Does Mandating Women on Corporate Boards Backfire?*

Bo Bian† Jingjing Li‡ Kai Li§

Abstract

We examine the labor demand-side consequences of mandating female representation on corporate boards. Using California's SB 826 as an exogenous shock and applying computational linguistic methods to job ads, we find that the mandate significantly reduced treated firms' demand for female labor. We also show that SB 826 led to fewer female new hires, with both effects more pronounced in high individualism, Republican-leaning, and high masculinity counties/firms. Additionally, the mandate resulted in poorer workplace treatment of women and increased female employee turnover. Experimental evidence points to psychological reactance and perceived violations of gender norms as key underlying mechanisms. Our findings complement Bertrand et al. (2019), which documents limited spillovers from a similar quota in Norway, by uncovering evidence of backlash in less gender-equal contexts. These results highlight how misalignment between policy goals and prevailing social norms can undermine the intended goals of gender quotas, underscoring the need for context-specific policy design.

Keywords: Female Directors, Board Gender Quotas, Backlash, Job Postings, Employee Turnover, Employee Reviews, SB 826, Computational Linguistic Methods.

JEL classification: *G30, G34, J16, J21, J23*

*We thank helpful comments from Audra Boone, Xiao Cen, Alex Edmans, Marina Gertsberg, Jing Gong, Todd Gormley, Shenje Hshieh, Wei Jiang, Ambrus Kecskes, Lenny Kostovetsky, Lena Larkin, Guangli Lu, Hamed Mahmudi, Feng Mai, David Matsa, Anya Mkrtchyan, Anh Nguyen, Alexandra Niessen-Ruenzi, Lin Peng, Karin Thorburn, Felix von Meyerinck, Aazam Virani, Qiang Wu, Liu Yang, Xingtan Zhang, Jing Zhao, and seminar or conference participants at Baruch College, Central University of Finance and Economics, CKGSB, City University of Hong Kong International Finance Conference, KAIST, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, Korea University, KU Finance Conference, Northern Finance Association Meetings, Peking University, SFS Cavalcade North America, Stevens Institute of Technology, Texas A&M Young Scholars Symposium, University of Calgary, University of Hong Kong, University of Waterloo, Wilfrid Laurier University, and York University. We thank Ruiqi Ma and Yile Chen for excellent research assistance. Approval for the experiment in this paper was obtained from the Institutional Review Board (IRB) at University of Virginia under UVA IRB-SBS #6912. This study is registered with the American Economic Association RCT Registry (AEARCTR-0015331). We acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (Grant Number: 430-2021-00562). Bo Bian acknowledges financial support from the Phillips, Hager & North Centre at UBC. Kai Li acknowledges financial support from the Canada Research Chair in Corporate Governance. This version: July 21, 2025.

†University of British Columbia, Sauder School of Business, bo.bian@sauder.ubc.ca.

‡University of Virginia, McIntire School of Commerce, jl9rf@virginia.edu.

§University of British Columbia, Sauder School of Business, kai.li@sauder.ubc.ca.

1 Introduction

Since Norway became the first country to legislate board gender quotas in 2003, such mandates have remained a contentious tool for promoting gender equality. The primary concern centers on potential unintended consequences. Existing literature largely focuses on the labor *supply* side—such as the limited pool of qualified female candidates—and warns that quotas may inadvertently reinforce stereotypes, discouraging women from investing in their careers if advancement is seen as quota-driven rather than merit-based (Bertrand et al., 2019). However, the labor *demand* side remains underexplored and can be equally important. Quotas may provoke backlash, reducing the likelihood of hiring women outside the mandated roles and undermining the quality of their treatment in the workplace. In light of the recent retrenchment in corporate diversity initiatives,[1] our paper provides timely and granular evidence on how social policies such as board gender quotas can trigger unintended changes in employer behavior in ways that may ultimately hinder their broader gender equality goals.

The idea that quotas can trigger backlash traces its roots to the psychology literature on reactance theory (Brehm, 1966) and the influential economics literature on the interplay between laws and social norms (Benabou and Tirole, 2011; Bowles and Polania-Reyes, 2012; Acemoglu and Jackson, 2017; Wheaton, 2022). These studies argue that when legislation conflicts with individuals' preferred norms, those affected may respond by acting in *direct opposition* to the law in an effort to preserve their norms and restore a sense of autonomy.

To test for potential backlash, we empirically examine a range of labor demand-side responses, from hiring preferences to the post-hiring treatment of female employees, through the lens of California Senate Bill 826 (SB 826). SB 826 required that all publicly listed companies with headquarters in California (CA) have at least one female director by the end of 2019. While SB 826 served as a plausibly exogenous shock, assessing its effect on firms' demand for female labor is challenging. Firms are not required to disclose the gender composition of their workforce, and the observed female share reflects both labor supply and demand factors, making it difficult to isolate employer intent.

To address this challenge, we apply computational linguistic methods to a dataset of 30 million online job ads over the period 2014–2020, treating gendered language usage in job advertisements as a proxy for firms' demand for female labor, an established tool firms use to shape their applicant pool.[2] We begin with the gendered word list from Gaucher et al. (2011) and construct our proxy measure as the share of feminine words among all

[1] For example, setbacks at firms such as Meta and McDonald's and policy reversals such as the rollback of Nasdaq's board diversity disclosure rules. See Forbes (2025) and WSJ (2024).

[2] Several companies such as Textio, Applied, and Ongig, now offer professional services to help firms revise job ads, particularly by modifying gendered language to better attract their target applicant pool. Laboratory experiments by Gaucher et al. (2011) show that job ads with more masculine than feminine wording are perceived less appealing by female applicants.

gendered words in a job ad. Additionally, we develop alternative measures by fine-tuning pre-trained Bidirectional Encoder Representations from Transformers (BERT, Devlin et al. (2018)) models to quantify the degree to which a job ad is female-tilted.

We implement a difference-in-differences design, comparing CA firms (treated) to non-CA firms (control) around the passage of SB 826. At the job ad level, we find that the adoption of SB 826 led to a four-to-five percentage-point drop in the share of feminine words among all gendered words, representing roughly a 10% reduction of the mean share. This result is robust to using alternative measures based on fine-tuned BERT models. Restricting the comparison to job ads with identical job titles—thereby holding the nature of the job constant—yields similar findings. To strengthen identification, we show that the negative effects are primarily driven by firms that were not in compliance with SB 826 before its enactment. We also estimate the dynamic effect of SB 826 and document little pre-trends but a clear negative post-trend.

While using job ads provides a clean measure of firms' demand for female labor, one may question whether gendered language in job ads meaningfully influences application behavior or actual hiring outcomes. To address this concern, we analyze the gender composition of new hires by aggregating individual LinkedIn profiles into a firm-month panel. We find that SB 826 significantly reduced the share of female new hires by around one percentage point (relative to the mean share of female new hires at 41%).[3] Importantly, the backlash extends beyond hiring. Following the passage of SB 826, compared to non-CA firms, CA firms received more negative employee reviews and lower employee ratings on workplace accommodations that are particularly important for female employees. These firms also experienced more frequent and serious gender-related labor violations, as well as increased turnover among female employees.

Thus far, we document widespread negative effects of SB 826 across nearly every stage of a female employee's experience with her employer—from job application and hiring to workplace treatment, benefits, and turnover. In supplemental analysis, using staggered adoptions of board gender quotas in Europe over the period 2008–2020, we uncover a similar negative effect on female employment. What drives these outcomes? Is there evidence consistent with the microfoundations of backlash? In our context, external pressure to appoint more female directors may be perceived as a threat to individual autonomy, prompting pushback even among those who support gender diversity in principle. At the same time, such mandates may conflict with prevailing gender norms, leading to resistance that undermines broader organizational diversity efforts. These two mechanisms—*psychological reactance* and *norm*

[3] Using the average annual increase in the share of female labor force participation over the past seven decades as a benchmark, this decline implies a setback equivalent to up to four years of progress toward closing workplace gender gaps.

violation—can reinforce each other: when a mandate like SB 826 simultaneously threatens autonomy and clashes with gender norms, reactance is amplified and resistance intensifies.

Our first piece of evidence in support of these economic mechanisms comes from examining how SB 826 affected gendered language in job ads across regions with varying cultural, political, and gender views. We begin by showing that in counties with stronger individualistic norms, the decline in feminine language in job ads following the enactment of SB 826 was 30% greater than in counties with weaker norms. Since individualism is closely associated with autonomy and a preference for personal agency over externally imposed rules, this pattern is consistent with greater psychological reactance to top-down mandates such as SB 826.

Moreover, if pushback against women empowerment was driven by perceived conflicts between SB 826 and individuals' preferred norms, we would expect stronger effects in Republican-leaning counties and those with stronger masculinity norms. Consistent with this conjecture, we find that the negative effect of SB 826 was nearly 70% larger in job ads posted in Republican counties compared to those in Democratic counties. We further show that the negative effect was over 35% larger in job ads posted in counties with above-median masculinity scores compared to those in counties with below-median scores. Notably, this heteregeneity extends to hiring outcomes: the decline in female new hires was concentrated in Republican-leaning firms and those with predominantly male workforces. These cross-county and cross-firm patterns provide suggestive evidence that resistance on the labor demand side was driven by reactance as well as a perceived misalignment between board gender quotas and prevailing social norms.

Our second piece of evidence on the underlying economic mechanisms comes from a randomized controlled experiment that elicits participants' views on board gender quotas. This approach provides direct evidence for our proposed mechanisms while ruling out alternative mechanisms. In the experiment, we manipulate the presence of a board gender quota policy to survey participants. We then ask sets of questions related to both our proposed mechanisms: psychological reactance and norm violation, and two alternative mechanisms: moral licensing—the idea that complying with board gender quotas diminishes individuals' motivation to support further gender equality—and job insecurity—the concern that quotas threatens existing roles or job opportunities.

We find that participants exposed to the quota treatment are significantly more likely to perceive the policy as restricting their decision-making autonomy and conflicting with established workplace norms than non-exposed participants, suggesting that the treatment triggers psychological reactance and perceptions of gender norm violation. In contrast, we find only small and insignificant differences between treatment and control groups on measures of moral licensing and job insecurity. Echoing our heterogeneity analysis using real-world data, we also find that psychological reactance and perceived norm violation are significantly stronger among Republican-leaning and male participants. These results suggest that psychological reactance and norm violation may interact and amplify resistance to board gender quotas.

Policy implications. The experimental evidence reinforces our interpretation that resistance to board gender quotas stems from *proactive* pushback—driven by threats to individual autonomy and/or perceived conflicts with prevailing gender norms. The backlash we document suggests that a one-size-fits-all quota may be counterproductive and that more flexible, context-specific policymaking is warranted. First, the way gender policies are framed matters: aligning them with shared organizational or societal values, rather than presenting them as top-down mandates. For example, creating opportunities for open dialogue and involving the public in the policymaking process can help reduce feelings of external imposition and foster a greater sense of ownership. Additionally, emphasizing the broader benefits of diversity—such as improved team performance and innovation—rather than focusing solely on compliance, may help shift perceptions and build support. Second, policies should be tailored to the political, cultural, and occupational contexts in which they are implemented, as we show that resistance varies significantly across settings. Third, implementation mechanisms matter. For example, setting voluntary targets may be more effective at promoting gender diversity in environments where strict quotas provoke backlash. Consistent with this, we find that backlash is largely absent in European countries with voluntary targets but is evident in those with state-imposed quotas.

Literature. Our paper makes a number of contributions to the existing literature. First, it adds to the growing body of work in economics and finance that examines board gender quotas. Prior research offers mixed evidence on stock market responses to board gender quotas and their effects on operating performance and corporate innovation (Ahern and Dittmar, 2012; Matsa and Miller, 2013; Greene et al., 2020; Gertsberg et al., 2021; Hwang et al., 2021; von Meyerinck et al., 2021; Eckbo et al., 2022).[4] However, relatively little attention has been paid to how board quotas influence labor market outcomes for women, especially those outside of leadership roles. Our study is among the first to use rich, multi-faceted data—from job ads to actual hiring and turnover outcomes—to provide novel evidence on whether, and how, mandates to increase female board representation may backfire. We do not evaluate the broader efficiency or firm value implications as the backlash we document appears to be driven by individual preferences and norm-based resistance than by economic

[4]Earlier pioneering studies focus on the relationship between endogenously formed gender-diverse boards and firm outcomes, see Adams and Ferreira (2009) and Levi et al. (2014).

rationale.

Our paper builds on Bertrand et al. (2019) who examine the effect of board gender quotas on female labor market outcomes in Norway, one of the most gender-progressive countries, and find no evidence that such quotas improve employment outcomes for women beyond those who enter boardrooms.[5] We extend this line of research in two key ways. First, by applying computational linguistic methods to job ads in the US, we isolate labor demand from supply and show that board gender quotas can provoke resistance from employers toward hiring female workers. We also show that this demand-side pushback extends beyond hiring into poorer treatment of female employees and adverse employment outcomes.[6] Second, through heterogeneity analyses and a randomized experiment, we identify psychological reactance and perceived violation of gender norms as key mechanisms underlying the backlash. By focusing on labor demand-side responses and providing microfoundations for resistance, our findings help explain the many negative point estimates reported in Bertrand et al. (2019), which they characterize as "counter-intuitive."

More broadly, our paper contributes to the economics literature on backlash against regulatory and legislative efforts aimed at advancing social change. A number of papers document unintended consequences of such policies, including evidence of backlash following major US laws such as the 1975 revision to the Voting Rights Act (Ang, 2019) and the state Equal Rights Amendments in the 1970s (Wheaton, 2022).[7] Related work also cautions how regulations targeting firms with social objectives, such as promoting gender equality can produce adverse outcomes if firms' responses offset or reverse the intended effects (Acemoglu and Angrist, 2001; Cullen and Pakzad-Hurson, 2023; Bailey et al., 2024). Our contribution lies in examining board gender quotas as a prominent but highly contentious policy tool aimed at empowering women in the corporate sector. Consistent with the economics literature cited above, we show that such quotas trigger backlash and lead to unintended labor market consequences for women, ultimately undermining the very goals they are intended to achieve.

Finally, our paper adds to the rapidly expanding literature that leverages granular job posting data and computational linguistic methods to analyze labor market dynamics. Pi-

[5]Similarly, Maida and Weber (2022) evaluate Italy's 2011 board gender quota and find no effect on the representation of women in top corporate positions or among top earners.

[6]It is worth noting that our findings are not inconsistent with Matsa and Miller (2011) and Tate and Yang (2015), who show that women in leadership can generate positive externalities for female employees. We suggest that such benefits may be preceded or offset by backlash resulting from the imposition of gender quotas, at least in the short run.

[7]The psychology literature also posits that reactance theory and its manifestation—the "forbidden fruit effect" can shape public responses to various policies, including backlash against environmental regulations and resistance to government health directives (e.g., anti-smoking efforts or mask mandates during the COVID-19 pandemic). See Mazis et al. (1973); Pechmann and Shih (1999) among others.

oneering studies such as Deming and Kahn (2018) and Hershbein and Kahn (2018) use job posting data to study how skill requirements across firms and labor markets evolve over time. Closer to our focus, Kuhn and Shen (2023) and Card et al. (2024) study how explicitly stated gender preferences in job ads affect applicant behavior. Building on techniques from the computer science literature (Tang et al., 2017; Cryan et al., 2020; Bohm et al., 2020), we extend this line of research by applying computational linguistic tools to tens of millions of job ads to capture firms' subtle, implicit gender preferences in labor demand. Our measure varies across different job positions within a firm, over time, and across job locations, thereby capturing firms' demand for female labor in a granular and timely way. The same method can be applied to job ads of private firms to study their labor demand and/or extended to other countries and in different languages.

2 Data and Measurement

2.1 Data

Our job posting data come from LinkUp, an employment intelligence company. LinkUp has scraped job postings directly from public companies' websites since 2007; its coverage has improved significantly since 2014. LinkUp provides reliable data for large companies that typically post jobs on their own websites. The data include job title, full-text description of each job, information about the employer, the date of the posting, the location in which a job is posted, and basic occupation information, such as the six-digit ONET Standard Occupation Classification (SOC) code.[8]

For each job ad, we match LinkUp data to Compustat to obtain firm financial information; to BoardEx to obtain board characteristics, including the share of female directors in the job posting month; and to ExecuComp to obtain information on (up to) five top executives. Our sample comprises 30 million job ads from over 3,000 unique public companies over the period 2014–2020. We discuss the coverage of LinkUp in Online Appendix E, and show that using LinkUp data allows us to capture the largest firms in the economy without any industry or geographic bias.[9] Table A1 in the Online Appendix provides the list of variables, their definitions, and data sources.

In addition to job ads, we construct firm-month female share in new hires and employee turnover based on LinkedIn profiles. We also construct measures of female employees' work-

[8]There are two advantages to obtaining job postings directly from employer websites as opposed to third-party job boards. First, employers update their own websites more regularly than they update job boards. Second, employers post a job opening only once on their company's website, as opposed to posting the same job multiple times across various job boards. As a result, LinkUp data suffer less from stale postings or duplicate postings compared to job posting data obtained from job boards.

[9]To facilitate regression analysis, most of our regression analyses employ a random one-fifteenth (1/15) of the full sample that comprises close to 2 million job ads. We confirm that the results remain robust across 1,000 random samples in Figure 3b.

place experience using employee reviews and ratings from Glassdoor, as well as gender-related labor violations from Violation Tracker. We link these datasets to publicly traded firms in Compustat.

2.2 Measuring gendered wording in job ads

Words matter. Prior research has shown that gendered language can influence job application behavior and reinforce existing gender imbalances in the labor market. In a series of lab experiments, Gaucher et al. (2011) demonstrate that even subtle wording differences in job ads can significantly change individuals' appraisals of relevant jobs.[10] In a large-scale field experiment conducted in Spanish, a language with gendered grammar, Del Carpio and Fujiwara (2023) find that replacing (default) male-oriented nouns with gender-inclusive alternatives in job ads, i.e., effectively making job ads more female-friendly, leads to a higher share of female applicants. Relatedly, Chaturvedi et al. (2025) show that young, skilled women in the urban Indian labor market are more likely to respond to job ads in which employers exhibit a female preference for low-wage jobs. Extending the importance of subtle employer signals beyond gender, Burn et al. (2025) find that age-related language in job ads, often reflecting ageist stereotypes, strongly deters older workers from applying.

Bag-of-words (BoW) approach. To capture gendered language usage in job ads, we employ the gendered word dictionary created by Gaucher et al. (2011), based on published lists of communal (e.g., committed, supportive) and agentic words (e.g., individualistic, competitive) as well as feminine (e.g., compassionate, understanding) and masculine trait words (e.g., ambitious, assertive).[11] Their dictionary contains words and lemmas. Each lemma is followed by a wildcard (asterisk) that captures all letters, hyphens, and numbers, allowing for the inclusion of inflectional and derivational forms of a word. For example, aggress* will match aggressive and aggressiveness. For our corpus of job ads, we take standard preprocessing steps, including lower-case transforming all words, removing trailing whitespace, line break, non-English words/posts, URLs, and special characters, etc.[12] We then use regular expressions to replace the aforementioned wildcard at the end of a lemma.

For each job ad, we count the number of feminine (masculine) words, and compute the share of feminine words out of the number of feminine and masculine words, as our proxy

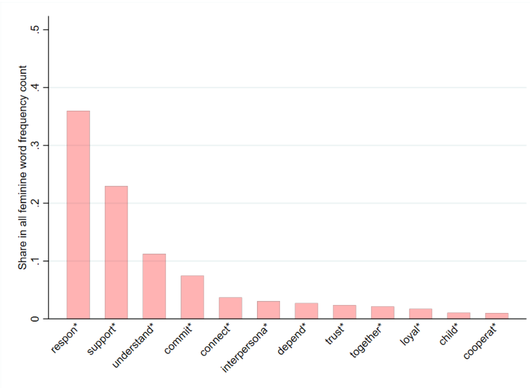
[10]For example, in their Study 4, job ads with masculine wording are ranked 10%-20% less favorably by female participants than those with feminine wording, while male participants show the opposite pattern. This suggests that gendered language can reduce the proportion of female applicants by 10%-20%. In their other experiments, when job ads include more masculine than feminine wording, participants perceive more men in those occupations (Study 3), and importantly, women find those jobs less appealing (Studies 4 and 5).

[11]These word lists are typically extracted from questionnaires given to college students to measure their self-concept and valuation of feminine and masculine characteristics (Bem, 1974; Rosenkrantz et al., 1968).

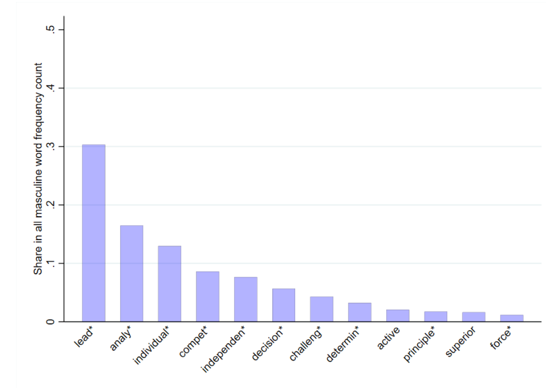
[12]Table G1 in the Online Appendix provides the complete dictionary of feminine and masculine words in Gaucher et al. (2011).

for female labor demand. Figure 1a (Figure 1b) shows shares of the most frequent feminine (masculine) words in job ads from the Gaucher et al. (2011) list over the sample period (2014–2020). Figures 1c and 1d further show relatively stable shares of the most frequent gendered words in job ads over time.

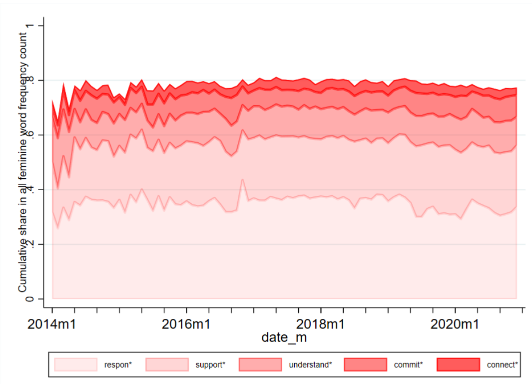
Figure 1: Top Feminine and Masculine Words



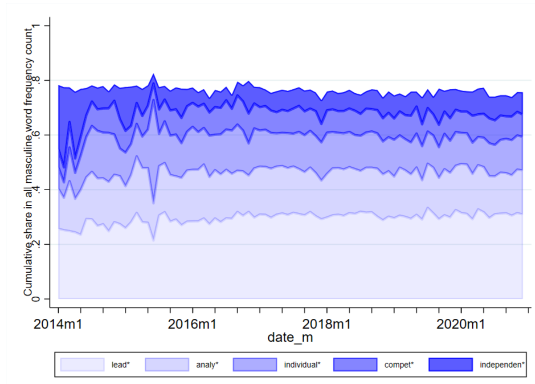
(a) Feminine Words (Share > 1%)



(b) Masculine Words (Share > 1%)



(c) Top 5 Feminine Words Over Time



(d) Top 5 Masculine Words Over Time

This figure shows the most frequent feminine and masculine words in job ads from the Gaucher et al. (2011) list. Figure 1a (Figure 1b) shows the share of the most frequent feminine (masculine) words out of all feminine (masculine) words over the sample period (2014-2020). Figure 1c (Figure 1d) shows the cumulative share of the top 5 feminine (masculine) words out of all feminine (masculine) words over time. The asterisk denotes the acceptance of all letters, hyphens, or numbers following its appearance.

BERT. While using the gendered word dictionary from Gaucher et al. (2011) is transparent and easily replicable, this approach is subject to some limitations. First, the dictionary could be incomplete and outdated. We address this limitation by augmenting the list with additional gendered words provided by a leading job ad optimization company.[13] Second, the BoW approach ignores the order or context of words and does not capture relationships between words, which can be crucial for understanding nuanced meaning and complex text

[13]See `https://www.ongig.com/gender-bias-in-job-descriptions`.

structure. To tackle this, we develop alternative measures of gendered wording in job ads based on BERT (Devlin et al., 2018), a pre-trained large language model that has demonstrated impressive performance on a wide range of natural language processing (NLP) tasks. To fine-tune a BERT model, we first form a representative sample of 4,500 job ads equally distributed in male-dominated, female-dominated, and gender-neutral occupations based on the share of female employees in a given occupation. We then have eight annotators to label them into feminine versus masculine job ads. Finally, we use 80% of the labeled data as training data and the remaining 20% as testing data to fine-tune the model. Our model is fine-tuned on JobBERT, pre-trained on 3.2 million sentences from job postings (Zhang et al., 2022).[14] The fine-tuned model assigns a probability between 0 and 1 indicating the likelihood of each job ad being more feminine. Our model has an area under the receiver operating characteristic curve (AUC) score of 94% in predicting the "feminine" probability. As expected, the BERT-based measure is positively correlated with the BoW-based measure (see Figure B1). For details of our BERT-based method and statistics, see Online Appendix B.

2.3 Employment outcomes: new hires and departures

We use job ads as our primary measure of gendered labor demand because they offer several distinct advantages. First, job ads are ubiquitous, making them especially valuable for studying potential backlash. This widespread availability allows us to examine hiring preferences across a broad range of occupations, firms, and industries. Second, job ads reflect pure demand-side preferences, untainted by supply-side factors that can confound other metrics such as gender gaps in employment.

A potential concern is whether gendered language in job ads truly deters applications or affects actual hiring outcomes. To investigate this, we complement our analysis with gender-specific firm-level employment outcomes, which offer a more direct and policy-relevant measure of women's labor market outcome. However, acquiring firm-level data on workforce composition poses a significant challenge, as companies are not required to publicly disclose such information.[15] To overcome this challenge, we use LinkedIn, which provides publicly accessible user profiles containing related information, including names and employment histories. We infer an individual's gender from their names and determine their employer at each point in time based on their career history. By aggregating these individual-level records, we construct monthly snapshots of gender-specific employee inflows and outflows at the firm level.

[14] For robustness checks, we also use DistilBERT, a distilled version of BERT that retains almost all the language understanding capabilities of BERT while being faster and smaller (Sanh et al., 2019).

[15] In the US, one potential source is the proprietary employer-employee matched data maintained by the Census Bureau, such as the Longitudinal Employer-Household Dynamics (LEHD) program.

We first apply several data cleaning steps to raw individual-level records. We start by restricting the sample to US-based employees and keeping only those profiles that report a valid job start date. Each employment spell is required to last at least three months to filter out short-term or temporary roles. To reduce noise from highly mobile individuals, we exclude those with more than 10 job entries or more than 5 distinct employers. We also remove profiles with overlapping employment spells or those indicating simultaneous employment at multiple firms, ensuring a clean, sequential employment history for individuals in our sample. After applying these filters, we construct the number of employees, and more importantly, the number of new hires, and the number of departed employees, for each firm-month.

We next sort the number of employees, new hires, and departed employees by gender using a probabilistic method based on an individual's first name. This approach relies on data from the US Social Security Administration, which reports the frequency of each name by gender over time. For each name, we calculate the probability that it is associated with a male or female, based on historical usage. For example, if 80% of the individuals named "Tracy" are females and 20% are males, the model assigns "Tracy" with a 0.8 probability of being a female and a 0.2 probability of being a male. To improve accuracy in our measures of gender-specific employee inflows and outflows, we include only individuals whose predicted female probabilities are below 0.1 (classified as a male) or above 0.9 (classified as a female). Given that job ads are naturally linked to hiring outcomes, our primary outcome of interest is the share of female new hires among all new hires for each firm-month. Another outcome variable we examine is the share of female departed employees among all departed employees.

2.4 Linking gendered job ads to employment outcomes

The existing literature already highlights the implications of gendered language usage for female labor market outcomes in experimental and real-world contexts. In our US setting, does the presence of more feminine language in job ads signal a greater demand for female employees? We present supporting evidence at both the occupation and the firm level.

Figure G1a in the Online Appendix presents a sample overview of gendered language usage in different occupations based on a job's two-digit ONET-SOC code. We show that across the six major occupational categories, the sales and office sector and the service sector, traditionally associated with high shares of female employment, have the highest shares of feminine words in job ads, whereas the military sector, traditionally associated with high shares of male employment, has the lowest share, suggesting that gendered language usage could be a meaningful proxy for gendered labor demand. Figure G1b examines the relationship between gendered language usage in job ads and the share of female employees at the six-digit ONET-SOC code level. We find a strong positive correlation: occupations with a higher share of female employees tend to use more feminine wording in job ads.

We next examine whether and how gendered language usage varies with occupation-level characteristics. Figure G2 in the Online Appendix splits the sample using four different occupation-level characteristics: job zone, mean wage, entry education requirement, and STEM occupations or not. We show that jobs requiring more preparation (Figure G2a), offering above-median pay (Figure G2b), requiring above-median level of education (Figure G2c), or being in STEM sectors (Figure G2d) tend to employ more masculine words in job ads, and likely more male employees, than jobs requiring less preparation, offering below-median pay, requiring below-median level of education, or not being in STEM sectors.

One potential concern about us relying on occupation-level variations is that gendered language usage in job ads may reflect the inherent nature of a job, such as job tasks or required skills, rather than employers' gender-specific hiring preferences. To address this, we exploit firm-level variations in gendered language usage and their share of female employees in Figure G3. Aggregating individual profiles from LinkedIn into a firm-month panel, we compute the share of female employees at each firm and associate it with the share of feminine words in the same firm's job ads using binscatter plots. Figure G3a shows a strong positive relationship in the raw data. Figure G3b through Figure G3d gradually add firm-level controls and fixed effects for time, industry, and firm headquarters state. Notably, we control for firm-level factors that are likely relevant to female employment, including the presence of female top executives and board gender diversity. The positive relationship remains, suggesting that gendered language usage in job ads has the potential to capture an important aspect of firms' gender-specific employment practices beyond what is explained by occupation-level variations or observable firm characteristics.

The analysis above is correlational in nature and does not identify the causal effect of gendered language usage, which is beyond the scope of this paper. Nonetheless, we uncover a consistent and positive association between gendered language usage in job ads and gender composition of the workforce at both the occupation and the firm level. This finding suggests that gendered language likely reflects employers' gender-related expectations or preferences for prospective employees, supporting its usage as a proxy for gendered labor demand.

2.5 Other female labor market outcomes: post-hiring experience

Using textual information from Glassdoor employee reviews, we construct measures that capture amenities important to women in the workplace. We also analyze employees' quantitative ratings of female-friendly benefits. Finally, we employ gender-related labor violation cases and penalties from Violation Tracker as supplemental measures of poor workplace treatment of female employees. A detailed description of these measures and datasets used is provided in Online Appendix F.

2.6 Summary statistics

Table 1 presents the summary statistics for observations at the job ad level. The mean/median share of feminine words in a job ad is 58%/57%. The mean/median number of feminine (masculine) words in a job ad is 4.5/4 words (3.7/2 words). These positions have an average hourly wage of $28, and approximately 7% are classified as STEM jobs. Out of the 11 directors in a median firm, 25% are females.[16] We also present the summary statistics for control variables used in the regression analysis of Equation 1. These firm characteristics are largely consistent with those of large Compustat firms, given that LinkUp's coverage leans toward large companies that typically post job ads on their own websites. Table G2 Panel A in the Online Appendix presents the summary statistics for variables constructed using LinkedIn profiles. At the firm-month level, a median firm has 842 LinkedIn employees and adds 13 new hires each month. The mean/median share of female new hires is 41% (39%). The mean/median share of female employee turnover is 40% (38%). Panels B to D cover employee reviews, ratings, and gender-related labor violations and penalties.

Table 1: Summary Statistics

Variable	N	Mean	Std. Dev.	Median	p1	p99
%Fem words	1,816,477	0.58	0.27	0.57	0.00	1.00
#Fem words	1,816,477	4.50	3.80	4.00	0.00	17.00
#Mas words	1,816,477	3.70	4.00	2.00	0.00	18.00
#Words	1,816,477	447.00	259.00	409.00	84.00	1217.00
Job zone	1,246,854	2.70	1.00	2.00	1.00	5.00
Mean wage	1,246,854	28.00	16.00	22.00	12.00	74.00
Entry education	1,246,854	1.90	2.00	1.00	0.00	7.00
STEM	1,246,854	0.07	0.25	0.00	0.00	1.00
#Dir	1,816,477	11.00	2.20	11.00	5.00	16.00
%Fem dir	1,816,477	0.25	0.11	0.25	0.00	0.50
Firm size (log assets)	1,816,477	9.60	1.90	9.40	5.70	15.00
Cash holdings	1,816,477	0.11	0.11	0.07	0.00	0.53
Tangibility	1,816,477	0.30	0.22	0.30	0.00	0.84
Tobin's Q	1,816,477	2.70	2.30	1.80	0.74	9.70
Leverage	1,816,477	0.52	0.56	0.36	0.00	2.40
ROA	1,816,477	0.16	0.13	0.13	-0.07	0.61
Capex	1,816,477	0.04	0.03	0.03	0.00	0.16
%Republican vote	1,583,031	0.42	0.16	0.41	0.12	0.80
Masculinity (0-1)	1,579,623	0.46	0.03	0.46	0.35	0.51
Gender wage gap	785,558	0.18	0.07	0.19	0.06	0.34

The table presents summary statistics for the main variables used in our analysis. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020.

3 Empirical Strategy

We exploit the passage of SB 826 in California on September 30, 2018 for identification. The bill required companies listed on major US stock exchanges and headquartered in the state

[16] This number is on the higher end because the summary statistics are calculated at the job ad level, and larger firms tend to post more job ads, which skews the sample toward larger firms.

(over 600 firms) to have at least one female director on their boards by the end of 2019.[17] We run the following difference-in-differences (DID) regression at the job ad level:

$$Y_{p,t} = \boldsymbol{\gamma_i} + \boldsymbol{\delta_{l,t}} + \boldsymbol{\alpha_{k,t}} + \boldsymbol{\vartheta_q} + \beta \times Flag\_CA_i \times Post\_2018m9_t + \boldsymbol{\theta' X_{i(p),t}} + \varepsilon_{p,t} \qquad (1)$$

where p denotes job ad, i denotes firm, l denotes two-digit SIC industry, k denotes region (state/zip), q denotes occupation, and t denotes year-month when a job is posted. The dependent variable, $Y_{p,t}$, is the share of feminine words among all gendered words in a job ad, an ad-level measure of gendered labor demand. Ad- and firm-level control variables, denoted by $X_{i(p),t}$, include log (# words in an ad), firm female leadership (i.e., women in C-suite), firm size, cash holdings, tangibility, Tobin's Q, leverage, ROA, and capex.

The indicator variable, $Flag\_CA_i$, takes the value of one for public firms headquartered in California (CA), and zero otherwise. The other indicator variable, $Post\_2018m9_t$, takes the value of one after September 30, 2018, and zero otherwise. The coefficient of interest, β, captures the (reduced-form) effect of SB 826. The DID regressions for employee reviews and ratings follow a similar structure, using data at the review level. Under the backlash interpretation, firms in California push back gender-balancing quotas in their boardrooms by reducing their demand for female labor elsewhere, we expect a negative β. In Table G3 in the Online Appendix, we cross-validate whether the passage of SB 826 resulted in an increase in the number/share of female directors for CA firms compared to non-CA firms. We find that indeed, the passage of SB 826 resulted in a 0.15 increase in the number of female directors and a 2 percentage-point increase in their share in CA firms compared to that in non-CA firms. These effects represent approximately 10% of the sample mean.

In Equation 1, the unit of observation is a job ad. One benefit of such a specification is that it allows us to include high-dimensional fixed effects to control for occupation-level characteristics and local labor supply shocks. Specifically, we include firm and industry $\times$ time fixed effects (γ_i and $\delta_{l,t}$) to absorb firm-invariant characteristics and time-varying industry-specific shocks, respectively. Since labor markets are segmented by region, we add region $\times$ time fixed effects ($\alpha_{k,t}$) to account for time-varying regional labor supply shocks. We define region at either the state or zipcode level.[18] Importantly, we also include occupation fixed effects (at the six-digit ONET-SOC code level) to allow for a comparison of job ads within a specific occupation between CA and non-CA firms.[19] As an alternative

[17]It further required that by the end of 2021, all firms must have at least one female director if their boards have four directors or fewer, two female directors if their boards have five directors, and three female directors if their boards have six directors or more.

[18]Note that information for a zipcode is not always available, resulting in a smaller sample size when job zip $\times$ year-month fixed effects are included in the regression.

[19]There are around 1,000 unique occupations at the six-digit ONET-SOC code level. See https://www.onetonline.org/find/all for a list of occupations.

to occupation fixed effects, we include more granular job title fixed effects to ensure that effectively we compare the same job to extract any possible change in the language used (instead of any change in the type of jobs advertised).

For analysis based on LinkedIn profiles, we run a similar DID regression with observations at the firm-month level:

$$Y_{i,t} = \boldsymbol{\gamma_i} + \boldsymbol{\delta_{l,t}} + \beta \times Flag\_CA_i \times Post\_2018m9_t + \boldsymbol{\theta' X_{i,t}} + \varepsilon_{i,t} \tag{2}$$

where i denotes firm, l denotes two-digit SIC industry, and t denotes year-month. $Y_{i,t}$ is either the share of female new hires among all new hires or the share of female departed employees among all departed employees, for each firm-month. The DID regressions for gender-related labor violations and penalties follow a similar structure, using data at the firm-year level.

4 Baseline Effects of SB 826 on Hiring Practice and Outcome

In this section, we first examine the impact of SB 826 on the use of gendered language in job ads, which reflects firms' gendered labor demand. We then examine whether these changes in language are followed by corresponding changes in firm-level hiring outcomes.

4.1 Gendered language in job ads

We employ a DID design in Equation 1 in which we compare firms headquartered *inside* and *outside* California, *before* and *after* the September 2018 law change. Table 2 presents the results. In column (1) with length of job ads and firm characteristics as controls, we find a negative and significant DID coefficient of -0.053, suggesting that SB 826 reduced the share of feminine words by 5.3 percentage points. In column (2), we compare within each occupation (at the six-digit ONET-SOC code level) for jobs posted by treated and control firms around SB 826. The estimated coefficient is similar in terms of magnitude and significance to that in column (1). To better control for local labor supply shocks, we add job zip × year-month fixed effects in column (3) and again find a similar effect. One may worry that the changes we observe may be driven by pre-existing gender-related culture rather than the effect of SB 826. To address this, we control for female executive presence in columns (4)-(6) and the estimated coefficients change little.[20]

Overall, we find that SB 826 reduced the share of feminine words by 4.2-5.4 percentage points. In terms of economic significance, the drop in the usage of feminine words by four to five percentage points represents about 10% of the mean share of feminine words in job

[20] We observe a negative coefficient on Female CEO/CFO, which may reflect a reduced need for firms to signal gender-specific inclusion through job ads when female leadership is already visible. We consider this as an alternative interpretation of the DID coefficient and rule it out by examining hiring outcomes in Section 4.2. We further note that cash holdings and ROA are negatively and significantly, whereas Tobin's Q and Capex are positively and significantly, correlated with the share of feminine words in job ads.

ads. We view this estimate as a crude proxy for the share of job ads tilting from femininely-worded to masculinely-worded. According to Gaucher et al. (2011), masculinely-worded job ads are ranked 10%-20% less favorably than femininely-worded job ads by female applicants. Combining both statistics suggests that the female applicant pool shrinks by 1%-2%. All else equal, this also implies that the share of female new hires would decline by approximately 0.5-1 percentage points, assuming a gender balanced applicant pool.

Table 2: Effect of California Statute on Gendered Language Usage in Job Ads

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)					
Flag_CA × Post_2018m9	-0.053***	-0.046***	-0.041***	-0.054***	-0.047***	-0.042***
	[0.008]	[0.008]	[0.008]	[0.010]	[0.011]	[0.010]
Log (#Words)	-0.036***	-0.024***	-0.023**	-0.037***	-0.026***	-0.025**
	[0.012]	[0.008]	[0.010]	[0.014]	[0.009]	[0.011]
Firm size	0.020*	0.017	0.014	0.016	0.012	0.009
	[0.011]	[0.010]	[0.009]	[0.013]	[0.013]	[0.011]
Cash holdings	-0.110***	-0.095***	-0.078**	-0.147***	-0.128***	-0.112***
	[0.028]	[0.028]	[0.032]	[0.030]	[0.031]	[0.030]
Tangibility	-0.112	-0.137	-0.125	-0.188**	-0.208**	-0.191**
	[0.082]	[0.087]	[0.085]	[0.082]	[0.090]	[0.086]
Tobin's Q	0.007***	0.008***	0.006**	0.009***	0.010***	0.007***
	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.003]
Leverage	0.003	0.011	0.007	-0.015	-0.003	-0.009
	[0.021]	[0.021]	[0.019]	[0.022]	[0.023]	[0.020]
ROA	-0.143**	-0.142**	-0.109**	-0.216***	-0.214***	-0.175**
	[0.063]	[0.057]	[0.051]	[0.077]	[0.070]	[0.065]
Capex	0.209	0.219	0.295**	0.395*	0.419*	0.456**
	[0.142]	[0.146]	[0.131]	[0.207]	[0.209]	[0.194]
Female CEO				-0.054**	-0.041	-0.036*
				[0.022]	[0.025]	[0.021]
Female CFO				-0.024	-0.025	-0.025*
				[0.015]	[0.015]	[0.013]
Female COO				0.007	0.013	0.003
				[0.021]	[0.022]	[0.021]
Firm FE	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES
Job State × Year-Month FE	YES	YES	-	YES	YES	-
Job Zip × Year-Month FE	NO	NO	YES	NO	NO	YES
Occupation (ONET 6 digit) FE	NO	YES	YES	NO	YES	YES
Obs	1,816,477	1,816,477	1,816,477	1,642,550	1,642,550	1,642,550
Adj R2	0.357	0.415	0.43	0.36	0.419	0.434

This table shows the effect of the California board gender diversity statute (SB 826) on gendered language usage in job ads. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

BERT-based approach. We use BERT-based gendered wording measures to show robustness of our main findings. In Panel A of Table G4, we present the DID results using the BoW and JobBERT measures. In columns (1) and (2), we start with the continuous measure and find a significantly negative effect of SB 826 using JobBERT. This effect becomes even stronger when we replace the continuous measure with quartile or decile indicators, as shown in columns (4) and (6). Panel B repeats the same analysis using the BoW and DistilBERT measures. Our main findings remain.

Treatment intensity. To address the concern that our DID results could be influenced by concurrent events such as the #MeToo Movement in 2017, we examine variation in treatment intensity. Specifically, we consider whether a CA firm was already in compliance with SB 826 prior to the bill's passage, or whether it experienced any increase in the number of female directors following the bill's passage.[21]

Table 3: Effect of California Statute on Gendered Language Usage in Job Ads – Treatment Intensity

	(1)	(2)	(3)	(4)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)			
Subsamples	Met the threshold in SB 826 before 2018m9?		Change in #female directors after 2018m9?	
	Just below the threshold	At the threshold	Increase	no Increase
Flag_CA × Post_2018m9	-0.041***	0.000	-0.049***	-0.006
	[0.008]	[0.015]	[0.010]	[0.015]
p-value (diff)	0.002		0.008	
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES
Occupation (ONET 6 digit) FE	YES	YES	YES	YES
Obs	377,637	630,272	512,965	1,082,906
Adj R2	0.545	0.357	0.463	0.448

This table shows heterogeneities in the effect of SB 826 on gendered language usage in job ads across firms that are differentially treated. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Columns (1) and (2) compare firms just below the threshold set by SB 826 (one female director short) and firms at the threshold before the treatment date. Columns (3) and (4) compare firms that experienced an increase in their number of female directors after the treatment date versus firms without an increase. p-value (diff) reports the p-value for testing the difference between the coefficients from different subsamples. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

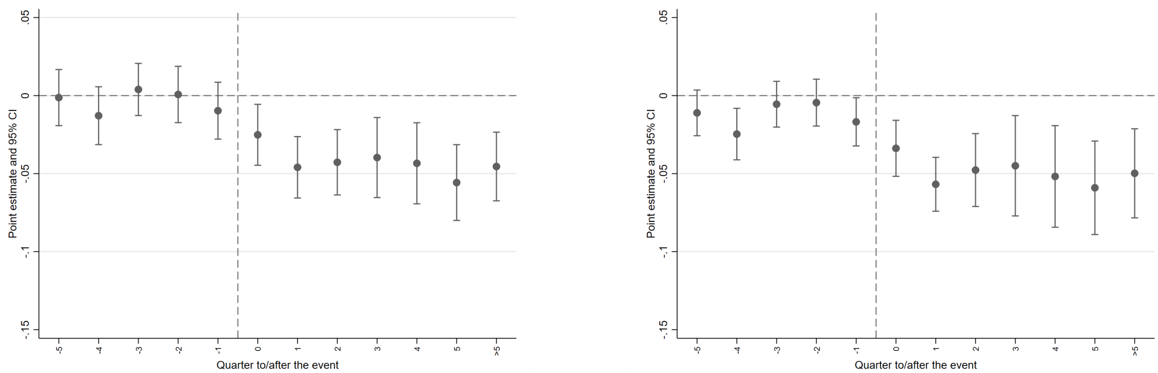
Our analysis begins by comparing firms that were just below the threshold set by SB 826 (i.e., one female director short) with those that just met the requirement prior to the treatment. This approach offers two key advantages over comparing firms based on varying degrees of noncompliance. First, both the number of female directors and the decision to comply with SB 826 were endogenous. It is not obvious that a greater shortfall would lead to more pressure to comply. One could argue that CA firms further away from meeting SB 826's requirement had less incentive to adjust and could potentially abandon efforts to meet the quota. Conversely, firms closer to the threshold may have more incentive to comply, resulting in a more intense treatment effect. Second, firms just below and just meeting SB 826's

[21] Noncompliance with SB 826 could result in financial penalties: a first violation could incur a $100,000 fine, and each subsequent violation could lead to a $300,000 fine. Additionally, failing to timely file board membership information with the Secretary of State could result in a $100,000 fine.

requirement before its passage are more comparable, allowing us to sharpen identification by focusing on firms that are similar in all respects except for their difference in pre-treatment compliance. In columns (1) and (2) of Table 3, we show a negative and significant effect for CA firms whose number of female directors fell short of the threshold. In contrast, those that just met the quota experienced little effect on the share of feminine words in their job ads.

Columns (3) and (4) further compare CA firms that experienced an increase in their number of female directors after the bill's passage (suggesting compliance) with those without. The treatment effect only showed up in the former. These findings further strengthen our identification strategy using SB 826.

Figure 2: California Statute and Gendered Language Usage in Job Ads – Dynamics



(a) Column (3) of Table 2 **(b)** Column (6) of Table 2

This figure shows the dynamic effect of SB 826 on gendered language usage in job ads. Each figure corresponds to the specifications with the most stringent fixed effects in Table 2. Every three months in a quarter are grouped together to reduce estimation error. Quarter 0 is the third quarter of 2018, when the bill was passed. The omitted category is < -5 or all quarters prior to quarter -5.

Dynamic effect. A key identifying assumption is the presence of parallel trends, i.e., in the absence of SB 826, the usage of gendered language in job ads would have followed similar trends for CA and non-CA firms. Figure 2 presents the temporal pattern based on regression results. To reduce estimation error, we group every three months within a corresponding quarter. We set quarter 0 to be the third quarter of 2018, when the bill was passed. We combine quarters prior to $t = -5$ into a single period and set it as the omitted category. All quarters after $t = 5$ are also combined. The dynamic DID coefficients confirm that the effect of SB 826 most materialized after the bill's passage. However, the coefficient for $t = -1$, which correspond to the second quarter of 2018, is negative and sometimes significant, suggesting anticipatory effects in the quarter immediately preceding the bill's passage. This likely reflects strong expectations that the bill would pass. Based on the bill's history from the state government website,[22] the first sign that it would become law appeared in April 2018, when it was read at the Assembly for the first time and passed with strong support.[23] The coefficients on all quarters prior to the second quarter of 2018 are mostly close to zero and insignificant. Importantly, there is a clear negative post-quota trend, suggesting significant pushback after the bill's passage. This effect does not dissipate over time, as the negative coefficients stabilize after two quarters, indicating a persistent rather than a transitory response from CA firms.

In 2019, three California taxpayers filed a lawsuit (*Crest v. Padilla I*), seeking to nullify SB 826. The plaintiffs' main argument was that the bill violated the Equal Protection Provisions of the California Constitution. The fact that the lawsuit was initiated shortly after the bill's passage is consistent with backlash. The Los Angeles Superior Court struck down SB 826 in May 2022. The California Secretary of State appealed and a final ruling is still pending. Table G3 columns (3) and (4) in the Online Appendix show that there was no reversal in the number/share of female directors in CA firms compared to their non-CA counterparts after the repeal, suggesting that it may be premature to explore the repeal's implications for female labor market outcomes in CA firms.

4.2 Hiring outcome by gender

Job ads represent a first step in the hiring process and are closely linked to eventual hiring outcomes. A natural question to ask is whether SB 826 led to a decline in the share of women among new hires at the affected firms. This analysis also allows us to assess the extent to which gendered language in job ads matters in shaping the gender composition of incoming employees.

Column (1) of Table 4 presents the results using the full LinkedIn sample. Consistent with our findings from the job ad analysis, we observe a negative and significant DID coefficient, indicating that SB 826 reduced the share of female new hires. In light of COVID-19's disruption to US labor market, we restrict the sample to the pre-COVID period in column (2) and the coefficient changes little. One limitation of LinkedIn data is its coverage bias. Not all firms are equally represented on the platform, particularly smaller or less tech-oriented firms. To address this concern, we apply a series of data filters to exclude firms with limited coverage from LinkedIn. In column (3), we exclude the 20 industries that are likely to be underrepresented from LinkedIn. In column (4), we drop firms with fewer than 200 identified employees from LinkedIn, which represent the bottom quartile. The DID

[22] https://leginfo.legislature.ca.gov/faces/billVotesClient.xhtml?bill_id=201720180SB826.

[23] Allen and Wahid (2024) further note a key information event in the second quarter of 2018, when SB 826 passed the Senate with a 22-11 vote.

coefficients in columns (3) and (4) become slightly stronger and remain significant at the 1% and 5% levels, respectively. To verify the robustness of our results, we apply alternative cutoffs based on industry or number of employees coverage. Columns (1) and (2) of Table G5 show that the results remain highly consistent.

Table 4: Effect of California Statute on Hiring

	(1)	(2)	(3)	(4)
Dep. Var.	Share of female new hires out of all new hires (%Fem new hire)			
Sample	Full	Excl. COVID	Excl. low coverage (industry)	Excl. low coverage (firm)
Flag_CA × Post_2018m9	-0.007**	-0.008**	-0.009***	-0.008**
	[0.003]	[0.003]	[0.003]	[0.003]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	81,308	71,830	64,203	69,785
Adj R2	0.291	0.294	0.285	0.426

This table shows the effect of SB 826 on the gender composition of newly hired employees. The outcome variable is the share of female new hires out of all new hires. The unit of observation is a firm-month. The sample includes public firms covered by LinkedIn over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Columns (1) employs the full sample. Columns (2) includes only the pre-COVID period (before March 2020). Columns (3) excludes industries known for low LinkedIn coverage. The following 20 industries are excluded: agricultural production crops, forestry, fishing, hunting, and trapping, metal mining, nonmetallic mineral mining and quarrying, heavy construction, construction special trade contractors, textile mill products, leather and leather products, stone, clay, glass, and concrete products, motor freight transportation and warehousing, water transportation, transportation by air, transportation services, wholesale trade-durable goods, wholesale trade-nondurable goods, food stores, automotive dealers and gasoline service stations, hotels, rooming houses, camps, and other lodging places, business services (certain sub-sectors), and motion pictures. Columns (4) excludes firms with fewer than 200 employees from LinkedIn (bottom quartile) to mitigate coverage bias in the LinkedIn data. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

In terms of economic significance, SB 826 reduced the share of female new hires by around 1 percentage point. Over the period 1950–2020, the share of females in the labor force rose from 28.6% to 46.8%, an annual increase of 0.26 percentage points.[24] A reduction of 1 percentage point suggests that SB 826 set back the progress in female labor market participation by three to four years. We also note that the effect here is larger than the effect (a 0.5-to-1 percentage-point drop) inferred in Section 4.1, which is based on the estimated drop in the share of feminine words in job ads. In practice, pushback can happen at different stages of the hiring process, such as candidate screening and interview evaluations, contributing to the larger drop in actual female hires. Importantly, these different magnitudes reaffirm the influential role of gendered language in job ads: changes in wording alone explain at least half of the overall drop in female new hires.

We also examine heterogeneity in treatment intensity based on CA firms' compliance with SB 826. In Table 5, columns (1) and (2) compare CA firms that were just below the

[24] See https://www.dol.gov/agencies/wb/data/lfp/civilianlfbysex for information from the US Bureau of Labor Statistics.

compliance threshold to those that were at the threshold prior to the bill's passage. We find a significantly larger decline in the share of female new hires among firms just below the threshold, while firms at the threshold show little to no effect. Columns (3) and (4) further confirm that the overall treatment effect was driven entirely by CA firms that added female directors following the bill's enactment. These results reinforce our interpretation that the observed labor market responses were driven by compliance with SB 826 rather than broader contemporaneous trends or external shocks.

Table 5: Effect of California Statute on Hiring – Treatment Intensity

	(1)	(2)	(3)	(4)
Dep. Var.	Share of female new hires out of all new hires (%Fem new hire)			
Subsamples	Met the threshold in SB 826 before 2018m9?		Change in female directors after 2018m9	
	Just below the threshold	At the threshold	Increase	No change
Flag_CA × Post_2018m9	-0.020***	0.007	-0.022***	0.001
	[0.005]	[0.006]	[0.006]	[0.004]
p-value (diff)	0.000		0.000	
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	24,985	15,942	25,827	42,048
Adj R2	0.346	0.488	0.407	0.436

This table shows heterogeneities in the effect of SB 826 on the gender composition of newly hired employees across firms that are differentially treated. The outcome variable is the share of female new hires out of all new hires. The unit of observation is a firm-month. The sample includes public firms covered by LinkedIn over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Columns (1) and (2) compare firms just below the threshold set by SB 826 (one female director short) and firms at the threshold before the treatment date. Columns (3) and (4) compare firms that experienced an increase in their number of female directors after the treatment date versus firms without an increase. p-value (diff) reports the p-value for testing the difference between the coefficients from different subsamples. The sample excludes firms with fewer than 200 employees from LinkedIn (bottom quartile) to mitigate coverage bias in the LinkedIn data. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

The negative effect of SB 826 on female hiring also helps rule out an alternative explanation for our job ad-level findings, namely, that the introduction of board gender quotas shifted beliefs and triggered a supply-side response. In other words, the policy might have signaled a cultural shift within affected firms, making them more attractive to female job seekers. Consequently, the increased supply of female applicants could have reduced the need for female-friendly language in job ads, as firms would no longer need to actively signal inclusiveness to attract female talent. However, the observed decline in the share of female new hires contradicts this interpretation.

4.3 Other possible explanations and robustness checks

Gendered job ads vs. gendered jobs One may argue that our job ad-based measure captures gendered jobs instead of gendered job ads. If so, our findings could be driven by SB

826 shifting the type of jobs advertised rather than changing the language used to describe a job. In the earlier specifications, we have included occupation fixed effects (at the six-digit ONET-SOC code level) to compare job ads within the same occupation. To further alleviate this concern, we limit the comparison to job ads with the same job title, which is substantially more granular than the six-digit ONET-SOC code. Column (1) of Table 6 includes full job title fixed effects, and the results change little. To alleviate the concern that the sample of job ads that fall within the common support of control variables and full job title fixed effects would be very thin, we also use the first three words in a job title to form fixed effects and find similar results in column (2).[25] In column (3), we further restrict the comparison to job ads with the same first three words in a job title, the same six-digit ONET-SOC code and, importantly, posted by the same firm. Comparing these nearly identical jobs yields similar results.

Table 6: Effect of California Statute on Gendered Language Usage in Job Ads – Gendered Job Ads vs. Gendered Jobs

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)					
	Same job title			Drop gendered jobs		Drop job-specific gendered words
Flag_CA × Post_2018m9	-0.049***	-0.047***	-0.075***	-0.034***	-0.042***	-0.029**
	[0.011]	[0.010]	[0.011]	[0.008]	[0.007]	[0.012]
Job Title FE (full)	YES	-	NO	NO	NO	NO
Job Title FE (first 3 words)	NO	YES	-	NO	NO	NO
Job Title × Occupation × Firm FE	NO	NO	YES	NO	NO	NO
Firm FE	YES	YES	-	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES	YES	YES
Occupation (ONET 6 digit) FE	NO	NO	-	YES	YES	YES
Obs	1,641,670	1,641,670	1,641,670	1,354,218	1,009,552	1,639,691
Adj R2	0.714	0.665	0.778	0.447	0.433	0.399

This table shows the effect of SB 826 on gendered language usage in job ads restricting the comparison to job ads with the same job title or in gender-neutral occupations. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Columns (1) to (3) include fixed effects related to job titles. Column (4) excludes jobs in occupations where the female worker share is below 25% or above 75%. Column (5) excludes jobs in occupations where the female worker share is below 40% or above 60%. Column (6) excludes the top 5 gendered words (in frequency) in a job ad when calculating the outcome variable. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

We next limit our analysis to gender-neutral jobs to ensure we capture gendered labor demand instead of gendered jobs. Specifically, we drop potentially gendered jobs based on the share of female employees in a given occupation.[26] Columns (4)-(5) present the results using different cutoffs for gendered occupations. We note that the estimated coefficients

[25] The most popular job titles include "restaurant general manager," "customer service associate," "retail sales associate," and "assistant general manager."

[26] We calculate the female share using the 2020 Labor Force Statistics from the BLS Current Population Survey. See Table 39 from https://www.bls.gov/cps/cps_aa2020.htm.

remain similar to the baseline estimate.

Finally, certain words in our gendered word list may be more problematic than others in the sense that they may reflect the nature of a job rather than the gendered language choice. For example, the word "analy*" shows up in all job ads concerning "analysts" and it is a masculine word. We thus re-compute our measure by dropping the five most frequent (likely job-specific) gendered words from each job ad. Column (6) in Table 6 presents the results. We note that the estimated coefficient remains similar to the baseline estimate, ruling out the possibility that our results are driven by high-frequency job-specific gendered words.[27]

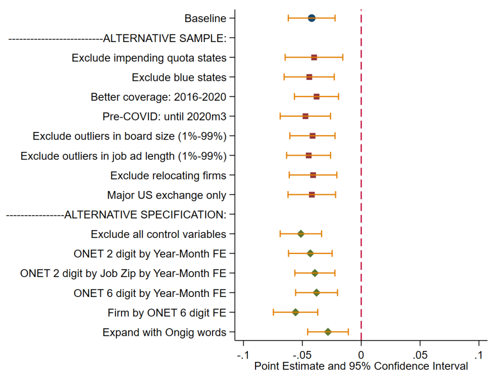
Spillover effects. One concern about DID analyses like ours is the potential for spillover effects. While firms headquartered in CA constitute the treated group by design, control firms may still be influenced by similar factors. For instance, these firms might be located in states with impending board gender quotas or in states that are likely to follow CA's lead in similar legislative efforts. Additionally, firms may anticipate the passage of such quotas and begin adjusting their behavior in advance. If this is the case, the estimated treatment effect could be biased, as the control group is partially treated. To assess the extent of this issue, we conduct two additional tests. First, we exclude firms headquartered in states that have proposed board gender quotas similar to SB 826 (i.e., MA, IL, NJ, NY, and WA). Second, we exclude blue states receiving more than 50% voter support for the Democratic candidate in the 2020 US presidential election, as these states are more likely to follow CA's legislative agenda. Figure 3a shows that the estimated coefficients using these alternative samples are similar to the baseline estimate, suggesting that any potential spillover did not significantly impact our main findings.

Alternative sampling criteria. We use subsamples starting in 2016 (to address the limited coverage of LinkUp in 2014 and 2015), including only the pre-COVID period (before March 2020, to eliminate the disruption of COVID-19 to US labor market), removing firms with a very small or large board size (below the 1st percentile or above the 99th percentile, to eliminate potential firm outliers), removing extremely long or short ads (to eliminate potential job ad outliers), removing CA firms relocated their headquarters (out of CA) over our estimation period (to ensure that we capture a clean treatment effect), or limiting to firms listed on NYSE, AMEX, or NASDAQ (to ensure that we capture actual treated firms). The estimated coefficients using these alternative samples are plotted in Figure 3a and they remain similar to the baseline estimate.

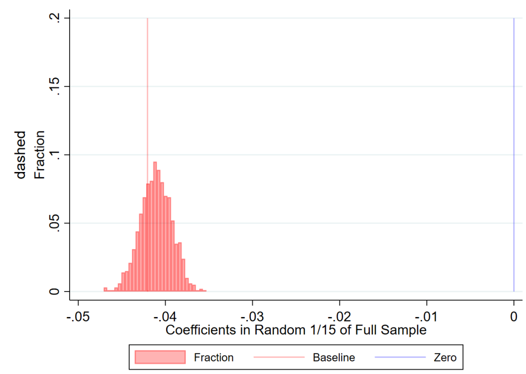
[27]To ensure that our results are not driven by a few selected feminine or masculine words, we recalculate our key measure of interest—the share of feminine words out of all gendered words in a job ad—by excluding each of the ten most popular feminine or masculine words in job ads. As shown in Figures G4, the estimated coefficients using these new measures are similar to the baseline estimate.

Alternative regression specifications. We also employ alternative regression specifications: excluding all control variables, including occupation by time fixed effects to control for time-varying occupation-specific shocks (at either the two-digit or six-digit ONET-SOC code level), including occupation by job location by time fixed effects, or including firm by occupation fixed effects to address the possibilities that our findings are driven by CA firms changing their business operations, job composition, or reducing female-dominant occupations. Figure 3a shows the estimated coefficients from these alternative specifications are similar to the baseline estimate.

Figure 3: California Statute and Gendered Language Usage in Job Ads – Robustness



(a) Alternative Samples/Specifications



(b) Random Sampling

This figure shows robustness checks on our main findings. Figure 3a shows the robustness of the baseline job ad level regression results in column (6) of Table 2 using alternative samples and specifications. Figure 3b shows the robustness of the baseline job ad level regression results in column (6) of Table 2 using alternative draws from the full sample. A random 1/15 of the full sample is drawn each time, for 1000 times. The top row in Figure 3a and the red line in Figure 3b indicate the baseline estimate. The blue line in Figure 3b indicates zero treatment effect.

Augmenting the feminine and masculine word list. We re-calculate the BoW measure expanding the Gaucher et al. (2011)'s list with a gendered word list provided by Ongig, an industry leader committed to eliminating biased job ads, and find similar results, as shown in the last row of Figure 3a.

Random sampling. In our baseline estimation, we use a randomly drawn subsample that is one fifteenth of the full sample. To confirm that our findings are not driven by chance, Figure 3b presents the distribution of the estimated coefficients for 1,000 randomly drawn subsamples, each corresponding to one fifteenth of the full sample, and shows that the estimated coefficients center around the baseline estimate and are far away from zero.

5 The Backlash Interpretation: Heterogeneous Effects of SB 826

If the increase in female leadership following SB 826 had signaled a stronger support for female employees, we would have expected to see a corresponding rise in gendered labor demand favoring women. Instead, we observe a significant decline in the use of feminine language in job ads and in the share of female new hires. To explain this seemingly counter-intuitive finding, we draw on two bodies of literature that offer microfoundations for backlash when externally imposed (legislative) changes conflict with prevailing social norms.

First, the reactance theory in psychology (Brehm, 1966; Brehm and Brehm, 2013) posits that individuals resist when they perceive a threat to their autonomy or freedom of choice. In our context, externally imposed pressure to appoint more women to corporate boards may be perceived as such a threat, leading to feelings of constraint and triggering reactance. In response, members of an affected firm may attempt to restore their autonomy by acting in direct opposition to such a constraint, resulting in a lower demand for female labor, as observed in our analysis.

Second, a long-standing economics literature examines how laws and institutions interact with social norms and can trigger backlash. Benabou and Tirole (2011) model the interplay between laws and norms, showing that laws can sometimes crowd out and undermine social norms, particularly when the expressive role of law is invoked.[28] Acemoglu and Jackson (2017) show that when laws conflict with prevailing social norms, for example, when they attempt to constrain behavior relative to the distribution of anticipated payoff-relevant behavior (i.e., social norms) in society, more people will elect to break the law. In a similar vein, Wheaton (2022) highlights a social form of crowding out (Bowles and Polania-Reyes, 2012): individuals resist legal changes in an effort to preserve their preferred norms. He cautions that aggressive legislative efforts to drive social change may incur significant costs in the form of cultural backlash. Empirically, previous studies have documented similar backlash following the implementation of major US social policy laws.[29] Mirroring those earlier findings in our context, board gender quotas may similarly clash with entrenched gender stereotypes, prompting behaviors that ultimately undermine efforts to promote gender equality in the workplace.

In summary, under either mechanism, efforts to empower women through board gender

[28] Pioneering legal scholarship has explored the expressive role of law and its effects on cultural norms and attitudes (Sunstein, 1996; Cooter, 1998; Posner, 2009). In short, beyond its functional role, law also serves to express and shape a society's values.

[29] Using historical data and surveys, Ang (2019) examines the long-run effects of federal oversight of election laws through the US Congress' 1975 revision to the Voting Rights Act (VRA) and provides evidence of a backlash among the white majority, who subsequently reduced their support for the Democratic party. Using survey data, Wheaton (2022) shows that after the passage of state-level Equal Rights Amendments (ERA) in the 1970s, men in states that adopted the ERA became less likely to support gender equality.

quotas may backfire. Although some studies emphasize the role of board expertise (Güner et al., 2008), our analysis does not rely on board members being directly involved in operational or hiring decisions. Instead, we view the board quota as a salient symbolic intervention that can provoke broader organizational backlash. It is also worth noting that psychological reactance and perceived norm violation can operate independently, or the latter may amplify the former, particularly when individuals view quotas as conflicting with deeply held gender beliefs. In such cases, this perceived misalignment can intensify feelings of threat or loss of autonomy.[30] We explore these potential mechanisms by examining heterogeneity in SB 826's effects across different cultural, political, and occupational contexts.

Table 7: Effect of California Statute on Gendered Language Usage in Job Ads – Heterogeneity: Individualism, Partisanship, and Masculinity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)							
Flag_CA × Post_2018m9	-0.037***	-0.040***	-0.021*	-0.038***	0.074	-0.037***	0.07	-0.035***
	[0.010]	[0.010]	[0.011]	[0.010]	[0.047]	[0.010]	[0.047]	[0.010]
Flag_CA × Post_2018m9 × TFE (in decades)	-0.005***						-0.003**	
	[0.001]						[0.001]	
Flag_CA × Post_2018m9 × > 20 years of TFE		-0.012***						-0.007**
		[0.003]						[0.003]
Flag_CA × Post_2018m9 × %Republican vote			-0.058***				-0.048***	
			[0.017]				[0.016]	
Flag_CA × Post_2018m9 × > 50% Republican vote				-0.026***				-0.024***
				[0.006]				[0.005]
Flag_CA × Post_2018m9 × Masculinity (0-1)					-0.254**		-0.203**	
					[0.096]		[0.092]	
Flag_CA × Post_2018m9 × > Median masculinity						-0.013**		-0.010**
						[0.005]		[0.005]
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES	YES	YES	YES	YES
Occupation (ONET 6-digit) FE	YES	YES	YES	YES	YES	YES	YES	YES
Obs	1,575,461	1,575,461	1,583,031	1,583,031	1,579,623	1,579,623	1,569,864	1,569,864
Adj R2	0.434	0.434	0.434	0.434	0.434	0.434	0.434	0.434

This table shows heterogeneities in the effect of SB 826 on gendered language usage in job ads across counties with differing levels of individualism, partisanship, and gender norms. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. In column (1), TFE (in decades) measures a county's total frontier experience in decades, following the definition in Bazzi et al. (2020). In column (2), >20 years of TFE is a dummy variable that equals 1 for a county (county of a job) with more than 20 years of total frontier experience (approximately the 75th percentile) and 0 otherwise. In column (3), %Republican vote is the percentage of voter support for the Republican candidate in the 2020 US presidential election in a county (county of a job). In column (4), >50% Republican vote is a dummy that equals 1 for a county (county of a job) receiving more than 50% voter support for the Replication candidate in the 2020 US presidential election, and 0 otherwise. In column (5), Masculinity (0-1) is Hofstede's 2010 cultural dimension index (feminine-masculinity) in a county (county of a job), based on the approach in McLean et al. (2023). In column (6), > Median masculinity is a dummy that equals 1 if a county's masculinity score is above the sample median, and 0 otherwise. Columns (7) and (8) jointly examine the heterogeneous effects of individualism, partisanship, and masculinity. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

[30] Both mechanisms can be understood as psychological responses rooted in deeply held values and beliefs about gender roles. Because such values and beliefs are long-lasting, these responses are typically triggered by institutional changes (e.g., the paper cited above; Di Tella et al. (2007), Lowes et al. (2017)) or disruptive events (e.g., terrorist attacks in Ahern (2018)).

Individualism. Cultural norms that emphasize personal autonomy influence how individuals respond to externally imposed mandates, such as gender quotas. We proxy these autonomy-oriented values using county-level Total Frontier Experience (TFE) from Bazzi et al. (2020), which captures historical exposure to the American frontier, a context that fostered enduring norms of individualism, self-reliance, and resistance to authority. These traits are closely linked to psychological reactance, a response in which individuals push back against perceived infringements on personal freedom (Iyengar and Lepper, 1999; Miron and Brehm, 2006; Steindl et al., 2015). Supporting this interpretation, Bian et al. (2022) find that individualistic norms undermine compliance with collective actions during the COVID-19 pandemic. To test whether SB 826 provokes stronger backlash in more individualistic counties, we interact the DID term with TFE in column (1) of Table 7. The negative and significant coefficient on the triple interaction term suggests that top-down gender mandates face greater resistance in counties where cultural norms emphasize autonomy and freedom of choice. In column (2), we use an indicator for counties with more than 20 years of frontier experience and find that the negative effect is nearly 30% greater in these counties, underscoring the role of psychological reactance. These effects remain robust when we use a more stringent specification with Firm $\times$ Year-Month fixed effects in Table G6, which helps control for firm-level time-varying confounders and isolates within-firm variation based on job location.

Partisanship. Partisan identity is a strong predictor of attitudes toward gender equality policies. Individuals with conservative views or those who are Republican-leaning are more likely to perceive gender quotas as conflicting with gender norms and principles of merit-based career advancement. As a result, such quotas are more likely to trigger both psychological reactance and norm-based resistance among those individuals. To test whether the effects of SB 826 vary by political context, we examine county-level partisanship. Specifically, we add an interaction term between the DID term and the percentage of county-level voter support for the Republican candidate in the 2020 US presidential election in column (3) of Table 7.[31] The triple interaction term yields a negative and significant coefficient, suggesting that the backlash effect of SB 826 was stronger in politically conservative counties. Notably, we still observe a negative and significant effect in Democratic-leaning counties, though the magnitude is substantially smaller. In column (4), we use an indicator for counties where the Republican vote share exceeded 50% and find that the negative effect increased by nearly 70% compared to less conservative counties, further supporting the role of partisanship in shaping SB 826's reception.

Does the same heterogeneous response extend to hiring? Given that we use a firm-month

[31] The results changed little if we use the 2016 US presidential election data.

Table 8: Effect of California Statute on Hiring – Heterogeneity: Partisanship and Employee Gender Composition

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Share of female new hires out of all new hires (%Fem new hire)					
Flag_CA × Post_2018m9	0.010	0.004	0.013	-0.004	0.031*	0.007
	[0.010]	[0.007]	[0.011]	[0.003]	[0.017]	[0.007]
Flag_CA × Post_2018m9 × %Republican contribution	-0.034**				-0.032*	
	[0.017]				[0.016]	
Flag_CA × Post_2018m9 × >50% Republican contribution		-0.014*				-0.013*
		[0.007]				[0.007]
Flag_CA × Post_2018m9 × %Male employee			-0.039*		-0.039*	
			[0.021]		[0.021]	
Flag_CA × Post_2018m9 × Top quartile %male employee				-0.017**		-0.016**
				[0.007]		[0.007]
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES
Obs	69,785	69,785	69,785	69,785	69,785	69,785
Adj R2	0.426	0.426	0.437	0.426	0.437	0.426

This table shows heterogeneities in the effect of SB 826 on the gender composition of newly hired employees across firms with differing levels of partisanship and existing employee gender makeup. The outcome variable is the share of female new hires out of all new hires. The unit of observation is a firm-month. The sample includes public firms covered by LinkedIn over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. In column (1), %Republican contribution is the share of a firm's political contributions to Republican candidates out of its total political contributions. In column (2), >50% Republican contribution is a dummy indicating whether more than 50% of a firm's political contributions go to Republican candidates. In column (3), %Male employee is the share of male employees out of all employees in the corresponding firm-month. In column (4), Top quartile %male employee is a dummy indicating whether the corresponding firm-month falls within the top quartile of the %Male employee distribution. Columns (5) and (6) jointly examine the heterogeneous effects of partisanship and employee gender composition. Stand-alone terms and other two-way interaction terms are suppressed in the table. The sample excludes firms with fewer than 200 employees from LinkedIn (bottom quartile) to mitigate coverage bias in the LinkedIn data. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

panel for the hiring analysis, we interact the DID term with a measure for firm-level political leaning. We expect Republican-leaning firms to exhibit stronger pushback compared to their Democratic-leaning peers. To capture a firm's political leaning, we collect information on its contributions to the Republican versus Democratic candidates based on data from the Federal Election Commission. In column (1) of Table 8, we find a negative and significant coefficient on the triple interaction term involving the share of a firm's contributions to the Republican candidates; the coefficient on the DID term is no longer significant. This suggests that the negative effect of SB 826 on female new hires was concentrated in Republican-leaning firms, consistent with the backlash interpretation. In column (2), we use an indicator variable, $>$ 50% Republican contribution, flagging firms that contributed more to the Republican candidates than to Democrat candidates. We again find a negative and significant coefficient on the triple interaction term.

Gender norms. Cultural norms, particularly those associated with gender, can shape how individuals respond to gender equality policies. In regions where masculine values such as competitiveness, assertiveness, and adherence to traditional gender roles are more prevalent, board gender quotas may be perceived as clashing with norms, triggering stronger resistance. To examine this, we use a county-level masculinity index ranging from 0 to 1, following the methodology of McLean et al. (2023), based on Hofstede's (Hofstede, 1980; Hofstede and Hofstede, 2001) cultural dimensions. Column (5) of Table 7 shows that the negative effect of SB 826 was significantly stronger in counties with higher masculinity scores, supporting the interpretation that norm violation and psychological reactance are more pronounced in such contexts. In column (6), we use an indicator for counties with above-median masculinity scores. The results show that the negative effect of SB 826 was more than 35% larger in these counties compared to those with below-median masculinity scores.

One key advantage of our job ad data is its granularity at the occupational level, which allows us to further examine heterogeneity across occupations. In particular, occupations with wider gender wage gaps may reflect more entrenched gender hierarchies in the workplace. In such contexts, board gender quotas are more likely to be perceived as violating prevailing norms, triggering resistance. To test this, we examine whether the effect of SB 826 varied with gender wage gap across occupations. In Table 9, we measure the gap using the 2020 Labor Force Statistics from the BLS Current Population Survey, defined as the difference in earnings between male and female employees within the same occupation. We find a negative and significant coefficient on the triple interaction term, indicating stronger resistance to SB 826 in occupations with greater gender pay disparities and thus more biases against female employees.

For firm-level hiring analysis, we turn to the gender composition of a firm's workforce as a proxy for its gender norms in the workplace. Specifically, we examine whether firms with a higher share of male employees, potentially reflecting more male-dominated workplace cultures, exhibit stronger pushback following SB 826. In column (3) of Table 8, we interact the DID term with a continuous measure of the male employee share and find a negative and significant coefficient. This suggests that the drop in female new hires was more pronounced in firms with male-dominated workforce. In column (4), we use an indicator for firms in the top quartile of the male employee share distribution and again observe a negative and significant effect, reinforcing the interpretation that gender norms within firms play an important role in shaping the intensity of the backlash. This finding is consistent with Deschamps (2023), who shows that following a 2015 French law mandating gender quotas in academic hiring committees, male-chaired search committees hired fewer women.[32]

[32]While sharing the insight that demand-side backlash can undermine gender quotas, our study shows that such resistance is broader and more pervasive. It extends beyond specific institutional settings to widely debated board gender quotas in the business world, permeating every stage of a female employee's experience with her employer, even though corporate directors are not directly involved in hiring decisions. Our study also unveils reactance and norm violation as the underlying mechanisms.

Table 9: Effect of California Statute on Gendered Language Usage in Job Ads – Heterogeneity: Occupation Level Gender Wage Gap

	(1)	(2)	(3)	(4)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)			
Flag_CA × Post_2018m9	0.003	-0.028***		
	[0.010]	[0.009]		
Flag_CA × Post_2018m9 × Gender wage gap	-0.205***		-0.252***	
	[0.040]		[0.032]	
Flag_CA × Post_2018m9 × > Median gender wage gap		-0.019**		-0.012**
		[0.009]		[0.006]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	-	-
Industry × Year-Month FE	YES	YES	-	-
Firm × Year-Month FE	NO	NO	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES
Occupation (ONET 6-digit) FE	YES	YES	YES	YES
Obs	785,558	785,558	785,558	785,558
Adj R2	0.472	0.472	0.508	0.508

This table shows heterogeneities in the effect of SB 826 on gendered language usage in job ads across occupations with different gender wage gap. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. In columns (1) and (3), Gender wage gap is the wage gap between male and female employees within the same occupation, based on the 2020 BLS Current Population Survey (CPS). In columns (2) and (4), > Median gender wage gap is a dummy that equals 1 if an occupation's gender wage gap is above the median, and 0 otherwise. Columns (3) and (4) include Firm × Year-Month fixed effects. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

One might argue that individualism, partisanship, and masculinity norms are correlated, and that the observed heterogeneity is driven by only one of those variables. To explore this possibility, columns (7) and (8) of Table 7 include triple interaction terms for these variables in the same regression specification. The results remain similar to those from the single triple interaction term specification for each variable, suggesting that values emphasizing autonomy, political ideology, and cultural attitudes toward gender roles independently influenced CA firms' responses to board gender quotas. The independent contributions of the latter two variables are further supported by the evidence on hiring outcome in columns (5) and (6) of Table 8.

Taking stock. This section presents evidence that the negative effects of SB 826 on gendered language in job ads and female hiring were more pronounced in contexts where backlash was more likely, such as in Republican-leaning counties or firms, and in counties or firms with high masculinity norms. These patterns are consistent with the interpretation that resistance to SB 826 was driven by its perceived violation of the prevailing social (gender) norms. Importantly, Table 7 points to a role for psychological reactance in response to gender quotas, as indicated by the negative and significant standalone coefficients in the even-numbered columns. This suggests that pushback occurred even in high individualism, Democratic-leaning, or strong masculinity counties, where the norm conflict is presumably lower—likely driven by a general resistance to externally imposed mandates. The observed heterogeneities across counties and occupations also offer rare insights into who is responsible for drafting job ads within firms. If job ads were centrally produced, we would not expect to observe systematic variation in language across counties with different cultural or political characteristics. Instead, our findings highlight the influence of local context and job-specific factors in shaping how firms communicate their labor demand.

6 Additional Evidence: Post-Hiring Experience

While much of the analysis thus far has focused on hiring, it is also important to examine whether pushback extends beyond women's entry into the labor market. To do so, we leverage granular data on employee reviews, ratings, and gender-related labor violations.

We first study workplace accommodations using the same DID approach and the results are presented in Panel A of Table 10. We find that employees mentioned female-friendly amenities less frequently (albeit not significant) in the pros section (column (1)) but more often in the cons section (column (2)). When we examine the difference between the pros and cons sections in terms of mentions of these amenities, we observe a negative coefficient that is significant at the 1% level (column (3)). This suggests that the overall sentiment regarding these amenities has become more negative. In column (4), we further take into account the length of the review text and use the weighted difference in word counts between the pros and cons sections. The coefficient is similar to that of column (3), confirming a deterioration in employees' perceptions of workplace accommodations toward female employees.

Panel B of Table 10 complements the above text-based measures and analysis by turning to more structured employee ratings to further investigate shifts in workplace sentiment after SB 826's passage. We find strong evidence that the bill resulted in a significant drop in employee ratings of female-friendly benefits in CA firms. The effect is economically large, with a 6.2% (0.174/2.81 where 2.81 is the mean value) drop in the mean rating for culture and values, for example. It is worth noting that the anonymity of Glassdoor reviews allows employees to share opinions about their current or former employers without fear of retaliation. Therefore, it is unlikely that our results are driven by (newly added) female directors (after SB 826) prompting employees to report poor gender-related practices or misconduct.

Beyond employee reviews and ratings, we also explore workplace treatments of female employees, including gender discrimination and sexual harassment that result in gender-related labor violations and penalties. Table G7 in the Online Appendix presents the results. We find that SB 826 led to a significant increase in both the number of gender-related labor violations and the penalties paid for those violations.

Table 10: Effect of California Statute on Employee Reviews and Ratings

Panel A: Reviews of Female-Friendly Amenities

	(1)	(2)	(3)	(4)
Dep. Var.	Coverage of 8 amenities in Glassdoor job reviews (word frequency)			
	Pros	Cons	Pros - Cons	Pros - Cons (weighted)
Flag_CA × Post_2018m9	-0.035	0.048*	-0.043***	-0.044**
	[0.023]	[0.028]	[0.012]	[0.020]
List of words for amenities	Anchor	Anchor	Anchor	Anchor
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	52,856	52,856	52,856	52,856
Adj R2	0.056	0.087	0.077	0.083

Panel B: Ratings of Female-Friendly Amenities

	(1)	(2)	(3)	(4)
Dep. Var.	Ratings (1-5)			
	Overall	Compensation and benefits	Culture and values	Work-life balance
Flag_CA × Post_2018m9	-0.052***	-0.111***	-0.174***	-0.079***
	[0.018]	[0.030]	[0.025]	[0.025]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	173,019	164,936	164,479	164,890
Adj R2	0.128	0.094	0.074	0.07

This table shows the effect of SB 826 on employee reviews and ratings of female-friendly amenities. The unit of observation is an employee review or rating. The sample includes all Glassdoor reviews in English from full-time former employees whose review text in the pros (cons) section has more than 50 characters. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Panel A shows the effect of SB 826 on employee reviews of female-friendly amenities. The outcome variable is the number of anchor words related to eight female-friendly amenities in the pros (cons) section in column (1) (column (2)). The outcome variable is the difference in the number of anchor words between the pros and cons sections in column (3). Columns (4) takes into account the length of review text in a pros (cons) section using the weighted difference in the number of anchor words between the pros and cons sections. The anchor words for all amenities are provided in Appendix Section I of Sockin (2022). Out of the 50 amenities, we identify eight amenities that are important for female employees in the workplace (5-Paid time off, 10-Work-life balance, 11-Hours, 12-Work schedule, 16-Teleworking, 19-Respect/abuse, 21-Support, and 31-Diversity/inclusion). Table F1 in the Online Appendix provides the anchor word list of the eight amenities. Panel B shows the effect of SB 826 on employee ratings of female-friendly benefits. The outcome variables are the overall rating, and ratings for compensation and benefits, culture and values, and work-life balance. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

So far, the evidence from three different data sources—employee reviews, ratings, and labor violations—consistently points to shifts toward a more hostile work environment for female employees after the passage of SB 826. Will these changes ultimately lead to higher exit rates of female employees in affected firms? Table 11 provides the answer. We show that SB 826 was followed by a 1 percentage-point increase in the share of females among departed employees. This finding remains excluding the COVID-19 period and restricting the sample to industries and firms with reasonable coverage in LinkedIn (columns (2) to (4)).

Table 11: Effect of California Statute on Employee Departure

	(1)	(2)	(3)	(4)
Dep. Var.	Share of female departed employees out of all departed employees (%Fem depature)			
Sample	Full	Excl. COVID	Excl. low coverage (industry)	Excl. low coverage (firm)
Flag_CA × Post_2018m9	0.010***	0.008**	0.015***	0.010***
	[0.004]	[0.004]	[0.004]	[0.003]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	78,384	68,977	61,626	69,289
Adj R2	0.273	0.273	0.269	0.376

This table shows the effect of SB 826 on the gender composition of newly departed employees. The outcome variable is the share of female departed employees out of all departed employees. The unit of observation is a firm-month. The sample includes public firms covered by LinkedIn over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Columns (1) employs the full sample. Columns (2) includes only the pre-COVID period (before March 2020). Columns (3) excludes industries known for low LinkedIn coverage. The following 20 industries are excluded: agricultural production crops, forestry, fishing, hunting, and trapping, metal mining, nonmetallic mineral mining and quarrying, heavy construction, construction special trade contractors, textile mill products, leather and leather products, stone, clay, glass, and concrete products, motor freight transportation and warehousing, water transportation, transportation by air, transportation services, wholesale trade-durable goods, wholesale trade-nondurable goods, food stores, automotive dealers and gasoline service stations, hotels, rooming houses, camps, and other lodging places, business services (certain sub-sectors), and motion pictures. Columns (4) excludes firms with fewer than 200 employees from LinkedIn (bottom quartile) to mitigate coverage bias in the LinkedIn data. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

7 Delineating The Underlying Mechanisms Through An Experiment

We interpret our main findings as evidence of proactive pushback, in which external pressure such as a board gender quota triggers psychological reactance due to perceived imposition and loss of autonomy. Additionally, such quotas may conflict with deeply held gender beliefs within organizations, further fueling resistance. Both mechanisms can lead individuals to resist such quotas by favoring male candidates in hiring decisions. However, alternative explanations exist. One possibility is moral licensing (Monin and Miller, 2001; Merritt et al., 2010; List and Momeni, 2021): "good" actions (e.g., supporting greater female representation on corporate boards) may serve to justify other actions that undermine gender equalities (e.g., reduced support for women in general workplace settings). Another explanation is that our findings simply reflect rational behavior by employees seeking to protect their own positions, because the board gender diversity mandate may lead employees to perceive (incoming) female hires as being "supported" or "favored," prompting resentment or defensive actions.

Although the heterogeneity analyses along political and gender lines in Section 5 provide some support for our proposed mechanisms, they cannot rule out these alternative interpretations. Differentiating among these mechanisms is particularly challenging, as individual motives are rarely observable in real-world data. To address this limitation, we conduct an experiment in which we manipulate the presence of a board gender quota policy and directly elicit responses tied to the competing mechanisms.

7.1 Experimental design

Employing a between-subjects design, participants were asked to imagine themselves as employees of a hypothetical firm to assist the recruitment of a specific position, and were then randomly assigned to either a treatment or control condition. In the treatment condition, participants read a policy vignette describing a government-mandated gender quota requiring a specified percentage of women on the company's board. In the control condition, participants read an otherwise similar vignette outlining a recruitment policy based solely on qualifications and experience, with no mention of quotas. By keeping the company context and vignettes similar in wording and length, aside from the quota element, this design ensures that any observed differences in responses can be attributed to the presence of the gender-quota policy. Table C1 provides the full text of both vignettes along with the pre-trial instructions.

We recruited 200 US-based participants through CloudResearch's Connect platform. To ensure demographic representiveness, we applied selection criteria aligned with the latest US Census benchmarks (Female = 50%; Republican = 50%; White = 78%, Black/African American = 14%, Other = 8%; Age 18-29 = 22%, Age 30-44 = 26%, Age 45-59 = 26%, Age 60+ = 26%). All participants had at least one year of professional work experience. To achieve geographic diversity, we drew the sample using state- and region-level sampling proportions based on population density.[33]

The experiment was approved by an institutional ethical review board and pre-registered with the American Economic Association's RCT Registry. Participants received a $1 incentive, and the average survey completion time was approximately four minutes. An attention check was embedded within the survey; two participants who failed this check were excluded from the final analysis.

Following a manipulation check, participants were asked to share their opinions about their hypothetical employer's recruitment policy (either treatment or control) by responding to questions adapted from prior research to evaluate four economic mechanisms: psychological reactance, norm violation, moral licensing, and job insecurity. These questions employed a multi-item, seven-point Likert scale, as shown in Table C2. To eliminate order effects, the presentation of mechanisms was randomized. Further details on our experimental design, including the manipulation and attention check questions, are provided in Online Appendix C. Table C3 confirms that the treatment and control groups are comparable across age, gender, race, work experience, and political view, supporting the validity of our randomization approach.

[33] Although our sample size may appear small, this biases against detecting significant effects. In contrast, larger, non-clustered experiments may yield significant results even for minor effects.

7.2 Experimental evidence on the mechanisms

Table 12 presents experimental evidence on the underlying mechanisms by comparing responses between the treatment and control groups. In column (1) of Panel A, we first report a significant manipulation effect, confirming that the treatment successfully primed participants as intended. Columns (2) to (5) of Panel A then examine the proposed economic mechanisms. Columns (2) and (3) assess whether the treatment induces psychological reactance and perceptions of gender norm violation. We find positive and significant coefficients at the 1% level. The economic magnitude is large, exceeding 50% of the sample mean for psychological reactance and over 25% for perceived gender norm violation. These findings suggest that treated participants are significantly more likely than control participants to view their employer's recruitment policy as infringing on their decision-making autonomy and conflicting with established workplace norms.

Columns (4) and (5) investigate the roles of moral licensing and job insecurity. For both mechanisms, we find small and insignificant differences between the treatment and control groups. The results remain consistent when individual-level control variables are excluded from the regressions, as shown in Panel A of Table G8. In addition, the R-squared values are minimal (less than 0.5%) without controls, indicating that these mechanisms do not meaningfully account for treated participants' reactions to the board gender quota.

One potential concern with our experimental design is that some participants may lack recruiting experience, which could limit the real-world relevance of their responses. To address this issue, we restrict the sample to individuals over the age of 30 with at least five years of professional work experience and re-run the analysis. Panel A of Table G9 presents the results, which are very similar to those in Table 12.

In Panel B of Table 12, we take the analysis a step further by examining whether the strength of participants' reactions depend on their gender-related views and preferences. The underlying idea is that the degree to which board gender quotas diverge from an individual's personal views determines the treatment intensity, allowing us to explore the heterogeneous treatment effects.

We first study the role of political orientation, comparing responses from Democrat- and Republican-leaning participants. To do this, we include an interaction term between the treatment and an indicator for Republican-leaning participants in the regression. As shown in column (1), Republican-leaning participants exhibit significantly stronger psychological reactance to the quota policy. This heightened reactance appears to be partially driven by the fact that these participants are more likely to view the policy as misaligned with workplace gender norms (column (4)). In contrast, Democrat-leaning participants show a much smaller increase in reactance (column (1)), and their perceived norm violation is

Table 12: Experimental Evidence on Mechanisms

Panel A: Comparing Treated and Control Participants

	(1)	(2)	(3)	(4)	(5)
		Mechanisms			
Dep. Var.	Manipulation check	Reactance	Norm violation	Moral licensing	Insecurity
Treatment	1.876***	1.473***	0.609***	0.261	0.201
	[0.222]	[0.247]	[0.214]	[0.186]	[0.206]
Controls	YES	YES	YES	YES	YES
Dep. Var. Mean	4.455	2.865	2.587	4.25	2.208
Obs	198	198	198	198	198
Adj R2	0.344	0.216	0.076	0.058	0.014

Panel B: Heterogeneity - Partisanship and Gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Mechanisms					
Dep. Var.	Reactance			Norm violation		
Treatment × Republican	1.796***		1.641***	1.869***		1.720***
	[0.470]		[0.484]	[0.420]		[0.439]
Treatment × Male		1.119**	0.737		1.097**	0.712*
		[0.488]	[0.492]		[0.433]	[0.428]
Treatment	0.621**	0.910***	0.323	-0.280	0.057	-0.568*
	[0.282]	[0.332]	[0.332]	[0.284]	[0.291]	[0.306]
Controls	YES	YES	YES	YES	YES	YES
Dep. Var. Mean	2.867	2.867	2.867	2.586	2.586	2.586
Obs	198	198	198	198	198	198
Adj R2	0.27	0.234	0.275	0.163	0.103	0.171

This table provides experimental evidence on the mechanisms. The unit of observation is a subject. The sample includes 198 participants following the US Census racial distribution who passed the attention test. Treatment is a dummy that equals 1 for the treatment group and 0 for the control group. Panel A reports the treatment effect estimated from the experiment. The outcome variable is in the seven-point Likert scale for the manipulation question in column (1) and for the mechanism questions in columns (2) to (5). Panel B reports heterogeneous treatment effects based on a subject's partisanship and gender. Republican is a dummy that equals 1 if a subject's political view is greater than 4 on a 1-7 scale (ranging from liberal to conservative), and 0 otherwise. Gender is a dummy that equals 1 for male participants and 0 otherwise. The following control variables are included: gender, partisanship, age, ethnicity, employment status, years of work experience. Robust standard errors adjusted for heteroskedasticity are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

insignificant (column (4)).

We next explore whether the strength of these mechanisms varies by gender. The results reveal clear differences: male participants exhibit significantly greater psychological reactance than female participants (column (2)), a pattern that aligns with their elevated perceptions of gender norm violations (column (5)). When political affiliation and gender are examined jointly in columns (3) and (6), both factors show independent effects on participants' reactions.[34] These findings suggest that psychological reactance and perceived norm violation may reinforce each other, contributing to labor demand responses to gender quotas.

[34] Panel B of Table G8 and Table G9 further show that these heterogeneous effects remain when additional controls are removed and when the sample is restricted to experienced participants, respectively.

The stronger effects observed among Republican-leaning and male participants align with the heterogeneity analysis by political orientation and gender norms presented in Section 5. Together, the experimental evidence complements our analyses with real-world data by unveiling psychological reactance and gender norm violation as key mechanisms driving our main findings. These findings have important implications for policy design. Understanding which groups are most prone to backlash allows policymakers to tailor communication strategies and implement early interventions to mitigate reactance and perceptions of norm violation. More broadly, rigid, state-imposed quotas may provoke stronger oppositions and risk undermining the intended policy goals. In contrast, more flexible approaches, such as "comply or explain" frameworks, may help mitigate these unintended consequences.[35] Supporting our US-based evidence, we find that backlash is largely absent in European countries with soft quotas. Specifically, when categorizing ten European quota policies by their enforcement regimes—state-mandated versus voluntary targets (see Table D1)—Panel B of Table D4 shows that adverse outcomes are concentrated in countries with hard quotas.

8 Conclusion

Using California's Senate Bill 826 enacted in 2018 as an exogenous shock that imposed board gender quotas on locally headquartered firms, we study the broader implications of such quotas for women in the corporate sector. Applying computational linguistic methods to job ads to construct a clean measure of firms' demand for female labor, and employing a difference-in-differences specification, we first establish that SB 826 led to a reduction in CA firms' demand for female labor. Using aggregated individual profiles from LinkedIn, we further show that following the passage of SB 826, CA firms significantly reduced their share of female new hires. To better understand these negative hiring outcomes, we examine cross-sectional variation and find that the adverse effects were significantly stronger in counties or firms characterized by high individualism, Republican-leaning, or strong masculinity norms. Beyond hiring, we also show that SB 826 led to poorer workplace treatment of female employees and a significant increase in female employee turnover. We then conduct a randomized controlled experiment and find that psychological reactance—driven by the top-down nature of the quota and the perceived misalignment with prevailing gender norms—are the likely explanations. We conclude that, despite their intentions, board gender quotas may backfire, resulting in worse labor market outcomes for women.

We caution that our study explores rather short-run outcomes, due to the repeal of SB 826 in 2022 and the onset of the COVID-19 pandemic in early 2020. One natural question

[35] Giannetti and Wang (2023) suggest that raising public awareness of gender equality issues can also serve as a policy instrument. Bennedsen et al. (2022) show that regulatory mandates on gender pay transparency are effective. Unlike a more forceful intervention such as quotas, these tools are generally perceived as neutral and procedural, making them less likely to trigger the mechanisms underlying backlash.

is whether the effects we document are transitory and ultimately outweighed by longer-term benefits of SB 826. While we do not directly assess long-term outcomes, several observations suggest that the negative effects we document remain important. First, prior research by Bertrand et al. (2019), along with our own cross-country evidence (Table D4), shows that board gender quotas can have persistent negative effects years after implementation. Second, the magnitude of the short-run backlash effect we estimate—equivalent to a reversal of three to four years of progress in female labor market participation—is economically significant. This underscores the potential for policy design to mitigate such costs and enhance the long-run net benefits. Third, the mechanisms we identify, rooted in human psychology and social norms, suggest that resistance may be long-lasting. Future research should build on these insights by examining longer-term effects and a broader range of institutional settings to deepen our understanding of the impact of gender diversity policies.

Taken together, our findings highlight the complex welfare implications of gender equality policies, especially when affected firms fail to internalize the social costs associated with such interventions. More broadly, our findings suggest that when governments introduce laws that deviate from prevailing social norms, the potential for societal pushback should be considered when evaluating the overall effectiveness of those laws.

References

Acemoglu, D. and J. D. Angrist (2001). Consequences of Employment Protection? The Case of the Americans with Disabilities Act. *Journal of Political Economy 109*(5), 915–957.

Acemoglu, D. and M. O. Jackson (2017). Social Norms and The Enforcement of Laws. *Journal of the European Economic Association 15*(2), 245–295.

Adams, R. B. and D. Ferreira (2009). Women in the Boardroom and Their Impact on Governance and Performance. *Journal of Financial Economics 94*(2), 291–309.

Ahern, K. R. (2018). The Importance of Psychology in Economic Activity: Evidence from Terrorist Attacks. Technical report, National Bureau of Economic Research.

Ahern, K. R. and A. K. Dittmar (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *Quarterly Journal of Economics 127*(1), 137–197.

Allen, A. and A. S. Wahid (2024). Regulating Gender Diversity: Evidence from California Senate Bill 826. *Management Science 70*(4), 2023–2046.

Ang, D. (2019). Do 40-year-old Facts Still Matter? Long-run Effects of Federal Oversight under the Voting Rights Act. *American Economic Journal: Applied Economics 11*(3), 1–53.

Bailey, M. J., T. Helgerman, and B. A. Stuart (2024). How the 1963 Equal Pay Act and 1964 Civil Rights Act Shaped the Gender Gap in Pay. *The Quarterly Journal of Economics 139*(3), 1827–1878.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020). Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States. *Econometrica 88*(6), 2329–2368.

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

Benabou, R. and J. Tirole (2011). Laws and Norms. Technical report, National Bureau of Economic Research.

Bennedsen, M., E. Simintzi, M. Tsoutsoura, and D. Wolfenzon (2022). Do Firms Respond to Gender Pay Gap Transparency? *The Journal of Finance 77*(4), 2051–2091.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *Review of Economic Studies 86*(1), 191–239.

Bian, B., J. Li, T. Xu, and N. Z. Foutz (2022). Individualism During Crises. *Review of*

Economics and Statistics 104(2), 368–385.

Bohm, S., O. Linnyk, J. Kohl, T. Weber, I. Teetz, K. Bandurka, and M. Kersting (2020). Analysing Gender Bias in IT Job Postings: A Pre-Study Based on Samples from the German Job Market. In *Proceedings of the 2020 on Computers and People Research Conference*, SIGMIS-CPR'20, New York, NY, USA, pp. 72–80. Association for Computing Machinery.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. *arXiv preprint arXiv:2108.12419*.

Bowles, S. and S. Polania-Reyes (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature 50*(2), 368–425.

Brehm, J. W. (1966). A Theory of Psychological Reactance.

Brehm, S. S. and J. W. Brehm (2013). *Psychological Ractance: A Theory of Freedom and Control*. Academic Press.

Burn, I., P. Button, T. Figinski, W. McGrew, and D. Nelson (2025). Help Really Wanted? Age Stereotypes in Job Ads. *Journal of Labor Economics 43*(2), 259–297.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with Multiple Time Periods. *Journal of Econometrics 225*(2), 200–230.

Card, D., F. Colella, and R. Lalive (2024). Gender Preferences in Job Vacancies and Workplace Gender Diversity. *Review of Economic Studies 91*(3), 1807–1848.

Chaturvedi, A., A. Mahajan, and Z. Siddique (2025). Gendered Language in Job Ads and Applicant Behavior: Evidence from India. *Labour Economics 83*, 102509.

Cooter, R. (1998). Expressive Law and Economics. *The Journal of Legal Studies 27*(S2), 585–607.

Cryan, J., S. Tang, X. Zhang, M. Metzger, H. Zheng, and B. Y. Zhao (2020). Detecting Gender Stereotypes: Lexicon vs. Supervised Learning Methods. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–11. New York, NY, USA: Association for Computing Machinery.

Cullen, Z. B. and B. Pakzad-Hurson (2023). Equilibrium Effects of Pay Transparency. *Econometrica 91*(3), 765–802.

Del Carpio, L. and T. Fujiwara (2023). Do Gender-Neutral Job Ads Promote Diversity? Experimental Evidence from Latin America's Tech Sector. Technical report, National Bureau of Economic Research.

Deming, D. and L. B. Kahn (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics 36*(S1), S337–S369.

Deschamps, P. (2023). Gender Quotas in Hiring Committees: A Boon or a Bane for Women? *Management Science*.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.

Di Tella, R., S. Galiant, and E. Schargrodsky (2007). The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters. *The Quarterly Journal of Economics 122*(1), 209–241.

Eckbo, B. E., K. Nygaard, and K. S. Thorburn (2022). Valuation Effects of Norway's Board Gender-quota Law Revisited. *Management Science 68*(6), 4112–4134.

Gaucher, D., J. Friesen, and A. C. Kay (2011). Evidence that Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality. *Journal of Personality and Social Psychology 101*(1), 109–128.

Gertsberg, M., J. Mollerstrom, and M. Pagel (2021). Gender Quotas and Support for Women in Board Elections. Technical report, National Bureau of Economic Research.

Giannetti, M. and T. Y. Wang (2023). Public Attention to Gender Equality and Board Gender Diversity. *Journal of Financial and Quantitative Analysis 58*(2), 485–511.

Greene, D., V. J. Intintoli, and K. M. Kahle (2020). Do Board Gender Quotas Affect Firm Value? Evidence from California Senate Bill No. 826. *Journal of Corporate Finance 60*, 101526.

Güner, A. B., U. Malmendier, and G. Tate (2008). Financial Expertise of Directors. *Journal of Financial Economics 88*(2), 323–354.

Hershbein, B. and L. B. Kahn (2018). Do Recessions Accelerate Routine-biased Technological Change? Evidence from Vacancy Postings. *American Economic Review 108*(7), 1737–72.

Hofstede, G. (1980). Culture and Organizations. *International Studies of Management & Organization 10*(4), 15–41.

Hofstede, G. H. and G. Hofstede (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations.* sage.

Hwang, S., A. Shivdasani, and E. Simintzi (2021). Mandating Women on Boards: Evidence from the United States. SSRN Scholarly Paper ID 3265783, Social Science Research

Network, Rochester, NY.

Iyengar, S. S. and M. R. Lepper (1999). Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation. *Journal of Personality and Social Psychology 76*(3), 349–366.

Jin, Y., A. Bhatia, and D. Wanvarie (2021). Seed Word Selection for Weakly-supervised Text Classification with Unsupervised Error Estimation. *arXiv preprint arXiv:2104.09765*.

Kuhn, P. and K. Shen (2023). What Happens When Employers Can No Longer Discriminate in Job Ads? Evidence from China. *American Economic Review 113*(4), 1055–1091.

Levi, M., K. Li, and F. Zhang (2014). Director Gender and Mergers and Acquisitions. *Journal of Corporate Finance 28*, 185–200.

List, J. A. and F. Momeni (2021). When Corporate Social Responsibility Backfires: Evidence from a Natural Field Experiment. *Management Science 67*(1), 8–21.

Lowes, S., N. Nunn, J. A. Robinson, and J. L. Weigel (2017). The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom. *Econometrica 85*(4), 1065–1091.

Maida, A. and A. Weber (2022). Female Leadership and Gender Gap within Firms: Evidence from an Italian Board Reform. *ILR Review 75*(2), 488–515.

Matsa, D. A. and A. R. Miller (2011). Chipping Away at the Glass Ceiling: Gender Spillovers in Corporate Leadership. *American Economic Review 101*(3), 635–639.

Matsa, D. A. and A. R. Miller (2013). A Female Style in Corporate Leadership? Evidence from Quotas. *American Economic Journal: Applied Economics 5*(3), 136–169.

Mazis, M. B., R. B. Settle, and D. C. Leslie (1973). The Elimination of Phosphate Detergents and Psychological Reactance. *Journal of Marketing Research 10*(4), 390–395.

McLean, R. D., C. Pirinsky, and M. Zhao (2023). Corporate Leadership and Inherited Beliefs about Gender Roles. *Journal of Financial and Quantitative Analysis 58*(8), 3274–3304.

Mensi-Klarbach, H. and C. Seierstad (2020). Gender Quotas on Corporate Boards: Similarities and Differences in Quota Scenarios. *European Management Review 17*(3), 615–631.

Merritt, A. C., D. A. Effron, and B. Monin (2010). Moral Self-licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass 4*(5), 344–357.

Miron, A. M. and J. W. Brehm (2006). Reactance Theory–40 Years Later. *Zeitschrift fÃŒr Sozialpsychologie 37*(1), 9–18.

Monin, B. and D. T. Miller (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology 81*(1), 33.

Pechmann, C. and C.-F. Shih (1999). Smoking Scenes in Movies and Antismoking Advertisements Before Movies: Effects on Youth. *Journal of Marketing 63*(3), 1–13.

Posner, E. A. (2009). *Law and Social Norms.* Harvard University Press.

Rosenkrantz, P., S. Vogel, H. Bee, I. Broverman, and D. M. Broverman (1968). Sex-role Stereotypes and Self-concepts in College Students. *Journal of Consulting and Clinical Psychology 32*(3), 287–295.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf (2019). Distilbert, a Distilled Version of bert: Smaller, Faster, Cheaper and Lighter. *arXiv preprint arXiv:1910.01108*.

Shoss, M. K. (2017). Job Insecurity: An Integrative Review and Agenda for Future Research. *Journal of Management 43*(6), 1911–1939.

Sockin, J. (2022). Show Me the Amenity: Are Higher-Paying Firms Better All Around?

Steindl, C., E. Jonas, S. Sittenthaler, E. Traut-Mattausch, and J. Greenberg (2015). Understanding Psychological Reactance. *Zeitschrift für Psychologie*.

Sunstein, C. R. (1996). Social Norms and Social Roles. *Colum. L. Rev. 96*, 903.

Tang, S., X. Zhang, J. Cryan, M. J. Metzger, H. Zheng, and B. Y. Zhao (2017). Gender Bias in the Job Market: A Longitudinal Analysis. *Proceedings of the ACM on Human-Computer Interaction 1*(CSCW), 1–19.

Tate, G. and L. Yang (2015). Female Leadership and Gender Equity: Evidence from Plant Closure. *Journal of Financial Economics 117*(1), 77–97.

von Meyerinck, F., A. Niessen-Ruenzi, M. Schmid, and S. Davidoff Solomon (2021). As California Goes, So Goes the Nation? Board Gender Quotas and Shareholders' Distaste of Government Interventions. SSRN Scholarly Paper ID 3303798, Rochester, NY.

Wheaton, B. (2022). Laws, Beliefs, and Backlash. The University of California Los Angeles Working Paper.

Zhang, M., K. N. Jensen, S. D. Sonniks, and B. Plank (2022). Skillspan: Hard and Soft Skill Extraction from English Job Postings. *arXiv preprint arXiv:2204.12811*.

For Online Publication:

Internet Appendix to "Does Mandating Women on Corporate Boards Backfire?"

This appendix has five sections. Section A contains additional information on variable definition and construction. Section B provides details of the BERT-based method and statistics. Section C discusses the details of the experimental design. Section D presents additional analysis based on board gender quotas in Europe. Section E discusses sample representativeness while Section F introduces datasets and measures used in the post-hiring analysis. Section G covers additional analysis, including tables and figures.

Table of Contents

List of Appendix Figures

List of Appendix Tables

A Definition of Variables

Table A1: List of Variables, Definitions, and Data Sources

Variable	Definition	Source
%Fem words	Share of feminine words out of all gendered words in a job ad	LinkUp
#Fem words	Number of feminine words in a job ad	LinkUp
#Mas words	Number of masculine words in a job ad	LinkUp
#Words	Number of words in a job ad	LinkUp
Job zone	Job zones: 1 (less preparation required) to 5 (most preparation required)	ONET 26.2 Database
Mean wage	Average hourly wage of an occupation	BLS OEWS
Entry education	Typical entry-level education requirement: 0 (no formal educational credential) to 7 (Doctoral or professional degree)	BLS OEWS
STEM	Equals 1 if an occupation is in STEM fields, and 0 otherwise	BLS OEWS
#Dir	Number of directors	BoardEx
%Fem dir	Share of female directors out of all directors	BoardEx
Firm size	Size of a firm: log of total assets (ta)	Compustat
Cash holdings	Cash holdings: che/at	Compustat
Tangibility	Tangibility: ppent/at	Compustat
Tobin's Q	Tobin's Q: (at+prcc_f*csho–ceq–txditc)/at	Compustat
Leverage	Book levarage: (dltt+dlc)/at	Compustat
ROA	Profitability: oibdp/at	Compustat
Capex	Capital Investment: capx/at	Compustat
TFE (in decades)	A county's total frontier experience in decades, following the definition in Bazzi et al. (2020)	
>20 years of TFE	Equals 1 for a county (county of a job) with more than 20 years of total frontier experience (approximately the 75th percentile), and 0 otherwise	
%Republican vote	Percentage of voter support for the Republican candidate in the 2020 US presidential election in a county (county of a job)	MIT Election Lab
>50% Republican vote	Equals 1 for a county (county of a job) receiving more than 50% voter support for the Replication candidate in the 2020 US presidential election, and 0 otherwise.	MIT Election Lab
Masculinity (0-1)	Hofstede's 2010 cultural dimension index (feminine-masculinity) in a county (county of a job)	McLean et al. (2023)
> Median masculinity	Equals 1 if a county's masculinity score is above the median, and 0 otherwise	McLean et al. (2023)
Gender wage gap	Wage gap between male and female employees within the same occupation	BLS Current Population Survey
> Median gender wage gap	Equals 1 if an occupation's gender wage gap is above the median, and 0 otherwise	BLS Current Population Survey
%Fem new hire	Share of female new hires out of all new hires	LinkedIn
%Fem departure	Share of departed female employees out of all departed employees	LinkedIn
%Republican contribution	Share of a firm's political contributions to Republican candidates out of its total political contributions	Federal Election Commission
>50% Republican contribution	Equals 1 if more than 50% of a firm's political contributions go to Republican candidates	Federal Election Commission
%Male employee	Share of male employees out of all employees	LInkedIn
Top quartile %male employee	Equals 1 if firm-month falls within the top quartile of the %Male Employee distribution	LInkedIn
Pros	Number of achor words related to the eight female-friendly amenities in the pros section of the review text	Glassdoor
Cons	Number of anchor words related to the eight female-friendly amenities in the cons section of the review text	Glassdoor
Pros - Cons	Difference in the number of anchor words related to the eight female-friendly amenities between the pros and cons section of the review text	Glassdoor
Pros - Cons (weighted)	Difference in the number of anchor words related to the eight female-friendly amenities between the pros and cons section, weighted by length of a pros (cons) section	Glassdoor
Ratings: Overall	Overall employee ratings (1-5)	Glassdoor
Ratings: Compensation and benefits	Employee ratings for compensation and benefits (1-5)	Glassdoor
Ratings: Culture and values	Employee ratings for culture and values (1-5)	Glassdoor
Ratings: Work-life balance	Employee ratings for work-life balance (1-5)	Glassdoor
#Violation	Number of gender-related labor violations	Violation Tracker
log (1 + $Penalties)	Log of 1 plus dollar amount of penalties incurred by gender-related labor violations	Violation Tracker

The table lists variables used in the job ad, firm-month, employee review, employee rating, and firm-year level analysis.

B BERT-based Gendered Wording Measures

Constructing the gendered wording measure involves classifying a job ad into high versus low female-oriented categories, making it a supervised classification task. Given our focus on text classification, we leverage two pre-trained BERT models to classify job ads based on their gendered language characteristics. JobBERT (Zhang et al., 2022) is a BERT model pre-trained on 3.2 million sentences from job postings. DistilBERT (Sanh et al., 2019), is a distilled version of BERT, known for its balance of performance and computational efficiency, retaining 97% of BERT's language understanding capabilities while being 60% faster and 40% smaller (Sanh et al., 2019).[36] This appendix outlines our methodology for constructing and validating the gendered wording measure, ensuring transparency and replicability for future research.

B.1 Data collection and annotation

To fine-tune a pre-trained BERT model for a specific classification task, we need a dataset of job ads with ground truth labels. We start with a representative sample, comprising 4,500 job ads (extracted from our LinkUp sample) evenly distributed in male-dominated, female-dominated, and gender-neutral occupations based on the share of female employees in a given occupation. Each category contains 1,500 randomly sampled job ads to ensure diversity across industries and roles. Additionally, for each category, we require an equal number of ads (500) to receive a high, medium, or low share of gendered words as identified by Gaucher et al. (2011).

Annotations are performed by eight trained research assistants, balanced by gender (four females and four males) and minority representation (four Whites and four minority individuals). A custom data labeling platform, hosted at https://nlpuva.lighttag.io/, is developed to manage the annotation workflow (This platform was discontinued in August 2024; all annotated data and metadata were preserved).

Job ads are evaluated using a single question adapted from Gaucher et al. (2011): "How many women work in the position being advertised?" Responses are recorded on a Likert scale ranging from 0 (0% women) to 20 (100% women) in 5% increments. Each job ad is labeled by three annotators, and disagreements are resolved by majority vote, followed by a final review conducted by a fourth annotator. Weekly meetings are held to monitor annotators' productivity, ensure annotation quality, address issues, and refine guidelines as needed.

To simplify the process, the continuous Likert scale is converted into binary labels. Job

[36] BERT (Devlin et al., 2018), a pre-trained large language model, has demonstrated superior performance on various text classification benchmarks, making it an optimal choice for our task compared to GPT, which excels in generating coherent text sequences but is less suited for text classification.

ads with scores above 50% are categorized as female-tilted (1), while those with scores of 50% or below are categorized as male-tilted (0). After removing incomplete entries, the final dataset comprises 4,436 labeled job ads, with 2,984 categorized as more femininely-worded and 1,452 as more masculinely-worded.

B.2 Data preprocessing and model selection

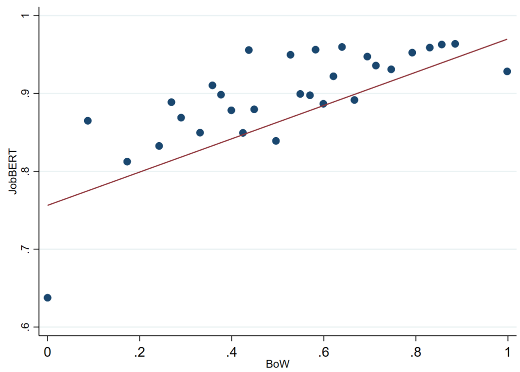
Before fine-tuning a BERT model, job ads are preprocessed for consistency. Tokenization and cleaning are performed using the Hugging Face tokenizer, ensuring compatibility with downstream tasks. The annotated dataset is split into training (60%), validation (20%), and testing (20%) subsets.

The training process involves fine-tuning JobBERT and DistilBERT to adapt them to the job ad context. Key hyperparameters include a batch size of 32 and 8 epochs. Optimization is performed using the AdamW optimizer with a linear learning rate schedule and 10% of training steps as warm-up steps, with an initial learning rate of 5e-5 and a weight decay of 0.01. Model performance is evaluated using standard classification metrics, including accuracy and AUC.

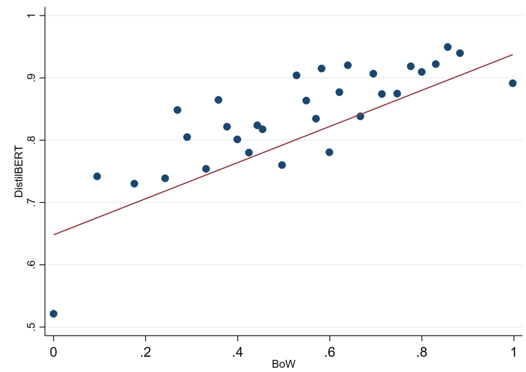
JobBERT achieves an accuracy of 88.84% and an AUC score of 94.00% while DistilBERT achieves an accuracy of 95.60% and an AUC score of 97.71%.

The fine-tuned JobBERT and DistilBERT models are then applied to our sample of job ads. At the job ad level, each model outputs a probability between 0 and 1; a probability closer to 1 suggests a more feminine ad, while a probability closer to 0 indicates a more masculine ad. Figure B1 illustrates the relationship between alternative measures from BERT and that from the bag-of-words (BoW) method used in the main specification, and we observe a strong alignment between the two approaches.

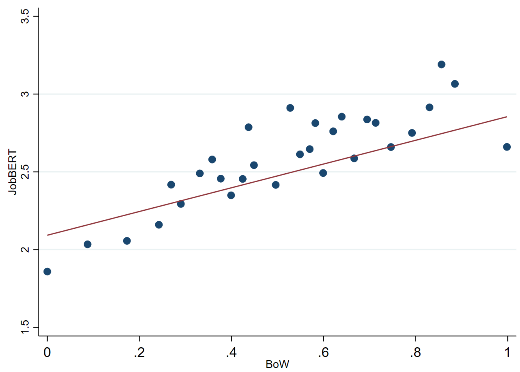
Figure B1: BERT (JobBERT and DistilBERT) vs. BoW



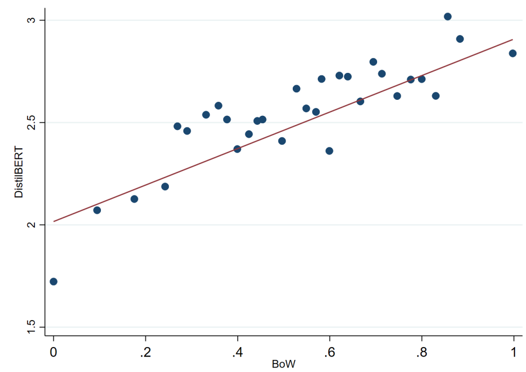
(a) JobBERT vs. BoW (continuous)



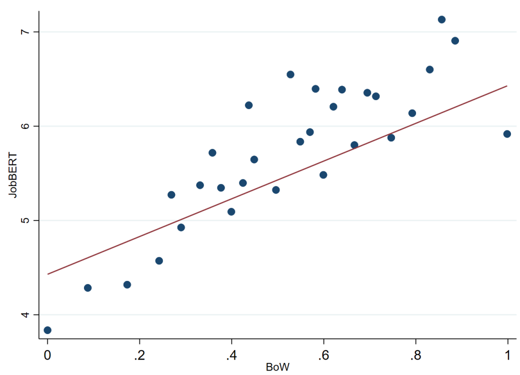
(b) DistilBERT vs. BoW (continuous)



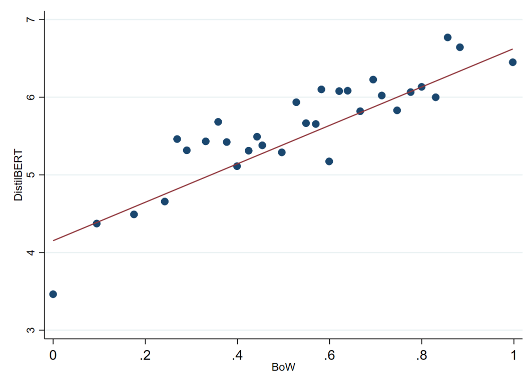
(c) JobBERT vs. BoW (4 Group)



(d) DistilBERT vs. BoW (4 Group)



(e) JobBERT vs. BoW (10 Group)



(f) DistilBERT vs. BoW (10 Group)

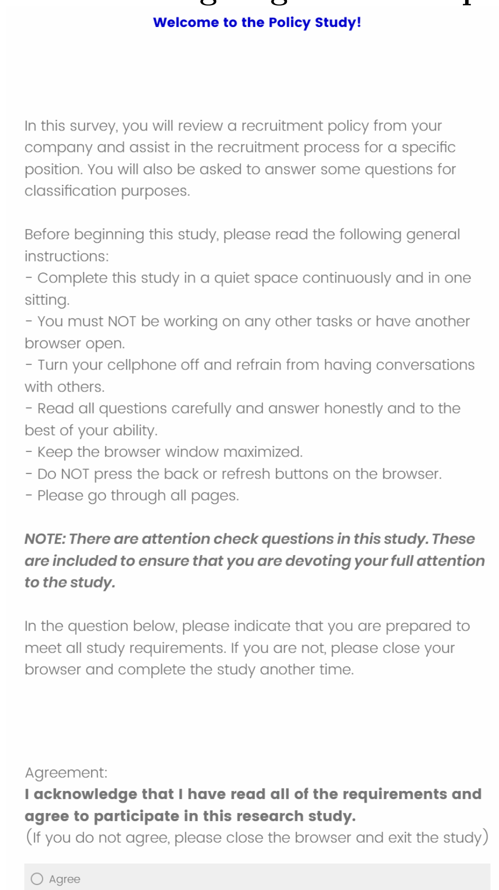
This figure illustrates the relationship between alternative measures from BERT and that from the bag-of-words (BoW) method used in the main specification. Figures B1a, B1c, and B1e use a fine-tuned JobBERT model. Figures B1b, B1d, and B1f use a fine-tuned DistilBERT model. Figures B1a and B1b use a continuous measure. Figures B1c and B1d (Figures B1e and B1f) categorize the continuous measure into 4 (10) groups.

C Experimental Design Details

C.1 Participants, manipulation, and survey questions

Our experiment employs a between-subjects design with two conditions: board gender quota policy (treatment) and no board gender quota policy (control). US-based participants are recruited via CloudResearch's Connect platform. We require participants to be over 18 years old and have at least one year of work experience. Our subject pool comprises 200 participants including 100 females, and 100 males. The median age is 46 (SD = 17.14), and age ranges from 19 to 81 years old.

Figure C1: Landing Page of the Experiment



Welcome to the Policy Study!

In this survey, you will review a recruitment policy from your company and assist in the recruitment process for a specific position. You will also be asked to answer some questions for classification purposes.

Before beginning this study, please read the following general instructions:
- Complete this study in a quiet space continuously and in one sitting.
- You must NOT be working on any other tasks or have another browser open.
- Turn your cellphone off and refrain from having conversations with others.
- Read all questions carefully and answer honestly and to the best of your ability.
- Keep the browser window maximized.
- Do NOT press the back or refresh buttons on the browser.
- Please go through all pages.

NOTE: There are attention check questions in this study. These are included to ensure that you are devoting your full attention to the study.

In the question below, please indicate that you are prepared to meet all study requirements. If you are not, please close your browser and complete the study another time.

Agreement:
I acknowledge that I have read all of the requirements and agree to participate in this research study.
(If you do not agree, please close the browser and exit the study)

○ Agree

This figure presents the landing page of the experiment when a subject starts the experiment.

Participants enter the experiment via an online survey. The experiment starts with a landing page shown in Figure C1.[37] Participants are asked to imagine themselves as employees of a fictitious company to assist the recruitment of a specific position. Each subject is then randomly assigned to one of the two policy conditions (see Table C1 for details). One condition (treatment) mentions and explains the board gender quota policy

[37] The full survey can be found at https://commercevirginia.co1.qualtrics.com/jfe/form/SV_4H3EnZwniPEpdZA.

while the other (control) does not. To ensure adequate attention paid to each condition, a delay is introduced so that participants spend at least five seconds reading the policy. To check participants' understanding of the policy, we include a manipulation check (using a three-item, seven-point Likert scale, see Table C2 and Figure C2a).

Table C1: Stimuli Used in the Experiment

Instruction
On the next page, you will read the recruitment policy from the company that you are working for. Then, you will be asked to share your opinions on it. While you are reading, please keep imagining yourself actually working for that company. Please note that there will be a short delay before the forward "≫" button appears. This is so you have sufficient time to read the scenario.
Treatment (with Board Gender Quota Policy)
The company you work for follows a **government-enforced gender quota policy that mandates a specific percentage of women must be included among its board members**. This policy applies strictly to leadership-level hiring decisions and introduces additional consideration of gender representation alongside qualifications and experience. Compliance with this policy is required and is closely monitored by external authorities. However, this policy **does not apply to non-leadership positions**, where hiring decisions are made independently of the quota.
Control (without Board Gender Quota Policy)
The company you work for follows a recruitment policy that **prioritizes selecting candidates based on their qualifications, professional experience**, and how well they align with the company's goals. This policy applies **equally to all positions, including both leadership and non-leadership positions**, and emphasizes the importance of individual merit in hiring decisions. The policy allows full flexibility for the company to select the best candidates for each role.

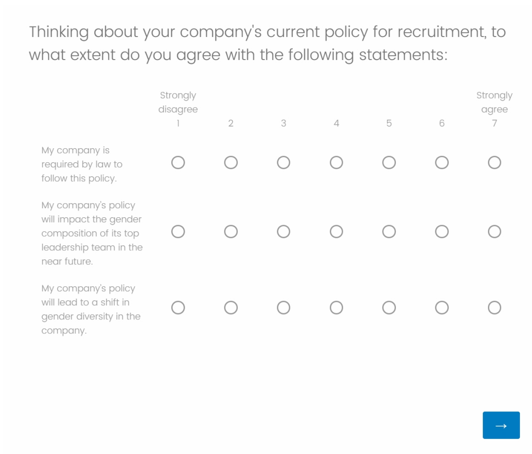
This table presents the instruction for participating in our experiment, followed by the two policy conditions (treatment vs. control) that are randomly assigned to participants.

After completing the manipulation check, participants are asked to share their opinions about the recruitment policy by answering a set of questions related to different mechanisms, listed in Table C2. We explore whether the board gender-quota policy, which may be perceived as a threat to or loss of freedom, triggers psychological reactance among participants. Reactance, in this context, refers to a defensive response when individuals feel their freedom is threatened or eliminated. The second mechanism is norm violation, i.e., participants view the quota policy (treatment) as a violation of gender norms in their workplace.

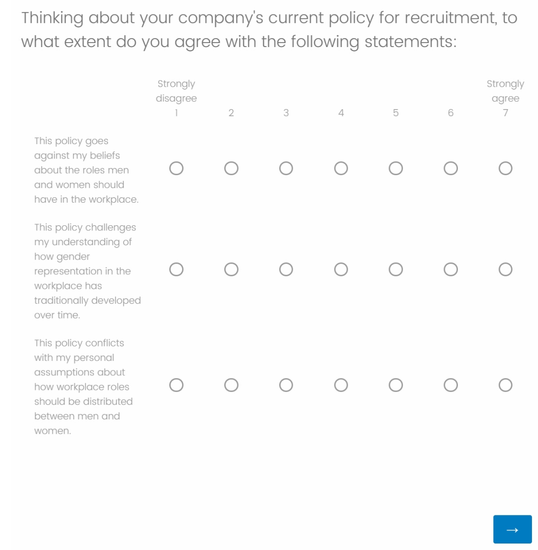
In addition to reactance and norm violation, we consider two other mechanisms that could shape participants' hiring preferences. Under moral licensing, complying with a quota policy in one context may lead individuals to feel justified in acting against gender equality practices in other contexts. We employ an adapted version of the scale in Merritt et al. (2010) to assess this mechanism (Figure C2c). Under job insecurity, the quota policy causes participants to worry about their own jobs or career prospects, particularly if they view the policy as favoring women at the expense of merit-based hiring or reducing opportunities for men. These concerns of job insecurity may lead participants to go against hiring female candidates in non-leadership roles as a way to safeguard their own positions or organizational stability. To evaluate this mechanism, we adopt the scale from Shoss (2017).

To eliminate order effects, we fully randomize the mechanisms in their presentation. We

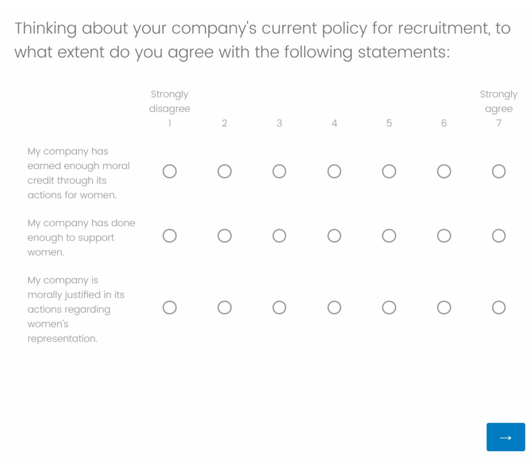
Figure C2: Screenshots of the Experiment



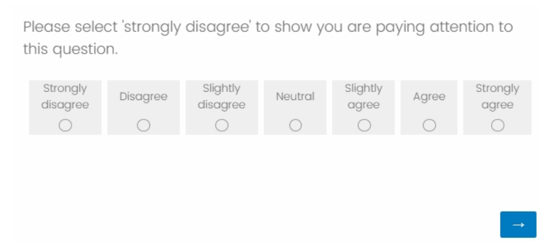
(a) Manipulation check



(b) Mechanism - Norm violation



(c) Mechanism - Moral licensing



(d) Attention check

This figure presents screenshots from various sections of the experiment. Figure C2a lists the manipulation check questions. Figure C2b lists the questions related to one of the mechanisms - norm violation. Figure C2c lists the questions related to one of the mechanisms - moral licensing. Figure C2d lists our attention check question.

also include a check requiring participants to select "Strongly Disagree" to ensure their attention (Figure C2d). We end the experiment by asking participants demographic questions, including age, gender, ethnicity, language, employment status, years of work experience, and political view (from liberal to conservative).

C.2 Validation of randomization and manipulation

Table C3 confirms our randomization design among participants. Participants in the treatment group are comparable to those in the control group across key characteristics, with no

Table C2: Measures Used in the Experiment

Measures	Questions and Scales
Manipulation check	Thinking about your company's current policy for recruitment, to what extent do you agree with the following statements: My company is required by law to follow this policy. My company's policy will impact the composition of its top leadership team in the near future. My company's policy will lead to a shift in gender diversity in the company. 1 = Strongly Disagree, 7 = Strongly Agree; Cronbach's $\alpha = 0.80$
Mechanism: Reactance	Thinking about your company's current policy for recruitment, answer the following questions: To what extent do you perceive the recruitment policy as a restriction of your freedom to make hiring decisions? Are you frustrated about the recruitment policy? How much does the recruitment policy annoy you? To what extent are you offended/disturbed by the recruitment policy? 1 = Strongly Disagree, 7 = Strongly Agree; Cronbach's $\alpha = 0.96$
Mechanism: Norm violation	Thinking about your company's current policy for recruitment, to what extent do you agree with the following statements: This policy goes against my beliefs about the roles men and women should have in the workplace. This policy challenges my understanding of how gender representation in the workplace has traditionally developed over time. This policy conflicts with my personal assumptions about how workplace roles should be distributed between men and women. 1 = Strongly Disagree, 7 = Strongly Agree; Cronbach's $\alpha = 0.86$
Mechanism: Moral licensing	Thinking about your company's current policy for recruitment, to what extent do you agree with the following statements: My company has earned enough moral credit through its actions for women. My company has done enough to support women. My company is morally justified in its actions regarding women's representation. 1 = Strongly Disagree, 7 = Strongly Agree; Adapted from: Merritt et al. (2010); Cronbach's $\alpha = 0.85$
Mechanism: Job insecurity	Thinking about your company's current policy for recruitment, to what extent do you agree with the following statements: Chances are, I will soon lose my job. I feel insecure about my future job. I think I might lose my job in the near future. 1 = Strongly disagree, 7 = Strongly agree; Adapted From: Shoss (2017); Cronbach's $\alpha = 0.95$

This table lists the multiple-item, seven-point Likert scale used in the manipulation check and mechanism tests. The mechanisms are randomized in their presentation to eliminate order effects.

significant differences in age, race, work experience, or political view. The treatment and control groups also spend a similar amount of time completing the survey.

Table C3: Covariate Balance between Treatment and Control Participants

	Control		Treatment			
	Mean	SD	Mean	SD	Diff	t-stat
Age	47.258	17.28	44.257	16.995	3.000	-1.231
Male (0/1)	0.505	0.503	0.495	0.502	0.010	-0.141
Race: White (0/1)	0.784	0.414	0.772	0.421	0.011	-0.189
Work experience (>5 years)	0.876	0.331	0.802	0.4	0.074	-1.426
Political view (1-7)	4.031	2.018	3.96	2.245	0.071	-0.233
Time to complete survey (in seconds)	230.258	130.92	244.663	145.864	-14.406	-0.732

This table shows the balance of covariates between participants in the control group and those in the treatment group.

The result in column (1) of Table 12 Panel A reveals a significant manipulation effect. Participants under the treatment condition view the policy as significantly more binding and impactful on board gender composition and diversity compared to those under the control condition. This manipulation effect is around 50% of the sample mean.

D External Validity: Aggregate Female Employment in Europe

In this section, we exploit the staggered adoptions of board gender quotas across European countries at different points in time to study the effect of board gender quotas on female labor market outcome in very different institutional settings.

Table D1: Board Gender Quotas in Europe

Country	Passage Date	Quota	Quota Law	Hard/Soft Quota?
NO	12/19/2003	40%	Norwegian Company Act	Hard
ES	3/12/2007	40%	Spanish Equality Act	Soft
IS	3/4/2010	40%	Icelandic Company Act	Soft
FR	1/13/2011	40%	French Cope Zimmerman Company Act	Medium hard
NL	6/6/2011	30%	Dutch Civil Code	Soft
IT	6/28/2011	33%	Italian Golfo Mosca Company Act	Hard
BE	6/30/2011	33%	Belgian Company Act	Hard
DE	3/15/2015	30%	German Company Act	Hard
AT	7/26/2017	30%	Austrian Company Act	Medium hard
PT	8/1/2017	33.30%	Portuguese Law	Hard

The table lists board gender quotas in Europe introduced over the period 2003-2017, including each quota's passage date and the relevant quota law. Whether a quota is hard, medium hard, or soft is classified by Mensi-Klarbach and Seierstad (2020).

Table D2: List of Variables, Definitions, and Data Sources

Variable	**Definition**	**Source**
%Fem emp	Share of female employment out of total employment at the country-industry (NACE)-year level	Eurostat LSF
Quota	Equals 1 if a country has implemented a board gender quota policy, and 0 otherwise	Legislative files
Hard quota	Equals 1 if a country has implemented a hard or medium hard board gender quota policy, and 0 otherwise	Legislative files
Soft quota	Equals 1 if a country has implemented a soft board gender quota policy, and 0 otherwise	Legislative files
Total emp (1,000)	Total employment at the country-industry (NACE)-year level	Eurostat LSF
%Fem population	Share of female population	WorldBank
Log (population)	Log of total population	WorldBank
Population growth	Year-on-year population growth	WorldBank
Log (GDP per capita)	Log of GDP per capita	WorldBank
GDP per capita growth	Year-on-year GDP per capita	WorldBank

The table lists variables used in the European country-industry-level analysis.

We obtain information on board gender quotas, including passage dates, enforcement mechanisms, and scope, from legislative offices. Based on the strength of the enforcement, the wording of the relevant law, and the institutional context in which the law was introduced, Mensi-Klarbach and Seierstad (2020) classify these board gender quotas into soft (e.g., Spain and Iceland) versus non-soft (i.e., hard or medium hard, e.g., Norway and France). Table D1 in the Online Appendix lists quotas in Europe introduced over the period 2003–2017. We obtain data on the share of female employment (men and women, age 15-64) from the Labor Force Survey (LFS) of Eurostat in each country-industry-year from 2008 to 2020. Table D2 in the Online Appendix provides the list of variables, their definitions, and data sources. Table D3 presents summary statistics.

Table D3: Summary Statistics: European Country-Industry-Year Level

Variable	N	Mean	Std. Dev.	Median	p1	p99
%Fem emp	29,734	0.39	0.23	0.36	0.03	0.91
%Fem dir	29,734	0.19	0.10	0.17	0.03	0.45
Qutoa	29,734	0.24	0.42	0.00	0.00	1.00
Hard quota	29,734	0.15	0.36	0.00	0.00	1.00
Soft quota	29,734	0.08	0.28	0.00	0.00	1.00
Total emp (1,000)	29,734	105.00	305.00	22.40	0.80	1409.00
%Fem population	29,734	0.51	0.01	0.51	0.50	0.54
Log (population)	29,734	16.00	1.44	16.00	12.70	18.20
Population growth	29,734	0.32	0.79	0.30	-1.67	2.40
Log (GDP per capita)	29,734	10.20	0.73	10.20	8.56	11.60
GDP per capita growth	29,734	0.98	3.70	1.38	-8.60	8.37

The table summarizes variables at the European country-industry-year level.

Panel A of Table D4 presents the results based on a staggered DID specification. We consistently find a negative and significant effect of a country's introduction of a board gender quota on its share of female employment in various specifications from columns (1) to (6).[38] In terms of economic significance, using the specification with the most comprehensive set of fixed effects and controls in column (6), we find that the adoption of board gender quotas leads to a 0.30 percentage-point drop in the share of female employment at the country-industry-year level. Given that the sample mean female share is 39%, this drop represents a 0.8% drop. At the aggregate level, the share of women in the labor force increases by less than five percentage points over the past 30 years, equivalent to an increase of about 0.15 percentage points per year. In other words, board gender quotas in Europe set back the progress in female labor market participation by two years. Figure D1 plots the dynamic effects of such quotas. We observe a clear post-quota downward trend and in the pre-quota years the estimated coefficients fluctuate around zero and are always insignificant.

Do all quotas have the same effect on female labor market participation? We divide all ten quotas into two regimes depending on whether a quota is enforced by the state as opposed to a voluntary target. Panel B of Table D4 presents the regression results. We show that the above effect is stronger when we focus on hard/medium hard quotas with strong and clear enforcement mechanisms. In contrast, the effect is close to zero when we focus on soft quotas. This contrast is consistent with backlash: strict quotas tend to provoke pushback compared to voluntary codes, and this can occur across countries with varying cultural and institutional backgrounds.

[38] Our findings remain when we cluster standard errors at the industry level or assign equal weight to each country-industry-year observation, as shown in columns (1) and (2) of Table D5. We also follow methodologies in Borusyak et al. (2021) and Callaway and Sant'Anna (2021) to re-estimate the average treatment effect in light of the recent criticism of two-way fixed effects regressions in staggered settings. We find similar results, see columns (3) and (4) of Table D5.

Table D4: External Validity – Effect of Board Gender Quotas in Europe on Female Labor Market Participation

Panel A: All Quotas

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Share of female employment out of total employment (%Fem Emp × 100)					
Quota	-0.289***	-0.252**	-0.228**	-0.358***	-0.294**	-0.299***
	[0.105]	[0.106]	[0.092]	[0.130]	[0.127]	[0.098]
%Fem population				0.268	0.276	0.156
				[0.268]	[0.238]	[0.139]
Log (population)				7.311**	6.052**	4.244*
				[2.991]	[2.834]	[2.160]
Population growth				0.066	0.025	0.088*
				[0.050]	[0.081]	[0.044]
Log (GDP per capita)				0.959	1.543*	0.991*
				[0.807]	[0.872]	[0.548]
GDP per capita growth				-0.002	0.004	-0.007
				[0.012]	[0.019]	[0.011]
Industry FE	YES	-	-	YES	-	-
Country FE	YES	YES	-	YES	YES	-
Region × Year FE	YES	YES	YES	YES	YES	YES
Industry × Year FE	NO	YES	YES	NO	YES	YES
Country × Industry FE	NO	NO	YES	NO	NO	YES
Obs	29,734	29,734	29,688	29,734	29,734	29,688
Adj R2	0.918	0.917	0.991	0.918	0.917	0.991

Panel B: Hard vs. Soft Quotas

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Share of female employment out of total employment (%Fem Emp × 100)					
Hard quota	-0.311**	-0.265**	-0.238**	-0.385***	-0.311**	-0.313***
	[0.117]	[0.110]	[0.095]	[0.139]	[0.130]	[0.101]
Soft quota	0.194	0.009	-0.013	0.093	-0.012	-0.06
	[0.119]	[0.277]	[0.307]	[0.163]	[0.286]	[0.300]
%Fem population				0.22	0.247	0.132
				[0.264]	[0.239]	[0.143]
Log (population)				7.074**	5.929**	4.132*
				[2.949]	[2.847]	[2.171]
Population growth				0.067	0.025	0.089*
				[0.048]	[0.081]	[0.044]
Log (GDP per capita)				1.046	1.605*	1.038*
				[0.834]	[0.887]	[0.555]
GDP per capita growth				-0.003	0.004	-0.007
				[0.012]	[0.019]	[0.011]
Industry FE	YES	-	-	YES	-	-
Country FE	YES	YES	-	YES	YES	-
Region × Year FE	YES	YES	YES	YES	YES	YES
Industry × Year FE	NO	YES	YES	NO	YES	YES
Country × Industry FE	NO	NO	YES	NO	NO	YES
Obs	29,734	29,734	29,688	29,734	29,734	29,688
Adj R2	0.918	0.917	0.991	0.918	0.917	0.991

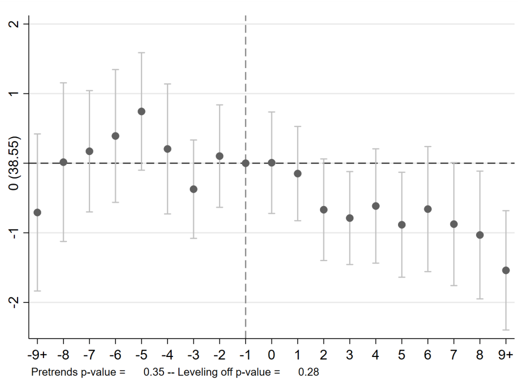
This table shows the effect of board gender quotas in Europe on the share of female employment. The outcome variable is the share of female employment out of all employment, scaled by 100 for easy interpretation of the regression coefficients. The unit of observation is a country-industry-year. The sample includes 89 two-digit NACE industries in 32 European countries from 2008 to 2020. Panel A includes board gender quotas adopted by the ten European countries over the period 2003-2017. Panel B compares board gender quotas with stronger enforcement and more precise policy wording (i.e., we combine hard and medium hard quotas into "hard" quotas) with the ones with weaker enforcement and less precise policy wording ("soft" quotas). All regressions are weighted by total employment in a country-industry-year. Robust standard errors clustered at the country level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table D5: Effect of Board Gender Quotas in Europe on Female Labor Market Participation – Robustness

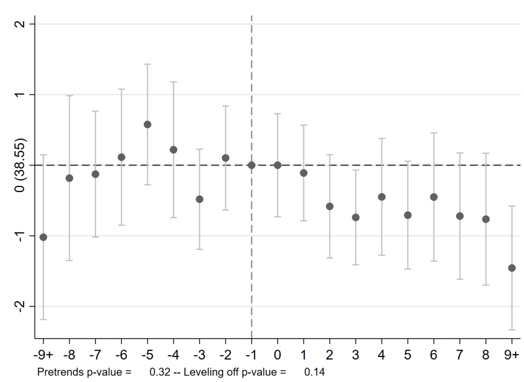
	(1)	(2)	(3)	(4)
Dep. Var.	Share of female employment out of total employment (%Fem Emp)			
Quota	-0.299**	-0.399**	-0.469**	-0.670**
	[0.119]	[0.185]	[0.225]	[0.311]
Alternative Estimator	Cluster by Industry	Equal-weighted	Borusyak et al. (2021)	CSDID
Controls	YES	YES	YES	YES
Region × Year FE	YES	YES	YES	NO
Industry × Year FE	YES	YES	YES	NO
Country × Industry FE	YES	YES	YES	YES
Year FE	-	-	-	YES
Obs	29,688	29,688	27,545	26,909
Adj R2	0.991	0.943	0.858	0.858

This table shows the effect of board gender quotas in Europe on the share of female employment using different specifications or alternative DID estimators. The outcome variable is the share of female employment, scaled by 100 for easy interpretation of the regression coefficients. The unit of observation is a country-industry-year. The sample includes 89 two-digit NACE industries in 32 European countries from 2008 to 2020. Column (1) uses alternative clustering at the industry level instead of the country level. Column (2) puts no weight to a country-industry-year in the regression analysis. Column (3) uses the imputation approach in Borusyak et al. (2021). Column (4) uses the estimator in Callaway and Sant'Anna (2021). Robust standard errors clustered at the country level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Figure D1: Effect of Board Gender Quotas in Europe on Female Labor Market Participation – Dynamics



(a) Without Controls



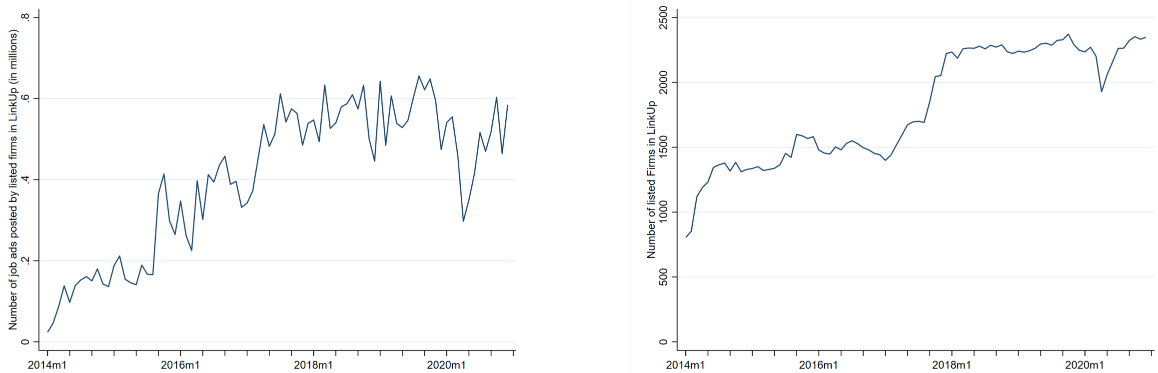
(b) With Controls

This figure plots the dynamic effect of board gender quotas in Europe on the share of female employment. Year minus one is the year before the announcement of board gender quotas and is the omitted category. The outcome variable is scaled by 100 for easy interpretation of the regression coefficients. Figure D1a (Figure D1b) presents the estimated coefficients without (with) control variables in the regression. p-values for testing the absence of pre-event effects and p-values for testing the null that dynamics have leveled off are included. Pointwise 95 percent confidence intervals are included.

E Coverage of LinkUp

Figure E1a shows that LinkUp's coverage of job ads grows over time. The number of job ads increases from fewer than 0.2 million per month in 2014 to around 0.6 million per month toward the later half of our sample period. Figure E1b shows LinkUp's monthly coverage of public firms. Their coverage grows from fewer than 1,000 firms at the beginning of 2014 to close to 2,500 firms at the end of 2020.

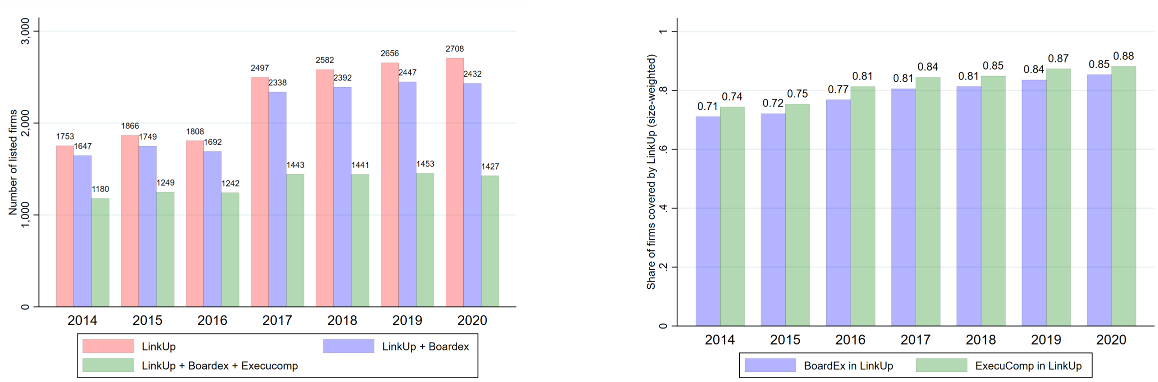
Figure E1: LinkUp Coverage from 2014 to 2020



(a) Number of Job Ads in LinkUp **(b) Number of Listed Firms in LinkUp**

This figure shows the coverage of LinkUp over the period 2014-2020. Figure E1a plots the number of job ads covered by LinkUp. Figure E1b plots the number of public firms covered by LinkUp.

Figure E2: LinkUp vs. BoardEx/ExecuComp



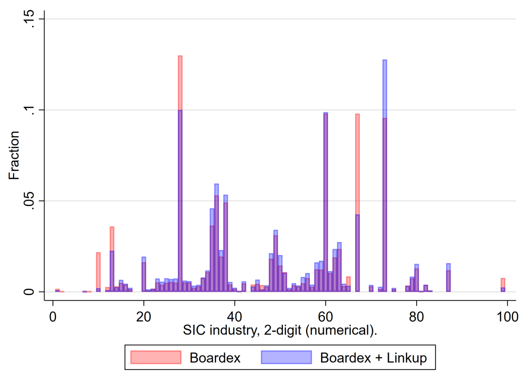
(a) Number of Public Firms in LinkUp and LinkUp ∩ BoardEx/ExecuComp **(b) Share of BoardEx/ExecuComp Firms Covered by LinkUp**

This figure shows the coverage of LinkUp relative to BoardEx and ExecuComp. Figure E2a shows the number of public firms in LinkUp relative to the number of firms covered by both LinkUp and BoardEx, or the number of firms covered by LinkUp, BoardEx, and ExecuComp. Figure E2b shows the assets-weighted share of BoardEx (ExecuComp) firms covered by LinkUp.

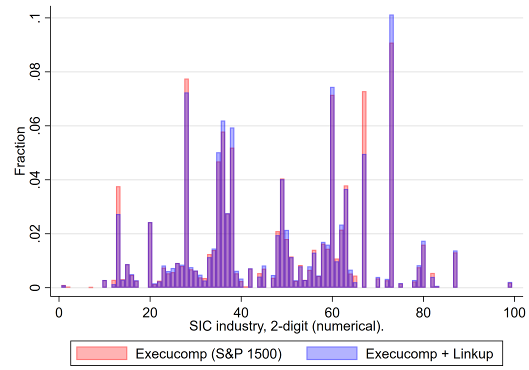
Figure E2a shows that the merged sample of LinkUp and BoardEx firms (blue bar) spans between 1,647 to 2,447 Compustat firms over the period 2014-2020, capturing between 37%

to 64% of BoardEx firms. While this coverage ratio does not seem particularly high, LinkUp tends to cover large companies that are also big employers. The blue bar in Figure E2b shows the assets-weighted share of public firms in the merged sample of LinkUp and BoardEx. This ratio varies between 71% to 85%.

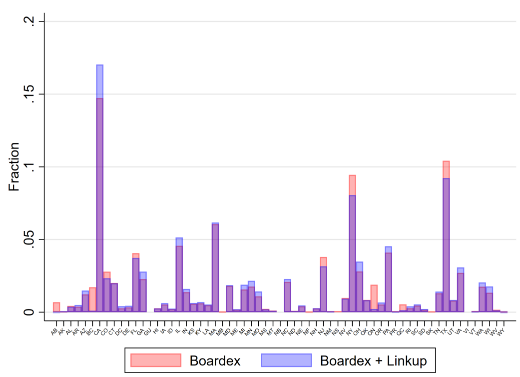
Figure E3: Sample Representativeness: LinkUp ∩ BoardEx/ExecuComp – Industry and Geography



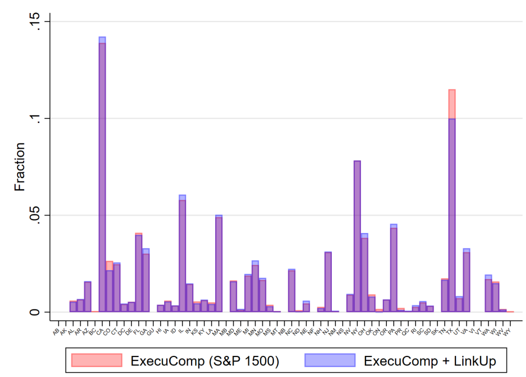
(a) Industry, BoardEx vs. BoardEx + LinkUp



(b) Industry, ExecuComp vs. ExecuComp + BoardEx + LinkUp



(c) HQ State, BoardEx vs. BoardEx + LinkUp



(d) HQ State, ExecuComp vs. ExecuComp + BoardEx + LinkUp

This figure shows the industry (two-digit SIC) and geographical coverage of the merged dataset of LinkUp and BoardEx (and ExecuComp). Figure E3a (Figure E3c) compares the fraction of public firms in each two-digit SIC industry (in each headquarters state) in the BoardEx dataset with the corresponding fraction in the merged dataset of LinkUp and BoardEx. Figure E3b (Figure E3d) compares the fraction of public firms in each two-digit SIC industry (in each headquarters state) in the ExecuComp dataset with the corresponding fraction in the merged dataset of LinkUp, BoardEx, and ExecuComp.

Next, when we focus on large companies by further merging the data with ExecuComp (ExecuComp covers public firms that are the constituents of S&P 1500), the resulting sample has between 1,180 to 1,453 S&P 1500 firms over the period 2014–2020, as shown by the green bar in Figure E2a. This represents a majority of S&P 1500 firms, both in terms of the number of firms and the assets-weighted share. The green bar in Figure E2b shows that between

74% to 88% of the aggregate assets value of S&P 1500 firms in ExecuComp are covered by LinkUp.

Our sample using LinkUp data captures the most significant firms in the economy. We next examine if there is any industry or geographic tilt. We compare the distribution of firms in our merged sample across industries and states with the distribution of firms in BoardEx or ExecuComp. Figure E3 suggests that the sample merged with LinkUp has very similar industry (Figures E3a and E3b) or geographic (Figures E3c and E3d) coverage to that of BoardEx and ExecuComp.

F Datasets and Measures Used in Post-hiring Analysis

Employee reviews of female-friendly amenities. Sockin (2022) identifies 50 amenities using textual analysis of Glassdoor employee reviews. For our purposes, we focus on eight amenities (5-Paid time off, 10-Work-life balance, 11-Hours, 12-Work schedule, 16-Teleworking, 19-Respect/abuse, 21-Support, and 31-Diversity/inclusion) that are important for female employees in the workplace. These "female-friendly" amenities not only reflect firm-level policies but also are shaped by supervisors and coworkers who interact with female employees. If work environment becomes more hostile for female employees, they may write more negative reviews about (lack of) workplace accommodations. Table F1 in the Online Appendix lists the eight amenities and their anchor words that we use to count different amenities in the pros and cons sections of Glassdoor reviews.

Table F1: Descriptions of Female-Friendly Amenities

#	Category	Amenity	Anchor words
5	Fringe benefits	Paid time off	vacation, pto, sick days, leave, pay time off
10	Working conditions	Work-life balance	work life balance, work life
11	Working conditions	Hours	hours, full time, part time
12	Working conditions	Work schedule	hours, shift, schedule, flex time
16	Working conditions	Teleworking	telecommute, telework, work home, home office, remote
19	Working conditions	Respect/abuse	respect, dignity, abuse, harass, hostile
21	Working conditions	Support	help, support, supportive, encourage
31	Working conditions	Diversity/inclusion	diversity, ethnic, multicultural, inclusive, lgbtq, inclusion, equality, diverse

This table lists the anchor words of the eight female-friendly amenities from Sockin (2022).

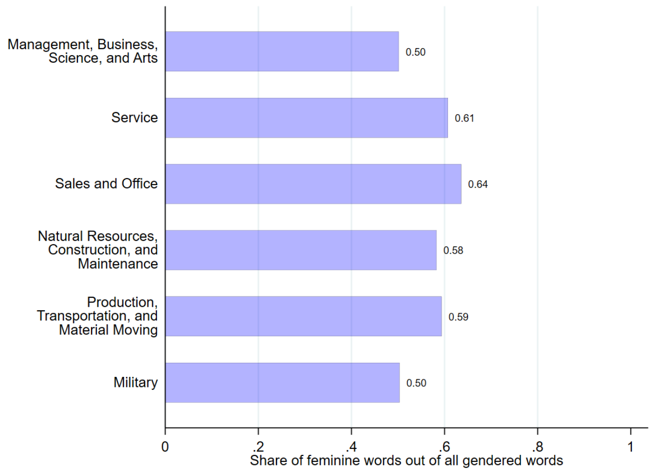
Employee ratings. Glassdoor also collects employee ratings on a scale of 1-5. We focus on a firm's overall and specific ratings for compensation and benefits, culture and values, and work-life balance—benefits that likely matter (more) for female employees in the workplace.

Gender-related labor violations. We start with all employment-related violation cases from Violation Tracker and employ two approaches to identifying gender-related labor violations. First, a violation is considered gender-related if it mentions any of the following seed words in the "description" or "secondary offense" fields: female, gender, hostile, hours, leave, sex, sexual, support, and vacation. Second, we expand this initial word list by adding words from the above two fields with high associations in terms of pmi-freq scores (Jin et al., 2021) with the seed words: applicants, women, job, positions, discrimination, environment, medical, sick, and harassment. We then aggregate case-level information to firm-year level and calculate the number of gender-related violations and the dollar amount of penalties paid for those violations.[39]

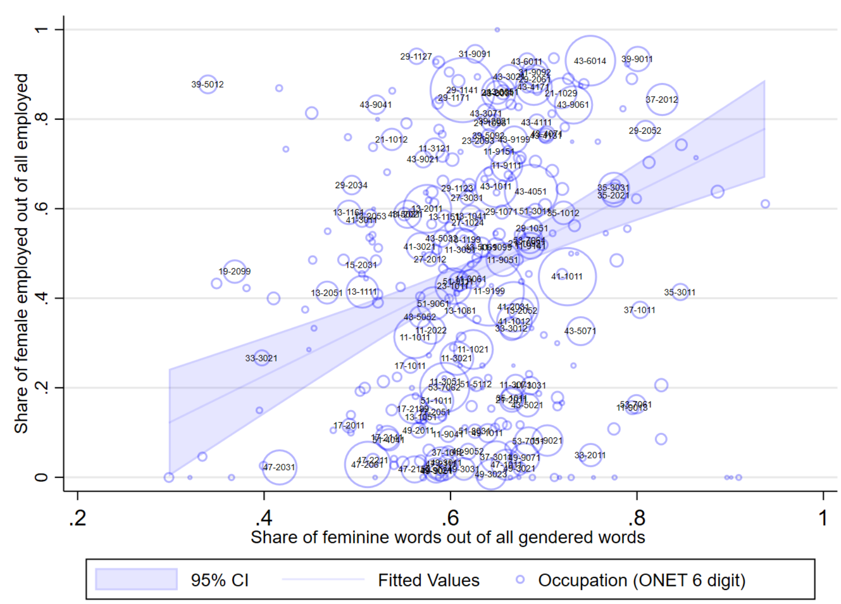
[39] These cases are mostly brought by the Equal Employment Opportunity Commission (EEOC) or the Office of Federal Contract Compliance Programs (OFCCP).

G Additional Figures and Tables

Figure G1: Gendered Language Usage in Different Occupations and Employment Outcomes



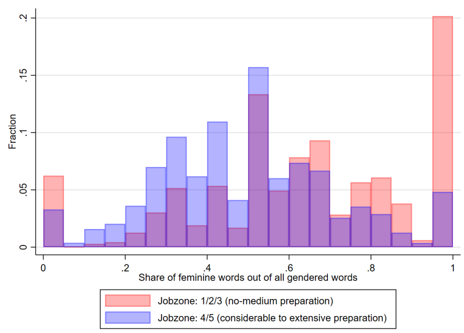
(a) Two-digit ONET-SOC



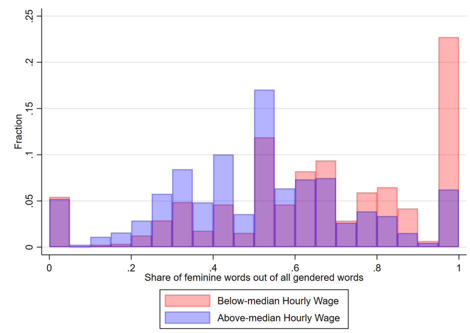
(b) Six-digit ONET-SOC

This figure shows gendered language usage in different occupations and employment outcomes. Figure G1a shows gendered language usage in six occupations, based on the two-digit ONET-SOC code (11-29, 31-39, 41-43, 45-49, 51-53, 55). The definitions are from Table 6 of the Standard Occupational Classification (SOC) User Guide. See https://www.aila.org/File/Related/17122005a.pdf. Figure G1b plots the share of females out of all employed persons against gendered language usage across all six-digit ONET-SOC codes (weighted by total number of employed persons). Information on employed persons by occupation and sex is from the 2020 Labor Force Statistics of the BLS Current Population Survey. See https://www.bls.gov/cps/aa2020/cpsaat11.htm.

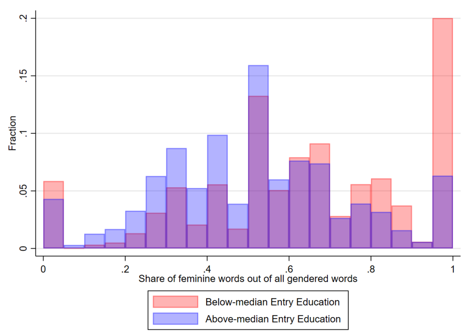
Figure G2: Gendered Language Usage in Different Occupations – Occupation-Level Characteristics



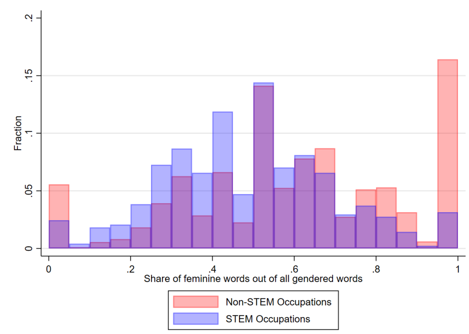
(a) Job Zone



(b) Mean Wage



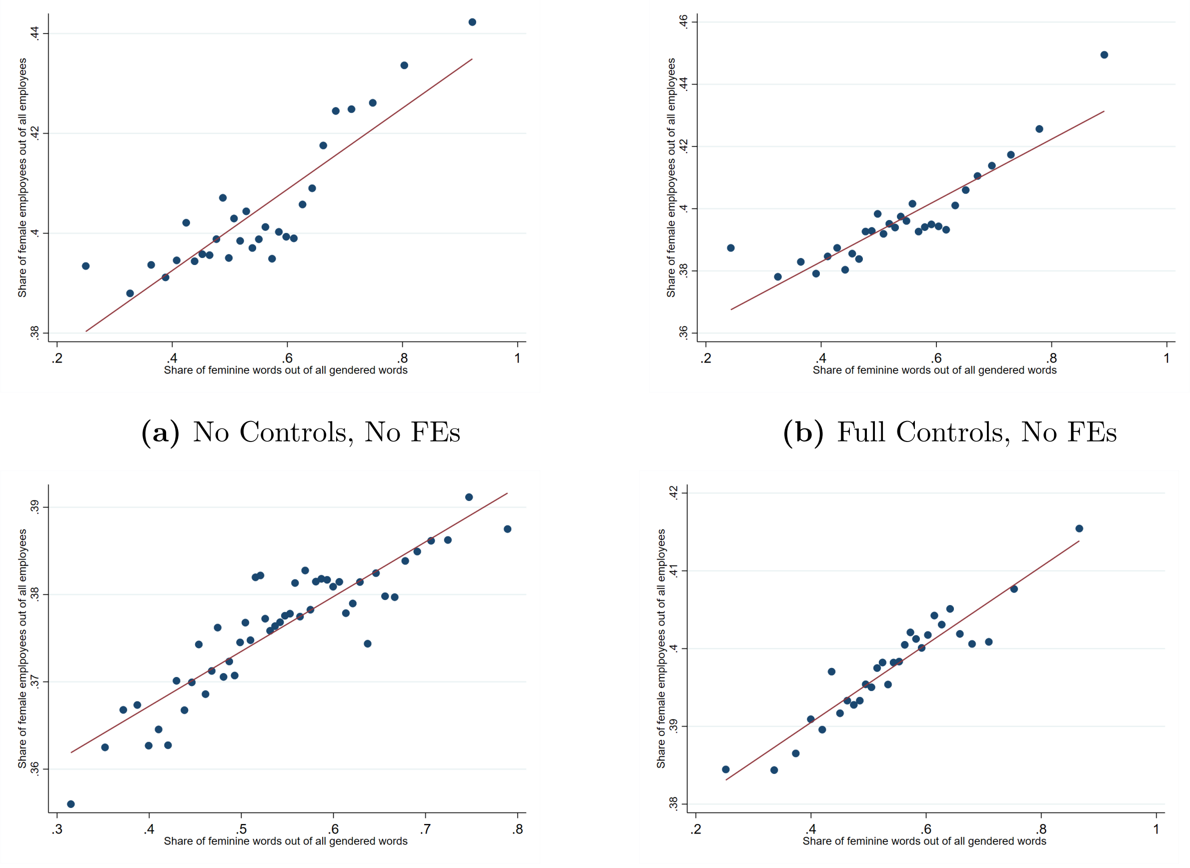
(c) Entry Education



(d) STEM

This figure shows how gendered language usage in job ads varies with occupation-level characteristics. Figure G2a splits the sample by the value of job zone, which is a group indicator from 1 to 5 for occupations that are similar in their education, experience, and training requirements. The higher the value of a job zone, the more complex skills a job requires. Information on job zone is from the ONET Database. Figure G2b splits the sample by the median of the average hourly wage for each occupation, using the 2020 national Occupational Employment and Wage Statistics (OEWS) datasets from BLS. Figure G2c splits the sample by the median of the entry education requirement for each occupation, using OEWS' typical entry-level educational requirement datasets. Figure G2d splits the sample into STEM and non-STEM occupations, using OEWS' STEM datasets.

Figure G3: Firm-level Gendered Language Usage and Share of Female Employees



(a) No Controls, No FEs

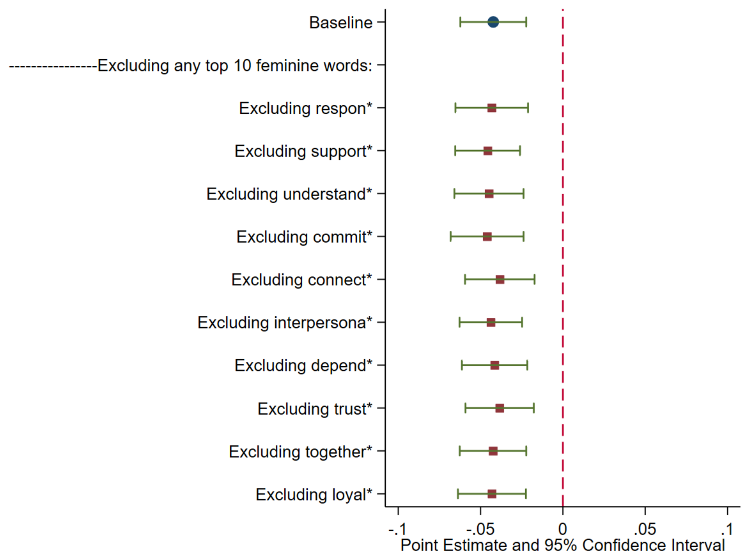
(b) Full Controls, No FEs

(c) No Controls, Time/Industry/State FEs

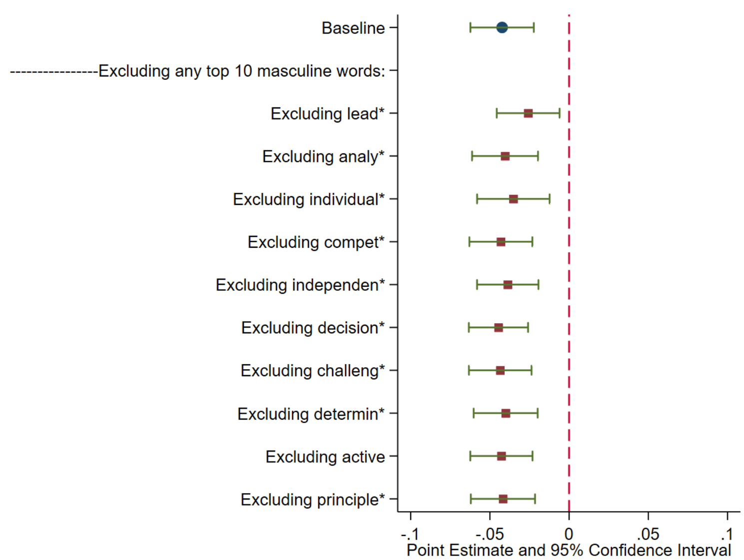
(d) Full Controls, Time/Industry/State FEs

This figure shows how gendered language usage in job ads varies with the share of female employees using a firm-month panel over the period 2014-2020. Share of female employees is calculated by aggregating individual profiles from LinkedIn. Figure G3a associates share of feminine words out of all gendered words in job ads with share of female employees out of all employees without any control variables or fixed effects. Figure G3b adds a full set of control variables including firm characteristics and average length of job ads. Figure G3c adds year-month, industry, and state (headquarter state of a firm) fixed effects. Figure G3d adds both control variables and fixed effects.

Figure G4: Robustness - Excluding Top Feminine/Masculine Words



(a) Feminine Words



(b) Masculine Words

This figure shows the robustness of the baseline job ad level regression results in column (6) of Table 2 excluding top feminine (Figure G4a) or masculine (Figure G4b) words when computing gendered language usage in each job ad. The asterisk denotes the acceptance of all letters, hyphens, or numbers following its appearance. The top row in each figure indicates the baseline estimate.

Table G1: Dictionary of Masculine/Feminine Words in Gaucher et al. (2011)

Masculine	Feminine
active	affectinate
adventurous	child*
aggress*	cheer*
ambitio*	commit*
analy*	communal
assert*	compassion*
athlet*	connect*
autonom*	considerate
boast*	cooperat*
challeng*	depend*
compet*	emotiona*
confident	empath*
courag*	feminine
decide	flatterable
decisive	gentle
decision*	honest
determin*	interpersonal
dominant	interdependen*
domina*	interpersona*
force*	kind
greedy	kinship
headstrong	loyal*
hierarch*	modesty
hostil*	nag
impulsive	nurtur*
independen*	pleasant*
individual*	polite
intellect*	quiet*
lead*	respon*
logic	sensitiv*
masculine	submissive
objective	support*
opinion	sympath*
outspoken	tender*
persist	together*
principle*	trust*
reckless	understand*
stubborn	warm*
superior	whin*
self-confiden*	yield*
self-sufficien*	
self-relian*	

The table lists the masculine and feminine words identified in Gaucher et al. (2011). The asterisk denotes the acceptance of all letters, hyphens, or numbers following its appearance.

Table G2: Additional Summary Statistics

Panel A: Employee Turnover from LinkedIn

Variable	N	Mean	Std. Dev.	Median	p1	p99
#Employees	87,495	3445.00	9,337	842	1	48,392
#New hire	87,495	60.40	200	13	0	840
#Departure	87,495	47.50	144	10	0	736
%Fem new hire	77,138	0.41	0.24	0.39	0.00	1.00
%Fem departure	74,356	0.40	0.25	0.38	0.00	1.00

Panel B: Employee Reviews of Female-Friendly Amenities

Variable (word frequency)	N	Mean	Std. Dev.	Median	p1	p99
Pros	54,814	0.49	0.83	0.00	0.00	4.00
Cons	54,814	0.61	1.08	0.00	0.00	5.00
Pros - Cons	54,814	-0.06	0.68	0.00	-2.50	1.50
Pros - Cons (weighted)	54,814	-0.14	0.82	0.00	-3.00	2.00

Panel C: Employee Ratings of Female-Friendly Benefits

Variable	N	Mean	Std. Dev.	Median	p1	p99
Ratings: Overall	180,247	3.15	1.30	3.00	1.00	5.00
Ratings: Compensation and benefits	172,010	3.06	1.49	3.00	0.00	5.00
Ratings: Culture and values	171,542	2.81	1.61	3.00	0.00	5.00
Ratings: Work-life balance	171,960	2.76	1.53	3.00	0.00	5.00

Panel D: Gender-Related Labor Violations and Penalties

Variable (word frequency)	List of words for violations	N	Mean	Std. Dev.	Median	p1	p99
#Violations	Seed word only	11,182	0.005	0.074	0	0	0
$Penalties	Seed word only	11,182	39,941	2,529,465	0	0	0
#Violations	Expanded word list	11,182	0.011	0.115	0	0	1
$Penalties	Expanded word list	11,182	1,520,086	130,000,000	0	0	7,939

The table presents additional summary statistics. Panel A presents summary statistics for variables constructed from the LinkedIn sample. Panels B and C present summary statistics for variables constructed from employee reviews and ratings of female-friendly amenities and benefits from Glassdoor. Panel D presents summary statistics for gender-related labor violations and penalties from Violation Tracker. Seed word only refer to violations that mention any of the following words in the "description" or "secondary offense" field: female, gender, hostile, hours, leave, sex, sexual, support, and vacation. Expanded word list refer to violations that mention any of the following words in the "description" or "secondary offense" field: any word from the seed word list, applicants, women, job, positions, discrimination, environment, medical, sick, and harassment. The newly added words are included as they have high associations with the seed words in terms of pmi-freq scores (Jin et al., 2021).

Table G3: Effect of California Statute on Female Directors

	(1)	(2)	(3)	(4)
Dep. Var.	#Fem dir	%Fem dir	#Fem dir	%Fem dir
Flag_CA × Post_2018m9	0.150***	0.020***		
	[0.028]	[0.003]		
Flag_CA × Post_2022m5			-0.014	0.002
			[0.020]	[0.002]
Dep. Var. Mean	1.766	0.178	2.857	0.287
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	144,877	144,877	54,455	54,455
Adj R2	0.842	0.809	0.884	0.855

This table shows the effect of SB 826, both its passage (columns (1) and (2)) and repeal (columns (3) and (4)), on the number or share of female directors. The unit of observation is a firm-month. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018 (passage date), and 0 otherwise. Post_2022m5 is a dummy that equals 1 after May 13, 2022 (repeal date), and 0 otherwise. In columns (1) and (3), the outcome variable is the number of female directors on a board. In columns (2) and (4), the outcome variable is the share of female directors on a board. Columns (1) and (2) use data over the period 2014-2020, and columns (3) and (4) use data over the period 2021-2023. Robust standard errors clustered at the state (headquarter state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G4: Effect of California Statute on Gendered Language Usage in Job Ads: BERT-based Measure

Panel A: JobBERT

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	BoW	JobBERT	BoW	JobBERT	BoW	JobBERT
	Continuous		4 Group		10 Group	
Flag_CA × Post_2018m9	-0.042***	-0.018***	-0.174***	-0.095***	-0.463***	-0.309***
	[0.010]	[0.007]	[0.043]	[0.035]	[0.101]	[0.090]
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES	YES	YES
Occupation (ONET 6-digit) FE	YES	YES	YES	YES	YES	YES
Obs	1,642,550	1,573,942	1,642,550	1,573,942	1,642,550	1,573,942
Adj R2	0.434	0.57	0.409	0.594	0.416	0.638

Panel B: DistilBERT

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	BoW	BERT	BoW	BERT	BoW	BERT
	Continuous		4 Group		10 Group	
Flag_CA × Post_2018m9	-0.042***	-0.067***	-0.174***	-0.148***	-0.463***	-0.422***
	[0.010]	[0.008]	[0.043]	[0.029]	[0.101]	[0.081]
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES	YES	YES
Occupation (ONET 6-digit) FE	YES	YES	YES	YES	YES	YES
Obs	1,642,550	1,606,790	1,642,550	1,606,790	1,642,550	1,606,790
Adj R2	0.434	0.482	0.409	0.448	0.416	0.489

This table shows the effect of SB 826 on gendered language usage in job ads using alternative measures based on BERT. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. Panel A compares the gendered language usage measure using the BoW method with that using a fine-tuned JobBERT model. Panel B compares the gendered language usage measure using the BoW method with that using a fine-tuned DistilBERT model. In each panel, columns (1), (3), and (5) use the measure based on the BoW method as in the main specification. Columns (2), (4), and (6) use the measure from the JobBERT/DistilBERT model. The outcome variable is a continuous measure between 0 and 1 in columns (1) and (2). Columns (3) and (4) ((5) and (6)) categorize the continuous measure into 4 (10) groups. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G5: Effect of California Statute on Hiring and Employee Departure – Robustness to Alternative Sampling Criteria

	(1)	(2)	(3)	(4)
Dep. Var.	%Fem new hire		%Fem depature	
Sample	Excl. low coverage (industry)	Excl. low coverage (firm)	Excl. low coverage (industry)	Excl. low coverage (firm)
Flag_CA × Post_2018m9	-0.008**	-0.010***	0.009**	0.008**
	[0.003]	[0.004]	[0.004]	[0.004]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES
Obs	77,732	79,328	74,933	77,202
Adj R2	0.283	0.314	0.267	0.286

This table shows the effect of SB 826 on employee hiring and departures using alternative sample criteria. The unit of observation is a firm-month. The sample includes public firms covered by LinkedIn over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. In columns (1) and (2), the outcome variable is the share of female employees among new hires. In columns (3) and (4), the outcome variable is the share of female employees among departures. Columns (1) and (3) exclude industries known for low LinkedIn coverage. The following 10 industries are excluded: agricultural production crops, forestry, fishing, hunting, and trapping, metal mining, nonmetallic mineral mining and quarrying, heavy construction, construction special trade contractors, textile mill products, leather and leather products, stone, clay, glass, and concrete products, and motor freight transportation and warehousing. Columns (2) and (4) exclude firms with fewer than 50 employees from LinkedIn (bottom decile) to mitigate coverage bias in the LinkedIn data. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G6: Effect of California Statute on Gendered Language Usage in Job Ads – Heterogeneity: Partisanship and Masculinity, Within Firm-Month Comparison

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Share of feminine words out of all gendered words (%Fem words)							
Flag_CA × Post_2018m9 × TFE (in decades)	-0.007***						-0.006***	
	[0.001]						[0.001]	
Flag_CA × Post_2018m9 × > 20 years of TFE		-0.012***						-0.010***
		[0.002]						[0.003]
Flag_CA × Post_2018m9 × %Republican vote			-0.034***				-0.021*	
			[0.010]				[0.011]	
Flag_CA × Post_2018m9 × >50% Republican vot				-0.019***				-0.017***
				[0.003]				[0.003]
Flag_CA × Post_2018m9 × Masculinity (0-1)					-0.229***		-0.211***	
					[0.042]		[0.045]	
Flag_CA × Post_2018m9 × > Median masculinity						-0.006*		-0.004
						[0.003]		[0.003]
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry × Year-Month FE	YES	YES	YES	YES	YES	YES	YES	YES
Job Zip × Year-Month FE	YES	YES	YES	YES	YES	YES	YES	YES
Occupation (ONET 6-digit) FE	YES	YES	YES	YES	YES	YES	YES	YES
Obs	1,575,461	1,575,461	1,583,031	1,583,031	1,579,623	1,579,623	1,569,864	1,569,864
Adj R2	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470

This table shows heterogeneities in the effect of SB 826 on gendered language usage in job ads including Firm × Year-Month fixed effects. The outcome variable is the share of feminine words out of all gendered words in a job ad. The unit of observation is a job ad. The sample includes a random 1/15 of the full sample of job ads over the period 2014-2020. Flag_CA is a dummy that equals 1 for public firms headquartered in California, and 0 otherwise. Post_2018m9 is a dummy that equals 1 after September 30, 2018, and 0 otherwise. In column (1), TFE (in decades) measures a county's total frontier experience in decades, following the definition in Bazzi et al. (2020). In column (2), >20 years of TFE is a dummy variable that equals 1 for a county (county of a job) with more than 20 years of total frontier experience (approximately the 75th percentile) and 0 otherwise. In column (3), %Republican vote is the percentage of voter support for the Republican candidate in the 2020 US presidential election in a county (county of a job). In column (4), >50% Republican vote is a dummy that equals 1 for a county (county of a job) receiving more than 50% voter support for the Replication candidate in the 2020 US presidential election, and 0 otherwise. In column (5), Masculinity (0-1) is Hofstede's 2010 cultural dimension index (feminine-masculinity) in a county (county of a job), based on the approach in McLean et al. (2023). In column (6), > Median masculinity is a dummy that equals 1 if a county's masculinity score is above the sample median, and 0 otherwise. Columns (7) and (8) jointly examine the heterogeneous effects of individualism, partisanship, and masculinity. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G7: Effect of California Statute on Gender-Related Labor Violations

	(1)	(2)	(3)	(4)
	Seed words only		Expanded word list	
Dep. Var.	#Violations	log (1 + $Penalties)	#Violations	log (1 + $Penalties)
Flag_CA × Post_2018m9	0.005**	0.056***	0.005*	0.128***
	[0.002]	[0.021]	[0.003]	[0.031]
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Industry × Year FE	YES	YES	YES	YES
Obs	10,967	10,967	10,967	10,967
Adj R2	0.126	0.106	0.134	0.117

This table shows the effect of SB 826 on gender-related labor violations and penalties. The unit of observation is a firm-year. We use case-level data from Violation Tracker and aggregate them to firm-year level. In columns (1) and (2), the outcome variable is the number of or the penalties incurred by labor violations that can be labeled as gender-related. These violations mention any of the following seed words in the "description" or "secondary offense" field: female, gender, hostile, hours, leave, sex, sexual, support, and vacation. In columns (3) and (4), the outcome variable is the number of or the penalties incurred by labor violations that can be labeled as gender-related using an expanded word list. The following words are added as they have high associations with the above seed words in terms of pmi-freq scores (Jin et al., 2021): applicants, women, job, positions, discrimination, environment, medical, sick, and harassment. Robust standard errors clustered at the state (headquarters state of a firm) level are in brackets . * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G8: Experimental Evidence on Mechanisms – Removing Control Variables

Panel A: Comparing Treated and Control Participants

	(1)	(2)	(3)	(4)	(5)
		Mechanisms			
Dep. Var.	Manipulation check	Reactance	Norm violation	Moral licensing	Insecurity
Treatment	1.952***	1.529***	0.670***	0.246	0.281
	[0.203]	[0.245]	[0.217]	[0.182]	[0.210]
Controls	NO	NO	NO	NO	NO
Dep. Var. Mean	4.456	2.867	2.586	4.246	2.212
Obs	198	198	198	198	198
Adj R2	0.316	0.160	0.041	0.004	0.004

Panel B: Heterogeneity - Partisanship and Gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Mechanisms					
Dep. Var.	Reactance			Norm violation		
Treatment × Republican	1.793***		1.615***	1.712***		1.561***
	[0.473]		[0.487]	[0.414]		[0.445]
Treatment × Male		0.991**	0.591		0.934**	0.547
		[0.477]	[0.484]		[0.426]	[0.444]
Treatment	0.663**	1.040***	0.460	-0.156	0.208	-0.353
	[0.297]	[0.328]	[0.331]	[0.285]	[0.295]	[0.300]
Controls	NO	NO	NO	NO	NO	NO
Dep. Var. Mean	2.867	2.867	2.867	2.586	2.586	2.586
Obs	198	198	198	198	198	198
Adj R2	0.228	0.204	0.249	0.130	0.077	0.140

This table reports the treatment effect estimated from the experiment without control variables. The unit of observation is a subject. The sample includes 198 participants following the US Census racial distribution who passed the attention test. Treatment is a dummy that equals 1 for the treatment group and 0 for the control group. Panel A reports the treatment effect estimated from the experiment. The outcome variable is in the seven-point Likert scale for the manipulation question in column (1) and for the mechanism questions in columns (2) to (5). Panel B reports heterogeneous treatment effects based on partisanship and gender. Republican is a dummy that equals 1 if a subject's political view is greater than 4 on a 1-7 scale (ranging from liberal to conservative), and 0 otherwise. Gender is a dummy that equals 1 for male participants and 0 otherwise. Robust standard errors adjusted for heteroskedasticity are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table G9: Experimental Evidence on Mechanisms – Restricting Participants to Experienced Employees

Panel A: Comparing Treated and Control Participants

	(1)	(2)	(3)	(4)	(5)
		Mechanisms			
Dep. Var.	Manipulation check	Reactance	Norm violation	Moral licensing	Insecurity
Treatment	2.022***	1.631***	0.609***	0.248	0.204
	[0.239]	[0.279]	[0.230]	[0.198]	[0.221]
Controls	YES	YES	YES	YES	YES
Dep. Var. Mean	4.374	2.748	2.435	4.245	2.086
Obs	149	149	149	149	149
Adj R2	0.404	0.215	0.073	0.084	0.007

Panel B: Heterogeneity - Partisanship and Gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Mechanisms					
Dep. Var.	Reactance			Norm violation		
Treatment × Republican	1.805***		1.668***	1.975***		1.824***
	[0.543]		[0.549]	[0.454]		[0.458]
Treatment × Male		1.211**	0.894		1.299***	0.984**
		[0.565]	[0.549]		[0.460]	[0.427]
Treatment	0.764**	1.034***	0.389	-0.344	-0.031	-0.757**
	[0.327]	[0.383]	[0.393]	[0.290]	[0.316]	[0.321]
Controls	YES	YES	YES	YES	YES	YES
Dep. Var. Mean	2.748	2.748	2.748	2.435	2.435	2.435
Obs	148	148	148	148	148	148
Adj R2	0.273	0.237	0.282	0.185	0.118	0.209

This table reports the treatment effect estimated from the experiment by removing participants with fewer than five years of work experience or below 30. The unit of observation is a subject. The sample includes 149 participants who passed the attention test. Treatment is a dummy that equals 1 for the treatment group and 0 for the control group. Panel A reports the treatment effect estimated from the experiment. The outcome variable is in the seven-point Likert scale for the manipulation question in column (1) and for the mechanism questions in columns (2) to (5). Panel B reports heterogeneous treatment effects based on partisanship and gender. Republican is a dummy that equals 1 if a subject's political view is greater than 4 on a 1-7 scale (ranging from liberal to conservative), and 0 otherwise. Gender is a dummy that equals 1 for male participants and 0 otherwise. The following control variables are included: gender, partisanship, age, ethnicity, employment status, years of work experience. Robust standard errors adjusted for heteroskedasticity are in brackets. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.