer programmers and women analogous to homemakers [8]. Other work showed when writing recommendation letters for faculty positions, more standout adjectives are used to describe male applicants than female applicants [38], and women are described as more communal and less agentic (assertive or competitive) [25]. In the context of job advertisements, researchers have shown that language used not only reflects such stereotypical views, but also reinforces the imbalance [2, 5], and that a conscious effort toward gender-fair language can help reduce it [23, 40]. Finally, much of the prior work in text classification rely on a supervised model with pre-labeled datasets, and is summarized nicely in a survey by Aggarwal et al [1].

Stereotypes can be captured by *gendered words* – terms describing socially desirable traits and behaviors of male or female genders [4, 36]. Gendered words are usually extracted from self-reported characteristics through questionnaires given to college students to measure their self-concept and valuation of feminine and masculine characteristics. The Personal Attributes Questionnaire (PAQ [42]) and Bem Sex Role Inventory (BSRI [4]) are two of the most representative questionnaires in early studies. The words extracted from BSRI and PAQ more typically associate females with more *communal* attributes (*i.e.,* gentle, warm) and men with more *agentic* attributes (*i.e.,* aggressive, competitive). Others generalized gendered words into *expressive* and *instrumental* traits [39]. Tying these together, aggregated lists of masculine and feminine characteristics have been compiled from previous studies, particularly through gendered wording in job advertisements [17]. Finally, Donnelly *et al.* found that women in recent years are less likely to endorse traditionally feminine traits in BSRI [14], indicating that gender norms may require an update of the masculine and feminine stereotyped characteristics.

Based on the lists encoded by Gaucher *et al.* [17], online services like Unitive[1] and Textio[2] use the words and phrases as a baseline to develop gender-neutralizing algorithms with help from machine learning classifiers. The algorithms are trained on internal application and hiring data, and aim at finding gendered wording in job advertisements before recruiters post online. These services represent the state-of-the-art for identifying gendered wording in job advertisements. Since these services run commercial, proprietary algorithms, it is cost prohibitive to evaluate our large job post dataset through their services. Instead, we designed our own algorithms using similar methodologies, and validate them against these online services using samples of test data.

2.3 Comparisons to Prior Work

The focus of our study is using large-scale data analysis to characterize gender bias across a comprehensive, longitudinal dataset. Our work was initially motivated by [17], which studied 4,000 job listings (most in a university setting) to characterize gender bias in job listings as an institutional-level mechanism of inequality maintenance. In contrast, we broadly characterize gender bias at scale, using a large dataset that consists of 17 million online job ads covering more than 140 industries. Our work also focuses on examining shifts in gender bias over ten years, and the impact it has on potential applicants' decision.

More recent work [21] identified gender/race discrimination on (performance) reviews in the online freelance (gig) marketplace, by correlating the review with the worker's gender and race. While their work targets reactions to worker output, ours focuses on job advertisements written by only the employer. While [21] analyzed keywords in review, their analysis was limited to sentiment analysis that identifies the attitude of the review, not gender bias.

---

[1]http://www.unitive.works/
[2]https://textio.com/
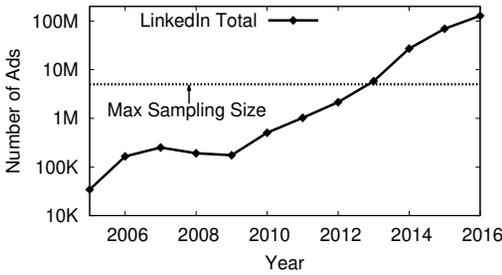
(a) Number of job ads per year
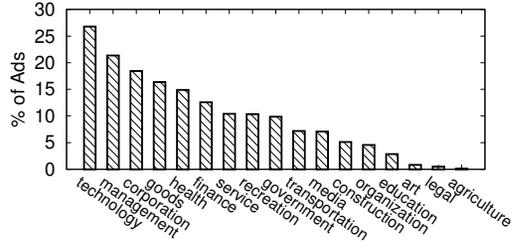
(b) Breakdown of sector groups
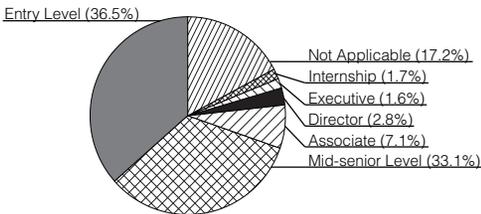
(c) Breakdown of employment type
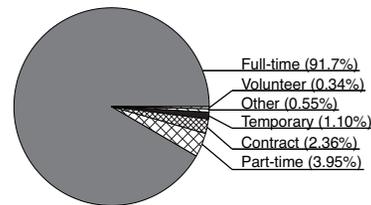
(d) Breakdown of seniority level

Fig. 1. Statistics of the LinkedIn dataset, 2005–2016.

3 DATA AND INITIAL ANALYSIS

We describe our data collection methodology and datasets, conduct preliminary analysis on our data, and present basic statistics to provide context for further analysis.

3.1 Data Collection

We collected a large sample of job advertisements from LinkedIn job posts over 10 years (from 2005 to end of 2016). LinkedIn job advertisements are fully public, and accessible online to any user without requiring account registration with LinkedIn. To retrieve a job advertisement, we simply queried known URLs and downloaded the webpages. LinkedIn keeps job advertisements available online for browsing even after the application window has closed. This allowed us to collect a significant longitudinal job advertisement dataset.

Job advertisements on LinkedIn are each assigned a unique ID, which increases monotonically over time. By the end of 2016, the maximum possible ID on LinkedIn reached above 253M, which means there are at most 253M job posts on LinkedIn. Since we limited our online query rate to avoid overloading LinkedIn's online services, we did not crawl all 253M job postings. Instead, we randomly sample 5 million job post IDs from each year, and only fetch job advertisements matching these IDs. Note that job listings in each year before 2013 contained less than 5 million ads. For these years, we fetched all available job advertisements. After filtering out job advertisements in languages other than English, our dataset contains 17,376,448 job advertisements in total.

Each job advertisement contains a job title, company name, company location, and the main descriptive content of the advertised position. In addition, LinkedIn also provides metadata, including job industry, job function, employment type, and seniority level. LinkedIn has 147 unique job industries[3], which are then further mapped into 17 sector groups, and 35 job functions which describe what activities a person is undertaking. Employment types includes 6 categories: full-time, part-time, temporary, contract, volunteer and other. Seniority level indicates the rank of the position within the business, ranging from entry-level (lowest) to executive (highest).

3.2 Preliminary Analysis

Number of Job Advertisements. We plot the number of LinkedIn job advertisements posted per year in Figure 1(a), as inferred by the total number of possible job IDs in LinkedIn matching a given year. For years up to 2013, our dataset closely follows that of the plotted LinkedIn results, with a small number of missing posts due to non-English listings and unavailable data errors for some of the oldest job postings (likely due to corrupted data). For years 2013–2016, we limited our sample set to 5 million postings per year. While our dataset captures only a limited sample from 2013 onwards, we believe a randomized sample set of 5 million ads is sufficient to capture a representative sample of job postings in any given year.

Sector Groups, Employment Type and Seniority Level. Next, we plot distribution of important metadata fields in Figure 1(b-d). Figure 1(b) shows the number of job posts in different job sector groups[4]. We found significant variation among the sizes of different groups: over 25% job posts belong to the largest group (technology) while less than 1% job posts in smallest group (agriculture). As for employment type (Figure 1(c)), most (91.7%) job listings seek full-time employment, while the rest are mostly split by part-time, contract and temporary (*i.e.,* seasonal). After 2013, LinkedIn introduced Volunteer as a new job sector, which accounts for a negligible portion of our total dataset. For seniority level (Figure 1(d)), we found a trend of fewer number of applicable job advertisements at higher levels of seniority. This matches our intuition about hierarchies in the job market.

4 QUANTIFYING GENDER BIAS

Our goal is to perform a large-scale analysis of the presence of gender bias over a large corpus of job listings covering the last 10+ years. Our first task is to develop a scalable algorithm to accurately quantify gender bias. In this section, we describe how we emulate the gender bias detection algorithms of two state-of-the-art recruitment assistance services, *Textio* and *Unitive*. We validate our approach by comparing our results against theirs on a small sample of 8,000 job ads. We follow up these results in the next section with a confirmatory user study.

4.1 Gender Bias Detection Algorithms

To develop a scalable gender bias detection algorithm, we start by developing metrics to accurately capture different aspects of gender bias. For guidance, we look to the two state-of-the-art recruitment assistance services that measure gender bias, Textio and Unitive. Textio and Unitive are the two largest web services today designed to help potential employers write better job advertisements. Each company curates their own algorithm to calculate whether a given job advertisement expresses masculine, feminine, or gender neutral language. The algorithms draw from an established baseline starting with gendered word lists [17]. We observe and adopt the metrics used by these two services, which we refer to as *Gender Target* and *Gender Tone*. Gender target follows Textio's methodology,

---

[3]https://developer.linkedin.com/docs/reference/industry-codes
[4]If a job belongs to multiple groups, we count the job in each group it belongs to.

(a) Textio score and Gender Target

(b) Unitive score and Gender Tone
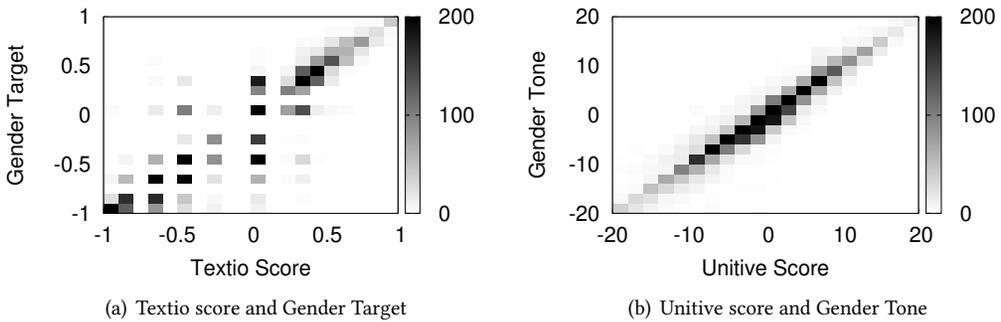
Fig. 2. Comparison of actual and approximated gender-bias metric.

which measures the intended audience gender reflected by a job advertisement, and falls into the range of -1 to 1, where -1 means the advertisement specifically targets male applicants, 1 means the advertisement specifically targets female applicants, and 0 means no gender preference is detected. Gender tone follows Unitive's methodology, which captures the extent to which a job advertisement is feminine- or masculine-phrased. It captures a cumulative effect, thus has no fixed range. A gender neutral advertisement has a tone of 0; the more masculine traits stated, the more negative the tone score is, and the reverse for feminine traits. We use the term *gender score* to refer to both metrics.

Gender target highlights feminine and masculine language in job listings. We calculate gender target by first calculating the number of gendered words, with feminine and masculine words canceling each other out. Job ads containing more feminine words than masculine words are considered to be targeting a female audience, and a final score is calculated by applying a sigmoid function on the remaining word count. The same procedure applies when masculine words outnumber feminine words, except that the result of this sigmoid function is reversed to fit into the range of -1 to 0. Finally, when the job ad has a perfectly balanced word count, it is considered to be gender neutral, with gender target score of 0.

In contrast, when calculating gender tone, we first categorize terms as *inclusive* (appealing) or *exclusive* (problematic). Prior research [3, 17] has shown a direct correlation with feminine bias from inclusive language, and between masculine bias and exclusive language. In addition, instead of simply counting words, we weights each words based on how gender specific they are. For example, the word "guy" carries a stronger gender implication than "ambitious." Thus, before calculating a cumulative score, gendered words are assigned weights depending on strength of their tone. A strongly masculine word has a strong negative weight, whereas a slightly feminine word has a weakly positive weight. We then add up the weights of all gendered words used in the ad.

Our first key challenge is obtaining an up-to-date list of biased words. To begin, we extracted 50,000 words with the highest frequency from all English LinkedIn job posts we collected, which cover 97.2% of all word occurrences. Since both services highlight words we can categorize with feminine or masculine bias, we queried the services with these words embedded in text, and examined the feedback. Textio annotated 296 gender-related key words, 150 masculine and 146 feminine. We also obtained 843 weighted keywords along with their weights annotated by Unitive, 445 with positive weights (feminine tone), 398 with negative weights (masculine tone). Note that since these two services picked their keyword independently, only 102 words overlap across services.
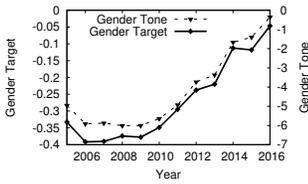
Fig. 3. Average gender bias score from 2005 to 2016.
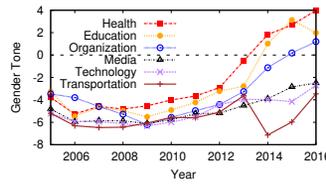
Fig. 4. Gender tone over time of top 3 and bottom 3 sectors in 2016.
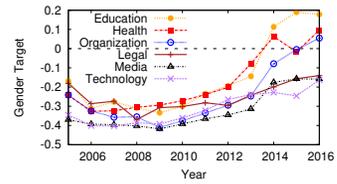
Fig. 5. Gender target over time of top 3 and bottom 3 sectors in 2016.

4.2 Algorithm Validation

We validate how well our techniques emulate these services, by comparing our results against theirs. We randomly selected 8,000 job advertisements from our dataset, and uploaded them to Textio and Unitive using their free online accounts. We compared the results they return to results from our own algorithms. We plot our results against results given by Textio and Unitive in Figure 2. For Textio, the scores are more scattered due to the use of discrete count before normalization. We found that 71.8% of the gender target scores are within a difference of 0.1 from Textio scores (an error rate of 10%). In the case of Unitive, the scores match along a straight line with a slight bias of -0.209 towards masculine tone, and 77.5% of the scores have an error of 1 or less. Since Unitive scores varied by as much as 10, this also represents an error rate of roughly 10%.

The error in our scores is due our inability to recover full keyword lists from both services, especially for phrases. Given the highly subjective nature of gender bias, our goal is not to generate "perfect" algorithms, but to obtain general and scalable algorithms with results that approximate public systems.

5 LONGITUDINAL ANALYSIS

In this section, we apply our gender bias detection algorithms on to our LinkedIn job post dataset. Our results show that in recent years, wording in job advertisements skews masculine, but the absolute level of bias is becoming more neutral. First, we identify a few factors that contribute to the trend. Different job functions across industry sectors distribute unevenly in terms of gender score. Although this uneven distribution of jobs across sectors results in an overall averaging of feminine and masculine bias scores, the effect is limited. Second, the masculine bias comes primarily from formal and long-term employment jobs, and appears more severe in senior level positions. Over the 11-year period, the number of entry-level jobs posted is increasing over time, which partially accounts for the decreasing masculine bias, as these positions predominantly skew feminine. Third, to quantify the effect of these factors, we formulate a regression analysis to predict gender score, which shows that the effect of all factors are significant. However, there is still an underlying trend of decrease masculinity after separating out the effect of these factors, indicating possible increasing awareness of using more gender neutral language. Finally, we identify a few gendered words that contribute the most in driving change in levels of gender bias.

5.1 Gender Bias Over Time

We begin by studying how gender bias in job postings changes over time. We are interested in whether the market as a whole (and perhaps as a proxy for the general population), is becoming more aware of gender bias. We compute two values for each job posting: a *gender target* score and a *gender tone* score. For each year, we compute the average scores and standard deviation of all job postings collected from that year. Figure 3 uses a dual Y-axis to compare the average score of the
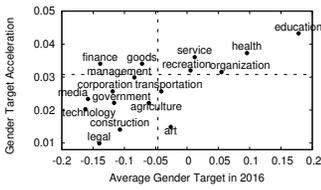
Fig. 6. Gender target in 2016 vs. gender target acceleration across sector groups.
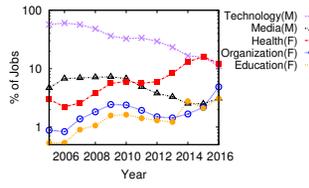
Fig. 7. Shift in distribution of 2 typically masculine and 3 typically feminine sector groups.
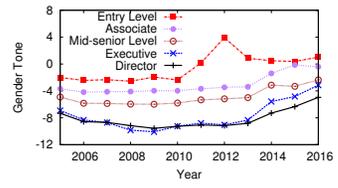
Fig. 8. Average gender bias score of different seniority level (*gender tone*).

two algorithms. The standard deviation values are similar over time and the two algorithms, thus they are omitted.

We make some key observations. First, the average gender scores remain consistently below 0 across all years, indicating that the job market, as captured by LinkedIn postings, is skewed towards masculine appealing positions. Second, an increasing absolute score over time suggests that the market is becoming more gender neutral. Third, while our two metrics use very different algorithms and their absolute scores are not directly comparable, their trends over time are almost identical. We performed another consistency check of these results using the gendered word lists from prior work [17], and the results are highly consistent. The frequencies of the three trends show strong correlation between each other (p-value < 0.0001), with Pearson correlation of more than 0.97. This confirms that the overall trend is fundamental to the job market, and the two metrics capture consistent views of the same phenomenon over time. To get a better understanding of where the masculine jobs originate and to explain the trend over time, we explore a variety of dimensions to better understand the underlying structure of the LinkedIn job market.

5.2 Gender Score over Job Sector Groups

We dive down to see how individual sector groups are changing over time with respect to gender bias. Recall that all together we have 147 distinct industries, mapped to 17 sector groups. While we have established that bias is decreasing over time for the entire job market, we want to observe any variance in dynamics across different job sectors.

We begin by looking at the top and bottom sectors sorted by gender scores. In Figure 4 and Figure 5, we plot the top 3 and bottom 3 sectors sorted by 2016 gender tone and 2016 gender target scores respectively. For each sector, we trace back their scores over past years. First, we note that gender tone and target are remarkably consistent in their top and bottom sectors. In both cases, Education, Health, and Organization are top sectors (more feminine), and all have risen consistently over time to current values above 0 (the dashed line represents the value 0). Media and Technology are the two sectors that appear as bottom sectors in both metrics. Their scores are rising, albeit at much slower rates, and occasionally experience short term dips (the start of the great recession 2008–2009). The Tech sector also experiences another dip around 2013–2015, showing that perhaps the most biased sectors might be more sensitive to economic downturns.

In Figure 6, we plot each sector's *acceleration* of gender scores over time, against its 2016 gender target value. Acceleration is computed as the slope of a linear regression of a sector's scores over time. The results are intuitive. The sectors slowest to reduce masculine bias (Tech, Legal, Construction) still have some of the most masculine biased gender target scores in 2016. Others like Education and Health have high rates of change towards more feminine wording, and as of 2016, are firmly on the side of feminine bias.
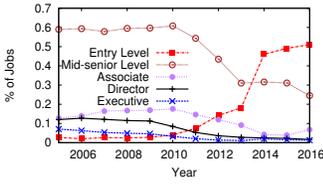
Fig. 9. Distribution of seniority level from 2005 to 2016.
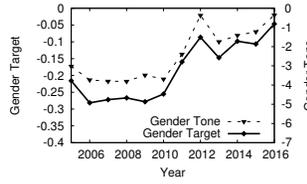
Fig. 10. Average gender score computed over distribution of seniority level in 2016.
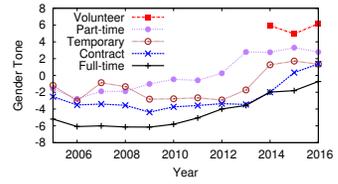
Fig. 11. Average gender bias score of different employment types (*gender tone*).

Dynamics of sector groups are correlated with the overall increasing gender score. One reasonable question is how much dynamics between sectors contribute to the overall gender bias trend. To answer this, we first calculate the ratio of gendered job postings inside each job sector. We find that the number of jobs is increasing in several stereotypically feminine sectors and decreasing in a number of stereotypically masculine sectors (see Figure 7). So, it is possible that the overall increasing gender score comes from a changing of sector distribution. To remove such effect, we recalculate the gender score across the entire 11-year period, but weigh each sector based only on the 2016 distribution of jobs across sectors. The result shows that the impact of shifting weights across sectors is small: gender tone only increases at most 0.56 under the new calculation (and gender target only increases by 0.034), much smaller than the overall increasing trend showed in Figure 3.

5.3 Gender Score over Seniority Levels

Next, we break down all job postings by their *seniority level*, and compute average scores in each category. Figure 8 shows the breakdown of results for different seniority levels, where there is a clear correlation between seniority ranking and the masculine tone of the job posting. We omit results of gender target here, since they show very similar trends. These results are consistent with prior work [3] that discusses how men hold an overwhelming majority of top management positions, and thus masculine traits are commonly associated with these higher-ranking positions. Due to a phenomenon called *ambivalent sexism* [18], attitudes regarding gender roles presume women historically belong in a domestic setting and are incompetent at holding positions of power. These unconscious biases may persist today, and are likely used to explain the gap in gender participation rates at more senior level positions.

Similar to trends across sector groups, we also find the distribution of different seniority levels changes over time (shown in Figure 9). It is clear that the increase in number of entry-level jobs corresponds to a decreasing proportion of mid-senior level jobs over time. Since entry-level jobs tend to be less biased towards masculine, the shift in distributions affects gender score. In Figure 10, we show the effect of removing this factor by computing the gender score using a fixed seniority level distribution from 2016. Compared to Figure 3, we get a similar increasing score trend, with a much smaller magnitude of masculinity. Thus, we conclude that the overall increasing trend comes from two parts: increasing lower-level jobs and increasing feminine language in each seniority level.

5.4 Gender Score over Employment Types

Figure 11 shows how the gender score is distributed over different employment types with respect to *gender tone*. Gender scores show clear and consistent trends in different employment types, and the more formal and long-term the employment, the more masculine tone in the job posting. Since

Feature	Coefficient	Feature	Coefficient
(Intercept)	-1.1380	Agriculture	-1.1434
Year after 2005	0.3064	Construction	-0.9402
Entry Level	2.9893	Management	-0.3033
Associate	0.6209	Finance	-0.3763
Mid-senior Level	-0.6825	Art	-0.0245*
Director	-4.6385	Organization	0.2408
Executive	-3.1562	Corporation	0.2867
Volunteer	2.4624	Legal	0.3372
Part-time	-1.8491	Recreation	0.4573
Temporary	-2.7402	Service	0.8327
Contract	-4.4472	Government	1.2678
Full-time	-5.0084	Goods	1.6498
Transportation	-3.6283	Education	2.4565
Technology	-1.5921	Health	3.9049
Media	-1.0557		

Table 1. Ordinal regression using *gender tone* as dependent variable. $p < 0.001$ applies for all entries except *, which has $p = 0.273$.

Feature	Coefficient	Feature	Coefficient
(Intercept)	-0.0510	Transportation	-0.0458
Year after 2005	0.0208	Corporation	-0.0432
Entry Level	0.1000	Legal	-0.0414
Associate	-0.0193	Construction	-0.0303
Mid-senior Level	-0.0885	Finance	-0.0243
Director	-0.2275	Management	-0.0156
Executive	-0.1589	Government	-0.0041
Volunteer	0.0442	Service	0.0150
Part-time	-0.1497	Organization	0.0203
Temporary	-0.1583	Goods	0.0328
Contract	-0.2670	Recreation	0.0401
Full-time	-0.2848	Art	0.0621
Technology	-0.0731	Health	0.1345
Media	-0.0620	Education	0.1974
Agriculture	-0.0571		

Table 2. Ordinal regression using *gender target* as dependent variable. $p < 0.001$ applies for all entries.
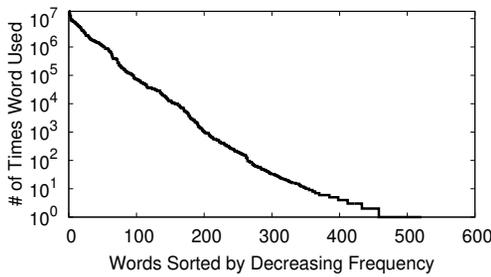
Fig. 12. Distribution of word frequency.
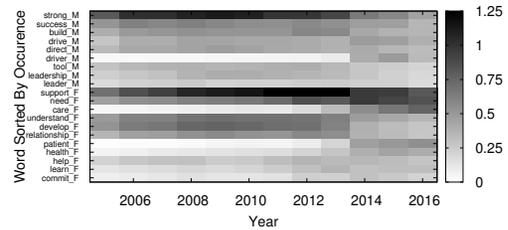
Fig. 13. Evolution of word frequency for top 20 words.

over 90% of jobs are full-time jobs across all years of our dataset, we do not investigate the effect of changing distribution in terms employment types.

5.5 Comparison of Gender Bias Contributors

Finally, after observing how these different factors affect the job market, we aim to quantify the effect of each factor. To do so, we formulate a regression analysis. We use seniority level, year, sector group and employment type as independent variables, to predict the gender score of a job advertisement. The reference categories for seniority level and employment type are "N/A" and "other," respectively. Since a single job can belong to multiple different sectors, there is no redundancy in sector groups.

Our result is shown in Table 1 and Table 2, for gender tone and gender target, respectively. We find that all the independent variables have statistically significant effects on gender score. The effects are consistent with the previous qualitative analysis, *i.e.,* gender scores vary over groups, and decrease with higher seniority level and more formal employment. However, after ruling out the effect of these factors, we still find an underlying increasing trend that is statistically significant. Although there could be other factors, we believe that awareness of using more inclusive language, and/or using less masculine language, is an important part of the change.

5.6   Changes in Word Use

Finally, we are interested in understanding how different words and phrases vary in their contribution to gender bias over time in job listings. We take the 500+ gender biased words from our dictionary, and plot their frequency of appearance (and therefore impact on gender scores) in Figure 12. We find that the frequency distribution of the most popularly used terms is exponential (and therefore it appears linear on a log plot).

Figure 13 plots the frequency of usage for top 20 words across the years, where we group the masculine and feminine words separately in the figure. We see that most of the masculine words show a stable or slightly decreasing trend[5], while the feminine words display more dramatic changes over time. Specifically, the most used masculine word, "strong", experienced a sharp decrease, potentially due to the growing awareness of biased wording. Among the top feminine words, "care," "patient" and "health" display significant increase, possibly driven by the growth in health-related jobs, while "understand," "develop" and "relationship" show a visible decline over the years. We also repeated our study on words with the largest changes in usage frequency, which produced similar results.

Finally, we compute the frequency of usage for each word across the years, and fit a trend line using linear regression. Of these terms, 145 words show statistically significant trends (p-value < 0.05). 75 of these are masculine toned words, and they are evenly divided between those growing in frequency and those dropping in frequency. Of the remaining 70 feminine toned words, the large majority (84%) showed an increasing trend.

6   USER STUDY: IMPACT OF GENDER BIAS

So far, we have quantified the level of gender bias in job listings over time, but we do not yet understand how these gender biases actually impact users (potential job applicants). To answer this question, we conducted a user survey, which we describe here. In short, we find that gender scores from our algorithms properly reflect perceived gender stereotypes associated with job postings, but that biased wording has limited effect on the perception of a job, compared to respondents' preconceived notion of the job type. We also find that male respondents are less willing to apply for stereotypically feminine jobs, while the reverse does not hold.

Survey Participants.   We recruit survey respondents from two different sources, Amazon Mechanical Turk (MTurk), and undergraduate students from UC Santa Barbara. In total, we received results from 469 distinct MTurk workers and 273 students, each job advertisement is evaluated by at least 20 different workers and 12 different students. Undergraduate students were volunteers who received necessary credit for their course work. Each worker was compensated $1 for finishing the task. To ensure the quality of replies, we require workers to have an 80% HIT approval rate, and have at least 50 HITs approved in the past. We also include a quality check (*i.e.,* gold standard) question in our survey question list, (*e.g.,* "Please answer A and D for this question.") to avoid low-quality/non-responsive workers. For respondents who failed our gold standard questions, their responses are not included in our analysis. In most cases, responses from MTurk workers and the students point to the same conclusion, and we thus combine their answers in such analysis. In cases when the responses differ, we analyze the respondent pools separately.

The demographics of our survey participants are as follows. Among all 469 Mturk workers, 54.6% indicated male and 45.4% indicated female. The majority of participant ages fall into the ranges of 21 to 30 (38.8%) or 31 to 40 (34.3%), with 1.49% younger than 21, 11.7% older than 50, and the rest fall between 41 and 50. Most participants work full-time (67.2%); and most hold a Bachelor's

---

[5]The only exception is "driver." This is probably because driver can also be a job title in the transportation sector which experienced a rapid growth since 2014.

(42.9%) or a Master's (24.5%) degree. Among the 273 college students, 201 (73.6%) are female and 72 (26.4%) are male. 209 (76.6%) of the students are younger than 21, and 64 (23.4%) are from 21 to 30.

Methodology. In our user study, we divided job advertisements into 3 categories: *masculine jobs*, *feminine jobs*, and *gender neutral jobs*. Feminine jobs are randomly sampled from job advertisements with the highest 10% gender score, as scored by both gender target and tone. Masculine jobs are similarly sampled from postings with the lowest 10% gender score, and neutral jobs are sampled from advertisements with scores nearest 0. We did not restrict the timespan of our advertisements, since we want to maximize user reaction on potential gendered language. When older job advertisements is included in our sample, we manually check their content to make sure there are no outdated words that will significant affect user's reaction.

For each advertisement, we created a second version of the job description by replacing keywords or phrases marked as gendered language from our dataset with more gender-neutral words or phrases not in our dictionary of gendered words. For example, substitutions included "workforce" replaced by "employees," and "collaborating" replaced by "working." Although we made efforts to consistently replace biased words with the same neutral alternative, some instances required more dynamic replacements to retain the readability and intent of the original post. For instance, depending on context, "engage" may be replaced with "participate," "employ," or "work." Other words, like "please," were simply removed. To ensure the biased language was not just replaced with different biased language, we calculated gender scores for the edited descriptions, and verified that the substituted descriptions received gender-neutral scores.

We replaced or removed as much gendered words and phrases as possible without changing the intended meaning of the original job posting. This provides a suitable baseline to isolate the impact of gender wording from people's inherent biases and stereotypes. This contrasts with prior studies [17] that analyzed masculine and feminine language without comparing against neutral wording as a baseline.

We asked each user to read three job advertisements, one from each category. After reading through the ads, we gathered their responses to the following questions:

- *Q1*: If you were fully qualified to apply for a job like this, how likely is it that you would apply for this particular position? Answers are measured by 5-level Likert Scale (1 indicates definitely would not apply and 5 indicates definitely would apply).
- *Q2*: By looking at the job description, what would you think to be the percentage of women currently working in this type of position?
- *Q3*: While reading the job description, to what extent did you feel that the advertisement would attract more male or more female applicants? Answers are measured by 5-level Likert Scale (1 indicates job attracts mostly males and 5 indicates job attracts mostly females).
- *Q4*: Please mark any words or sentences that you do not feel comfortable with.

Q1-Q3 in the above questions are multiple-choice, and Q4 is open-ended. At the end of the survey, we collected user demographic information. We consulted our local IRB[6] and obtained approval before conducting the user study. Note that while Q3 could be asked differently, *i.e.,* , ask users to rate the attractiveness of the job and compare results between male and female respondents, we chose this version so users would focus on the effect of wording rather than allowing other, random or uncontrolled factors to influence their "broad" evaluation of a post.

The three job advertisements were randomly selected from the three categories (masculine, feminine, neutral), with equal likelihood of choosing an edited or raw version for each category. We used a pilot test of Amazon Turk users to determine if question order impacted user response.

---

[6]The IRB protocol number is 29-17-0259.
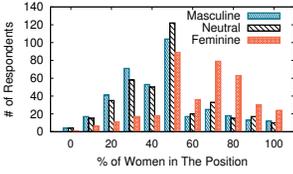
Fig. 14. Responses on Q2 (percent of female in the position).
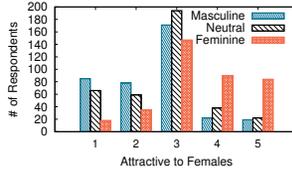
Fig. 15. Responses on Q3 (attracting female applicants).
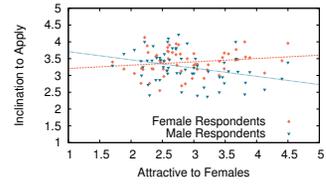
Fig. 16. Average user response on Q1 against average user response on Q3.
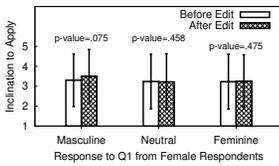
(a) Female respondents
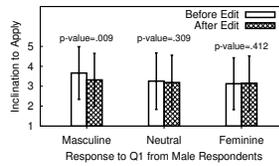
(b) Male respondents

Fig. 17. Responses from university students on Q1 (inclination to apply), breakdown by types of jobs. Differences measured with Mann-Whitney U-test.
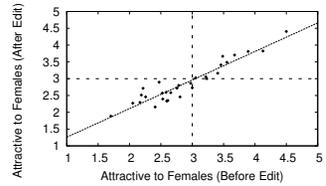
Fig. 18. Average answer to Q3 (attracting female applicants) before and after word substitution.

After controlling for other factors, results showed a decreased likelihood of job application for the second and third ads. Thus we presented the three sampled advertisements to participants in random order.

6.1 Preliminary Task: Do gender scores accurately reflect user perceptions of gender bias

We first validate our algorithms designs with the user study. Here, we compared user responses to questions Q2 and Q3 for different types of job advertisements before replacing or removing gendered words. The results are shown in Figure 14 and Figure 15. In each figure, we analyzed responses for each question choice, across all advertised job positions. We find that people presume more female workers in supposed feminine positions, suggesting feminine toned job advertisements appear more attractive to women. Corresponding findings apply to masculine toned job advertisements, as well, thus validating our algorithms.

We also conduct Mann-Whitney U-tests on the distribution of responses between groups, and the difference is statistically significant with p-value less than 0.001 across all types except between masculine and neutral ads. Interestingly, all distributions show an artificial peak at the most neutral answer. Many respondents selected 50% for Q2, where we asked about the percentage of women working in the position, and for Q3, many selected the option "attracting male and female applicants equally." Respondents who selected the neutral choice often provided reasons related to "equal chance" or "equal rights," showing a conscious awareness of, or desire for, equal gender participation in the workplace.

6.2 Principle Task: Quantifying effects of gendered wording on job application rates

Perceived occupational gender bias affects decisions to apply.   Given the correlation between word choice and perceptions of gender bias, we next sought to examine the extent to which this perceived bias influences one's decision to apply. We show that male respondents express noticeably diminished inclination to apply for jobs perceived to predominately attract females. We quantify the level of perceived gender bias by averaging responses to Q3. In Figure 16, we plot the perceived bias of a job against the average tendency of female and male applicants to apply, indicated by average response to Q1. By applying linear regression on both male and female applicants, we discovered that female applicants do not show any preference with respect to gender distribution, with near zero slope (0.0989) and a p-value of 0.190. Meanwhile, male applicants displayed a preference against applying for female-dominated jobs, with a slope of -0.245 and a p-value of 0.0113. This contradicts prior work [17], where female applicants found masculine worded occupations significantly less appealing. One explanation is that our gender bias is naturally embedded in the job posts, and thus likely to be of a lower intensity than artificial job ads composed specifically to contain gender bias. Additionally, female perception of and reaction to gender bias may have shifted since the 2011 study.

From Figure 16, we can observe a high degree of variance, indicating that willingness to apply for a job may be affected by other external factors besides gender neutrality. When we asked their reason for why they will or will not apply for a position, we found a few frequently mentioned reasons, including the anticipated salary, benefits, location, workload, potential of career development, and whether the field of job appeals to the respondent.

Changes in gendered wording have limited effect on predisposition to apply.   The ultimate question remains as to whether a recruiter can change the wording in a job advertisement and increase the likelihood of potential applicants to apply for the job. Thus, we seek to quantify the causal effect of wording on users' decisions to apply for a job or not. For female and male applicants, we compared the predisposition to apply for a job, measured by averaging answers to Q1, before and after word substitution.

These results are different between the two pools of respondents. For students, wording change in masculine-worded ads does affect application decisions, as shown in Figure 17. Removing male-biased words from job advertisements leads to less male applicants and marginally more female applicants expressing an inclination to apply. When performing Mann-Whitney U-test on responses of MTurk workers, the p-value are above 0.05 for all three job types with both male and female respondents.

This shows that the effects of word use are observable, but somewhat limited. We then sought to break down the effect, pinning down whether wording actually *causes* a perceived bias, by comparing respondents' reported perception before and after word substitution.

We plotted the average responses to Q3 before and after word substitution. If wording is the sole cause for gender bias, then by removing the biased words, all jobs advertisements should appear with similar level of perceived gender bias, thus yielding a slope of 0. In contrast, if wording has zero impact on gender perception, it will show a slope value of 1. In Figure 18, we can see that the perceived bias persists even after word substitution, with a linear regression yielding a slope of 0.850 and p-value of 0. Similar results are observed for Q2, showing a slope of 0.825 and p-value of 0. This indicates that there certainly exist other properties affecting gender perception more influential than changes in wording.

Preconceived notions of occupations predominately affect user perceptions.   To better understand what factors influence user perception, we examined the reasons given in our survey responses. In the survey, we asked users to explain their reasoning and mark any words or phrases

in the job advertisements that made them feel uncomfortable. With this exercise, we hoped to gain insights into current perceptions that may be missing from previous studies or even current available services.

We found that many explanations given in our responses include preconceived ideas of the described job function. For example, in response to a job providing technical support for customers in a cable television company, one user believes that 20% of the workers in this position consist of women and therefore presumes the position will attract primarily male applicants, indicating as the reason "It's a technology job." Some respondents even expressed strong gender stereotypes, making statements like, "Low wage jobs tend to hire women, men try to get better jobs." Similar stereotypes also affected users reading posts for jobs perceived to be suitable for female applicants.

Many responses associated a particular gender with specifics characteristics they assumed would best fit the job. For instance, some highlighted phrases such as "bringing accountability, decency, and humor to the job," with explanations stating that these expectations would appeal primarily to male applicants, and women may not like the position. We infer that these users think women are less likely to possess such attributes, making them unfit for the job. These words were not included in our gendered language dataset, indicating modern perceptions of occupational gender bias.

Other responses focused more on preconceived notions of job functions. One response to a position requiring business travel with the company CEO described how they couldn't imagine a man doing this kind of assistant job, demonstrating an inherent stigma against men performing clerical work. Other respondents indicated that non-assistant positions requiring travel or "with little supervision" were better fits for male applicants who may feel more comfortable traveling than women, perhaps due to traditional views that women should or would want to stay at home. Additionally, users suggested that job descriptions requiring an ability to lift up to 50 lbs. or unloading trucks skewed towards male applicants who would be more likely to be capable of such physical activity. On the other hand, many male respondents expressed no interest or consideration for a beauty consultant position because they perceived it as a field of work for females, with some users describing a beauty school degree requirement as simply, "sounds sexist." Most surprisingly, several responses ironically stated that using the phrase "Equal Opportunity Employer" felt insincere and directly singles out females or minorities.

We originally intended to use these questions to better identify specific gendered words or phrases. Surprisingly, we instead gained insights about the role that inherent gender bias plays in the job marketplace.

7 DISCUSSION

Through our data analysis, we observed an increasing shift away from masculine-biased job postings over the years, and that employers today use less gendered wording than they did 10 years ago. However, the results of our user study also indicate that this trend towards gender neutral wording does not correlate with a perception of gender neutrality in the job market.

Surprisingly, user responses to our survey showed significant gender bias in the responders to specific job positions. Despite the correlation we found between gendered wording and perceived bias, users' explanations show their underlying biases were bigger determinants of their likelihood to apply than any gendered wording. Even after removing all gendered language from the job advertisements, these trends remained in the responses (see Figure 18). Gender bias is present in the responder's own perception, independent of the language used in job posts. The implication is that completely removing gendered wording will have limited impact in forming a more gender neutral workforce. This echoes observations made in prior work [21] that inherent user bias was pervasive in the job marketplace. Ultimately, we need to address inherent gender bias in the applicants themselves to significantly improve gender neutrality.

7.1 Limitations

Our methodology and data source introduced a number of limitations in our study. First, our data analysis shows a trend of increasing gender neutrality over the years, and we examined the impacts from different factors. However, all the observations are based on co-occurrences, so we cannot make claims of strong casual relationships in our high-level results. Second, our job postings dataset is from a single source, LinkedIn. While it is arguably the largest job listing site online, it is still prone to hidden biases, *e.g.,* towards US based positions or more technology sectors. Unfortunately, other large job listing web services like Monster or Indeed do not provide historical job postings for analysis.

For our user study, there are potential biases in our sampling. In our user study we recruited 25 users to evaluate each job advertisement, but we only studied 30 job advertisements in total. A small number of job advertisements may not represent the largely diverse pool of all the job advertisements. In our user study, we are limited to Amazon Mechanical Turk workers and undergraduate college students, neither of which are representative of a highly diverse workforce in the general population. In addition, since the workers do not necessarily evaluate jobs from their own area, some respondents expressed unfamiliarity with terminology (acronyms, corporate jargon) specific to the field of work described. Finally, answers to our questions showing gender bias may in fact be reflecting personal familiarity of respondents with assumed statistics in a given industry.

While quantifying and understanding these limitations will require further studies, we believe our analyses provides an early empirical perspective on the shifting dynamics of gender bias in the American workplace. We hope our work will encourage further studies of large-scale gender bias, and help identify the key factors that will lead to improved gender quality in the workplace.

ACKNOWLEDGMENTS

REFERENCES

[1] Charu C Aggarwal and ChengXiang Zhai. 2012. A survey of text classification algorithms. *Mining text data* (2012), 163–222.

[2] Inger Askehave and Karen K Zethsen. 2014. Gendered constructions of leadership in Danish job advertisements. *Gender, Work & Organization* 21, 6 (2014), 531–545.

[3] Leanne E Atwater, Joan F Brett, David Waldman, Lesley DiMare, and MaryVirginia Hayden. 2004. Men's and women's perceptions of the gender typing of management subroles. *Sex Roles* 50, 3 (2004), 191–199.

[4] Sandra L Bem. 1974. The Measurement of Psychological Androgyny. *Journal of Consulting and Clinical Psychology* 42, 2 (1974), 155–162.

[5] Sandra L Bem and Daryl J Bem. 1973. Does sex-biased job advertising "aid and abet" sex discrimination? *Journal of Applied Social Psychology* 3, 1 (1973), 6–18.

[6] Marianne Bertrand and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* 94, 4 (2004), 991–1013.

[7] Francine D Blau and Lawrence M Kahn. 2016. *The gender wage gap: Extent, trends, and explanations.* Technical Report. National Bureau of Economic Research.

[8] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in Neural Information Processing Systems.* 4349–4357.

[9] Anna Brown and Eileen Patten. 2017. The narrowing, but persistent, gender gap in pay. http://www.pewresearch.org/fact-tank/2017/04/03/gender-pay-gap-facts/. (2017).

[10] Stephen J Ceci, Donna K Ginther, Shulamit Kahn, and Wendy M Williams. 2014. Women in academic science A changing landscape. *Psychological Science in the Public Interest* 15, 3 (2014), 75–141.

[11] Mary A Cejka and Alice H Eagly. 1999. Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin* 25, 4 (1999), 413–423.

[12] Jacob Clark Blickenstaff*. 2005. Women and science careers: leaky pipeline or gender filter? *Gender and education* 17, 4 (2005), 369–386.

[13] Karen L Cropsey, Saba W Masho, Rita Shiang, Veronica Sikka, Susan G Kornstein, and Carol L Hampton. 2008. Why do faculty leave? Reasons for attrition of women and minority faculty from a medical school: four-year results. *Journal of Women's Health* 17, 7 (2008), 1111–1118.

[14] Kristin Donnelly and Jean M Twenge. 2016. Masculine and feminine traits on the Bem Sex-Role inventory, 1993–2012: a cross-temporal meta-analysis. *Sex Roles* (2016), 1–10.

[15] Martha Foschi, Larissa Lai, and Kirsten Sigerson. 1994. Gender and double standards in the assessment of job applicants. *Social Psychology Quarterly* 57, 4 (1994), 326–339.

[16] Paul E Gabriel and Susanne Schmitz. 2007. Gender differences in occupational distributions among workers. *Monthly Labor Review* 130 (2007), 19.

[17] Danielle Gaucher, Justin Friesen, and Aaron C Kay. 2011. Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology* 101, 1 (2011), 109.

[18] Peter Glick and Susan T Fiske. 1996. The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology* 70, 3 (1996), 491–512.

[19] Linda S Gottfredson. 1981. Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling psychology* 28, 6 (1981), 545.

[20] The World Bank Group. 2017. Many societies gradually moving to dismantle gender discrimination, yet more can be done, says World Bank Group president Jim Yong Kim. http://www.worldbank.org/en/news/press-release/2013/09/24/societies-dismantle-gender-discrimination-world-bank-group-president-jim-yong-kim. (2017).

[21] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. 2017. Bias in online freelance marketplaces: evidence from TaskRabbit and Fiverr. In *Proc. of CSCW*.

[22] Karen Holtzblatt, Aruna Balakrishnan, Troy Effner, Emily Rhodes, and Tina Tuan. 2016. Beyond the pipeline: addressing diversity in high tech. In *Proc. of CHI Extended Abstract*.

[23] Lisa K Horvath and Sabine Sczesny. 2016. Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements. *European Journal of Work and Organizational Psychology* 25, 2 (2016), 316–328.

[24] Kshiti D Joshi and Nancy L Schmidt. 2006. Is the information systems profession gendered?: characterization of IS professionals and IS career. *ACM SIGMIS Database* 37, 4 (2006), 26–41.

[25] Juan M Madera, Michelle R Hebl, and Randi C Martin. 2009. Gender and letters of recommendation for academia: agentic and communal differences. *Journal of Applied Psychology* 94, 6 (2009), 1591.

[26] Cynthia M Marlowe, Sandra L Schneider, and Carnot E Nelson. 1996. Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology* 81, 1 (1996), 11.

[27] Lisa A McLoughlin. 2005. Spotlighting: Emergent gender bias in undergraduate engineering education. *Journal of Engineering Education* 94, 4 (2005), 373.

[28] David I Miller and Jonathan Wai. 2015. The bachelor's to Ph. D. STEM pipeline no longer leaks more women than men: a 30-year analysis. *Frontiers in psychology* 6 (2015), 37.

[29] Harriet Minter. 2017. Why so few women in tech? Your answer depends on your gender. https://www.theguardian.com/careers/2016/dec/12/why-so-few-women-in-tech-your-answer-depends-on-your-gender. (2017).

[30] David Neumark, Roy J Bank, and Kyle D Van Nort. 1996. Sex discrimination in restaurant hiring: An audit study. *The Quarterly Journal of Economics* 111, 3 (1996), 915–941.

[31] Matthew L Newman, Carla J Groom, Lori D Handelman, and James W Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* 45, 3 (2008), 211–236.

[32] U.S. Bureau of Labor Statistics. 1998. Median weekly earnings of full-time wage and salary workers by detailed occupation and sex. https://www.bls.gov/cps/aa1998/CPSAAT39.PDF. (1998).

[33] U.S. Bureau of Labor Statistics. 2017. Median weekly earnings of full-time wage and salary workers by detailed occupation and sex. https://www.bls.gov/cps/cpsaat39.htm. (2017).

[34] Peter A Riach and Judith Rich. 2006. An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market. *The B.E. Journal of Economic Analysis & Policy* 5, 2 (2006), 1–22.

[35] Cecilia L Ridgeway. 1997. Interaction and the conservation of gender inequality: Considering employment. *American Sociological Review* 62, 2 (1997), 218–235.

[36] Paul Rosenkrantz, Susan Vogel, Helen Bee, Inge Broverman, and Donald M Broverman. 1968. Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology* 32, 3 (1968), 287.

[37] Myra Sadker and David Sadker. 2010. *Failing at fairness: How America's schools cheat girls.* Simon and Schuster.

[38] Toni Schmader, Jessica Whitehead, and Vicki H Wysocki. 2007. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles* 57, 7-8 (2007), 509–514.

[39] Stephen A Schullo and Burton L Alperson. 1984. Interpersonal phenomenology as a function of sexual orientation, sex, sentiment, and trait categories in long-term dyadic relationships. *Journal of Personality and Social Psychology* 47, 5 (1984), 983.

[40] Sabine Sczesny, Magda Formanowicz, and Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology* 7 (2016), 25.

[41] Steven Davidoff Solomon. 2017. Why So Few Women Reach the Executive Rank. https://dealbook.nytimes.com/2013/04/02/why-so-few-women-reach-the-executive-rank/. (2017).

[42] Janet T Spence, Robert Helmreich, and Joy Stapp. 1975. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology* 32, 1 (1975), 29.

[43] Rhea E Steinpreis, Katie A Anders, and Dawn Ritzke. 1999. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles* 41, 7 (1999), 509–528.

[44] Donald Tomaskovic-Devey. 1993. *Gender & racial inequality at work: The sources and consequences of job segregation.* Number 27. Cornell University Press.

[45] Eileen M Trauth, Curtis C Cain, KD Joshi, Lynette Kvasny, and Kayla M Booth. 2016. The influence of gender-ethnic intersectionality on gender stereotypes about IT skills and knowledge. *ACM SIGMIS Database* 47, 3 (2016), 9–39.

[46] Samuel F Way, Daniel B Larremore, and Aaron Clauset. 2016. Gender, productivity, and prestige in computer science faculty hiring networks. In *Proc. of WWW*.