Raising the Bar:
The Backlash of Gender Quotas

Juan B. González Alejandro Martínez-Marquina *

June 4, 2025

PRELIMINARY AND INCOMPLETE DRAFT, PLEASE DO NOT DISTRIBUTE

Abstract

Gender quotas are widely used to address gender disparities, but they may trigger backlash that undermines their effectiveness. In an online experiment simulating hiring decisions, we find clear evidence of such backlash. When participants acting as recruiters are required to hire an additional female candidate, they offer reduced salaries and lower hiring rates to other women–but only when female candidates underperform relative to males. Hence, quota backlash exists, but it is performance-specific. The presence of a quota raises the bar for evaluating other women, inadvertently intensifying scrutiny of the targeted group.

JEL classifications: D9, J16, J23, J71, M51
Keywords: gender; quotas; backlash; experiment

In recent years, gender quotas have emerged as a prominent policy tool to address gender disparities in various spheres of society, particularly in political and corporate domains. These quotas, whether legislative mandates or voluntary initiatives, aim to increase female representation and participation in decision-making roles. While gender quotas have shown promise in enhancing gender equity and diversity, they have also sparked significant backlash. Just in California, Proposition 209 banned race-based affirmative action in 1998 at public universities, and the Supreme Court recently struck down the recent mandate on women on corporate boards. One stark example from another country is South Korea, where the former prime minister recently ran a successful presidential campaign focusing on anti-feminism backlash and the promise of abolishing all gender quotas[1]. This backlash, often unexpected and multifaceted, has drawn attention to the complexities

*González: Department of Economics, University of Southern California, juanbgon@usc.edu. Martínez-Marquina: Marshall School of Business, University of Southern California, am04817@usc.edu. We thank Muriel Niederle, and the many Stanford SITE, and USC seminar participants for their invaluable feedback. We also thank all the participants from the LAX conference and Caltech workshop.

[1] https://time.com/6156537/south-korea-president-yoon-suk-yeol-sexism/

surrounding the implementation and effectiveness of gender quota policies.

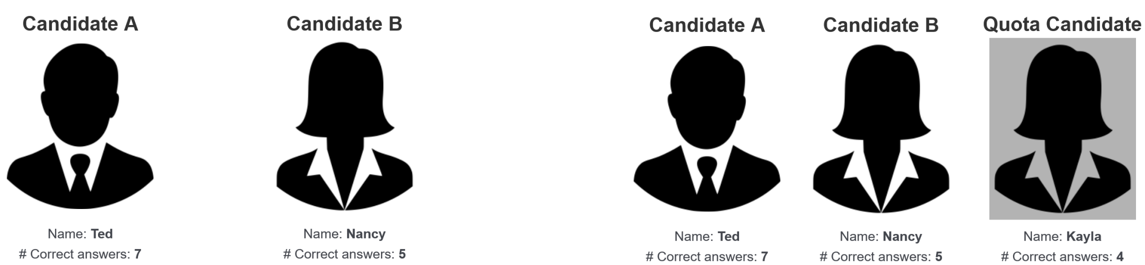
Despite the growing acceptance and adoption of gender quotas, particularly in politics, a notable trend of resistance and opposition is emerging from various quarters. This backlash is not limited to conservative or traditionalist factions but has also been among some groups for which these policies intend to benefit. Critics argue that gender quotas may lead to tokenism, undermine meritocracy, and perpetuate stereotypes about women's abilities and qualifications. See, for example, Bleemer (2022), as a case where URM students also show a dislike for affirmative action quotas. Moreover, concerns about fairness and the unintended consequences of quotas, such as resentment and backlash against women who benefit from them, underscore the need for a nuanced understanding of the dynamics at play. A quota policy could have well-intended consequences, but the counteraction to it could undermine the policy's primary purpose.

In this paper, we provide experimental evidence that quota backlash exists and its consequences for policies promoting equality. To do so, we have created a new experimental design where participants, acting as recruiters, send offers to available candidates.[2] Our main treatment innovation is comparing a situation with only two candidates, e.g., a man and a woman, with another where there are three possible candidates, e.g., a man and two women, where we force participants to hire one woman. Effectively, this leads recruiters in both conditions to choose between two candidates, a man and a woman, with the only difference being having to hire an extra candidate, which we refer to as the quota candidate.

Figure 1: Main Experimental Design - First Decision

(a) Baseline:
$7 to split across two candidates

(b) Quota Treatment:
$12 to split across three candidates
Must hire quota-candidate for $5



Notes: Shaded candidate corresponds to the quota candidate that participants must hire for $5.

Figure 1 summarizes our main between-subject treatment variation. In the *Baseline*, participants must split $7 between two potential candidates, with the option to either split the money between the two or uniquely pay one of them. If a candidate is hired, the recruiter receives $1 for each correct answer the candidate obtains in the last half of a 20-question quiz. In addition to

[2]See Bohren et al. (2022) and Exley and Nielsen (2024) for other examples in the literature where participants act as recruiters.

the gender of the candidates, we provide a measure of the candidate's proficiency at the task by displaying how many out of the first 10 questions they got right in the quiz. Hence, we provide an objective measure of productivity that is highly predictive of future performance. In the *Quota Treatment*, we increase the budget for hiring to \$12, but \$5 of those funds must be used in hiring the quota candidate. Hence, in both treatments, participants effectively have \$7 to be split among the two potential candidates. In the first part of the experiment, participants face 12 hiring decisions, similar to the one in Figure 1, where we randomized candidates' gender and performance. This provides variation in gender composition and relative candidate performance, which will be crucial in dissecting the quota backlash.

Our central hypothesis, referred to as the *Quota Backlash*, is that forcing the hiring of a female quota candidate (Kayla) creates a retaliatory response towards other female candidates who do not benefit directly (Nancy) compared to those in the *Baseline*. Thereby, diminishing the effectiveness of the quota in promoting the hiring of women. To rule out alternative explanations, such as recruiters having a preference for hiring a person from each gender, we run an additional treatment where the quota candidate is male.

We find strong evidence of quota backlash. In the first decision, where all participants encounter the same non-quota candidates and face the same incentives (Figure 1), the average wage gap between male and female candidates is \$2.49 in the *Baseline*, but increases to \$3.09 when a quota is introduced. Expanding the analysis to all 12 decisions, we find that the quota backlash emerges only when female candidates underperform relative to their male counterparts, as in the initial decision. In such cases, the effect is comparable to what we observe in the first round. However, when female candidates perform as well as or better than male candidates, the quota has no discernible impact. Hence, quota backlash exists, but it is performance-specific. The presence of a quota appears to raise the bar for evaluating other women, inadvertently subjecting them to heightened scrutiny.

An analysis of individual offers reveals that the quota backlash primarily acts through the extensive margin. Quotas increase by 50% the likelihood that a low-performing female candidate receives no offer at all. This effect is not limited to male-female comparisons—it also occurs when both candidates are female. Therefore, it is unlikely to stem from a preference for male candidates or a desire for gender diversity. Instead, the results point to a taste-based backlash against underperforming female candidates. By showing quotas don't lower the expected performance of female candidates and repeating our analysis with only participants who answered all comprehension questions correctly, we can rule out statistical discrimination and general confusion as alternative explanations.

Finally, we study how competition among recruiters affects gender disparities and the effectiveness of quotas. In the second part of the experiment, each participant is matched with two others, and only the highest offer among the three results in a hire. Contrary to the predictions of Becker's model of market discrimination (Becker, 2010), we find that competition amplifies gender discrimination. Even without quotas, female candidates receive lower offers and are more likely to receive no offer. When a woman is the worst candidate, the probability of receiving no offer rises from 15% in the non-competitive setting to 23% under competition ($p-$value $= 0.003$). In contrast, competition reduces the likelihood that an underperforming male candidate receives no offer falls from 22% to 13% ($p-$value < 0.001). Despite this, competition improves the effectiveness of quotas by limiting backlash against non-quota female candidates. By design, the quota policy should mechanically increase the probability of female candidates receiving an offer. However, in non-competitive settings, quota backlash undermines this effect by 36% for all women, and by 68% for low-performing ones. Under competition, the reduction in effectiveness caused by quota backlash falls to 11% and 19%, respectively. Thus, while competition intensifies baseline discrimination, we find that it is precisely in these settings where the quotas are most effective to promote female hiring.

This paper contributes to a large literature on the effects of gender quotas and related policies on labor market and political outcomes. Although quotas are intended to improve representation and aggregate outcomes for women, empirical findings are mixed. Some studies document gains in representation (Beaman et al., 2012; De Sousa and Niederle, 2022), while others find limited or unintended effects. Bertrand et al. (2019) study Norway's board quota and find that, despite increases in female representation, there is no clear improvement in firm performance or outcomes for non-appointed women. These limited effects suggest the presence of backlash or other frictions. Lee and Zanella (2024) document a political backlash in South Korea, where gender quotas reduce nominations of women to non-quota races, although this pattern reverses with exposure to competent female politicians. In education, Arcidiacono and Lovenheim (2016) show that affirmative action policies can affect admission standards and shape perceptions of merit, sometimes fueling resentment. These findings raise questions about how quotas affect perceptions and behaviors toward non-beneficiaries. By examining this channel directly, we find that quota backlash – retaliation against non-targeted individuals – significantly reduces their effectiveness.

Our design builds on the literature showing how experiments have proven particularly valuable in studying gender gaps, due to their ability to isolate causal mechanisms. Niederle and Vesterlund (2007) and Buser et al. (2014) demonstrate robust gender differences in competitiveness, which correlate with career choices. Exley et al. (2020) and Exley and Nielsen (2024) show that women strategically manage their behavior in anticipation of biased evaluations, and that evaluators often fail to adjust for gendered patterns in confidence. A closely related literature highlights the role of beliefs in perpetuating discrimination: Bohren et al. (2022) and Bohren et al. (2025) show that evaluators hold inaccurate beliefs about female performance and update asymmetrically, sustaining gender gaps in evaluation. Our work complements these findings by documenting how inaccurate performance beliefs and evaluators' stereotypes mediate reactions to gender quotas.

A growing literature documents how well-intentioned equity policies can backfire due to endogenous responses by firms and evaluators. For example, Cullen and Pakzad-Hurson (2023) show that pay transparency laws reduce wages as firms preempt costly renegotiations, and Passaro et al. (2023) find that equal pay mandates in the UK increased occupational segregation, unintentionally widening the gender pay gap. In the context of education, Otero et al. (2021) show that affirmative action policies can negatively affect outcomes for non-targeted individuals. We add to this literature by showing that gender quotas can elicit backlash from employers – particularly against lower-performing women who are not direct beneficiaries – thereby reducing the policy's effectiveness despite improving overall representation.

We organize the paper as follows. Section II describes the experimental design. Section III presents our main results on the existence of quota backlash. In Section IV, we examine the role of competition and the effects of quotas on female compensation, showing that, despite the backlash, quotas increase overall representation and pay. Section V addresses alternative explanations and demonstrates the robustness of our findings. Section VI discusses the underlying mechanisms and broader implications. Section VII concludes.

2 Experimental Design

Conceptual Framework - Defining the Quota Backlash

We formalize a simple framework to guide our empirical analysis. Consider a recruiter who evaluates two candidates, one male and one female, each with a signal of productivity s_g, for gender $g \in m, f$. After observing the signal s_g, the recruiter makes a wage offer w_g to each candidate, and hires a candidate if this offer is higher than their reservation wage r_g. If a candidate is hired, the recruiter receives their productivity p_g and pays them their reservation wage r_g. When making an offer, the recruiter holds some belief over each candidate productivity and reservation wage. We denote the marginal CDF of the posterior belief over reservation wages as $F_r(.)$.

To allow for gender preferences, we incorporate taste-based discrimination through a taste function $\theta_g(s_g)$ that weights the utility of hiring candidates of different genders and performance. The recruiter chooses wages to maximize expected utility subject to a fixed hiring budget $\mathcal{B}$:

$$\max_{w_m, w_f} \sum_{g \in \{m,f\}} \theta_g(s_g) F_{r_g}(w_g|s_g) E[p_g - r_g | s_g, r_g < w_g] \quad \text{subject to} \quad w_f + w_m \leq \mathcal{B} \tag{1}$$

Let w_m^* and w_f^* denote the optimal offers. We define the baseline wage gap as $\Delta_{\text{Base}} = w_m^* - w_f^*$. This gap may arise from expected productivity differences, reservation wage disparities, or taste-based bias.

Now consider the case of a quota policy where the recruiter must hire an additional female candidate at fixed cost w_q, and receives a budget increase of the same amount $(\mathcal{B} + w_q)$, leaving the discretionary budget unchanged. We denote the new optimal offers $w_m^{q\,*}, w_f^{q\,*}$. Although the budget is the same in both environments, the optimal offers might differ if the quota policy changes beliefs or taste. We define the wage gap in this environment as $\Delta_{\text{Quota}} = w_m^{q\,*} - w_f^{q\,*}$.

Our central hypothesis, which we refer to as *Quota Backlash*, posits that the introduction of the quota exacerbates the wage gap: $\Delta_{\text{Quota}} - \Delta_{\text{Base}} > 0$. That is, quotas worsen outcomes for non-beneficiary women, potentially due to increased scrutiny or reduced perceived marginal returns from hiring them. Importantly, this hypothesis does not require a baseline gap to exist. If $\Delta_{\text{Base}} = 0$, any positive wage gap under quotas would indicate backlash. Conversely, if baseline gaps exist (e.g., due to statistical or taste-based discrimination), they do not, on their own, imply backlash unless these gaps widen in response to the quota. Equation 1 illustrates the mechanisms through which quotas might produce backlash: by *changing* beliefs over productivity or reservation wages, and by *changing* preferences for candidates.

Our experimental design allows us to empirically test for such differential responses by holding budgets constant and systematically varying candidate performance and the gender of the quota beneficiary. We show that quota backlash exists, that it arises primarily when female candidates underperform ($s_f < s_m$), and that is driven by changes in taste rather than in beliefs.

Design Overview

Our experimental design involves two main types of participants: "workers" and "recruiters." Workers face a general knowledge quiz with 20 questions (including verbal and quantitative questions, like GRE or SAT standardized tests), for which they receive 10 cents per correct answer. Additionally, we also elicit their preference for being paid based on the quiz or a specific monetary amount, which we refer to as their *reservation wage*. Participants acting as recruiters choose which workers to hire. At each hiring decision, they receive a fixed budget and are presented with at least two candidates to whom they can extend offers. They observe the gender, proxied by a name, and the previous performance of each candidate, i.e., how many out of 10 questions they answered correctly in the quiz. Offers work as follows: a recruiter offers an amount within the budget. If the amount is larger or equal than the worker's *reservation wage* (unknown to the participant), they hire the worker and pay them their *reservation wage*, keeping for themselves the remaining amount. Therefore, an offer reflects the maximum willingness to pay to hire someone. For a hired candidate, the recruiter receives \$1 for each correct answer the candidate obtained in the remaining 10 questions of the quiz. If the worker is not hired, the recruiter just keeps the offered money for themselves. Since any amount of the budget that is not used is lost, they have no reason not to use all their funds to making offers. In fact, we later use this condition as an indicator of participants' confusion.

In our Baseline treatment, participants receive \$7 and see two potential candidates, selected at random from the pool of workers. Their option are then to split the money among the two or uniquely pay one of them. Our main treatment variation consists of introducing an additional candidate, referred to as the *Quota Candidate*. In this condition, all quota candidates are female, a requirement that we later relax in additional robustness treatments. In the *Quota* treatment, we increase the budget for hiring to \$12, but they must hire the quota candidate for \$5, rendering their available budget back to \$7. Hence, recruiters are always choosing how to split \$7 among the two potential candidates in both treatment arms. Our main hypothesis is that participants in the *Quota* group will negatively react to being forced to hire a woman by reducing offers to other female candidates, effectively diminishing the effectiveness of the quota in promoting women's hiring.

Table 1: Summary of Experimental Design

Part I	12 hiring decisions First decision is the same non-quota candidates for everyone Candidates' performance and gender of the remaining 11 decisions randomized Drawn from a 100 gender-balanced pool of workers
Part II	12 hiring decisions, competing with two other participants Hire if the offer is higher than the competitors'
End	Avg.Performance Elicitation, strategy description, demographics and policy views

The experiment is divided into two parts. In part I, participants face 12 hiring decisions, with two non-quota candidates available to hire at each and no feedback between decisions. Candidates are randomly selected from the pool of workers, which introduces variation in gender and performance. The only decision not randomized is the first one, which presents the candidates from Figure 1. Part II of the experiment aims to measure how recruiter competition interacts with quota backlash. For that purpose, we match each recruiter with two other participants competing for the same candidates. Now, a recruiter only manages to hire a candidate if their offer is better than the others. This introduces an additional layer of strategy, which could either increase or decrease gender differences: If the recruiter realizes that others might be discriminating against women, it might be more profitable to try harder to hire the candidate suffering discrimination. Alternatively, a discriminating participant might discriminate even more to hire their preferred candidate. To ensure the quality of their responses, participants must correctly answer several comprehension questions during the instruction period and spend at least 15 seconds in each decision before submitting their offers. At the end of the study, we also asked participants to describe their strategies and to estimate the average performance of each group, men and women, in addition to eliciting their demographic characteristics and support for quota policies.

Participants and Procedures.- Subjects were recruited online on Prolific. We recruited a gender-balanced sample of XX participants with unique IP addresses, all of whom reside in the US. Participants must have an approval rate above 90% on the platform and have done at least 100 previous submissions. Additionally, the sample is politically balanced, comprising half self-declared Democrats and half self-declared Republicans[3]. A total of 405 participants took part in the two main treatments and XX in the additional ones. During the instruction period, participants were required to correctly answer several comprehension questions before proceeding to the main decisions. We record the number of errors each subject makes in the comprehension questions and use this as a control in our analysis. Participants seem to understand the instructions: 47% of them make no mistakes in the comprehension questions, and 75% make no more than one. As a robustness check, we also report results restricting the analysis to those who answer all questions correctly. Subjects who finish the entire experiment are paid a $4 participation fee and a bonus based on their performance. One decision in Part I and another in Part II are randomly selected for payment. In addition, subjects are paid 50 cents if they correctly guess the number of correct answers for each group, men, women, and quota candidates, if applicable. The average subject made $12 and took 30 minutes to complete the survey.

Additional Treatment.- As a robustness check, we implement one additional treatment, *Quota Male*, where the designated quota candidate is male. This treatment allows us to test whether backlash stems from a general aversion to quotas or a preference for diversification, rather than gender-specific bias.

Why an Experiment

Despite its simplicity, our design enables us to isolate the causal effect of introducing gender quotas while ruling out several confounding factors common in observational settings. In practice, the

[3]See Table ?? for demographic information and the sample balance.

implementation of quotas often reduces opportunities for non-targeted groups unless capacity or funding is simultaneously increased. For instance, a college that reserves a fixed number of spots or scholarships for a particular group may do so at the expense of others unless it expands. Moreover, it is often ambiguous who directly benefits from the quota, as this information is typically unobservable. Individuals from the targeted group may have qualified independently of the policy, leading to uncertainty and potential heterogeneity in beliefs and priors. Our experimental setting addresses this challenge by explicitly identifying the candidates who benefit from the quota, even when all belong to the targeted group. Additionally, we provide an objective measure of each candidate's productivity, allowing us to examine how outcomes vary with relative performance and minimizing the scope for statistical discrimination. This enables a clearer identification of taste-based responses.

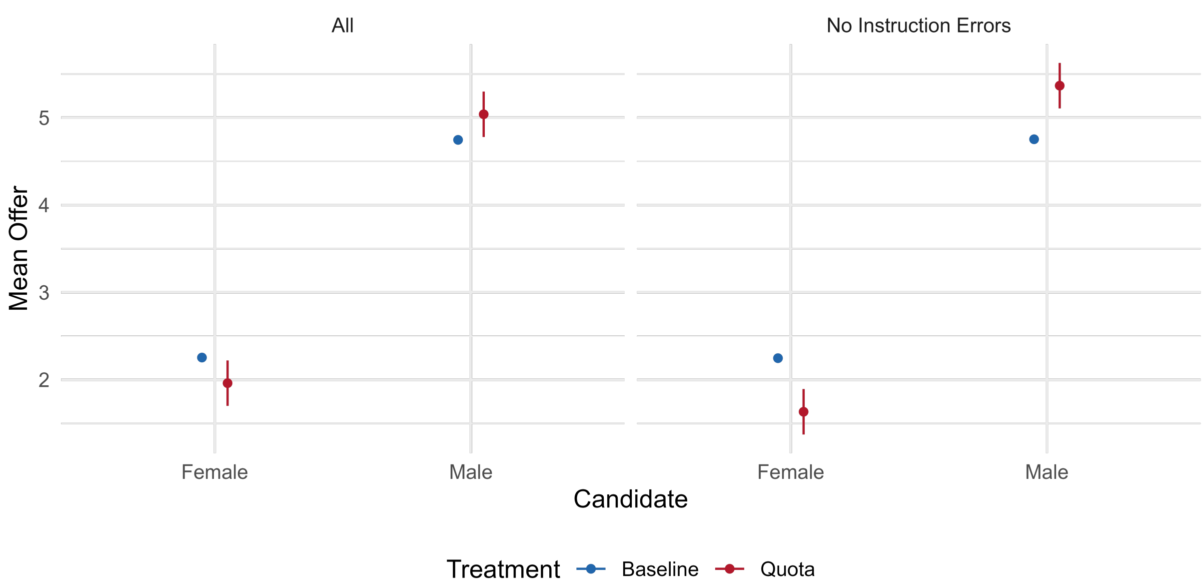
Finally, our design isolates the effect of quotas while holding the candidate selection constant. In empirical contexts, the candidate pool may change following the introduction of a quota, either by encouraging applications from the targeted group or by discouraging them in the absence of a quota. For example, Bleemer (2022) documents a decline of 250 Black and 900 Hispanic applicants per year to the UC system following the passage of Prop 209, which banned race-based affirmative action. Conversely, De Sousa and Niederle (2022) finds that the introduction of a gender quota in French chess teams led to a trickle-down effect, encouraging more women to join. While these dynamics are important, they complicate the identification of the causal effect of quota backlash, as the observed effect is a composite of both changes in the applicant pool and the response to the quota itself.

3 Main Result - The Backlash of Gender Quotas

This section presents our evidence on the backlash of gender quotas, meaning that when participants are forced to hire a quota candidate who is female, we expect them to retaliate against other female candidates not targeted by the quota.

We begin by examining the gender gap in the first hiring decision, where the non-quota candidates are the same for all participants. In this decision, the male candidate has a performance advantage (7 vs. 5 correct answers), so we expect a gender gap to exist. What we find is that the presence of a quota candidate substantially increases the gender gap: wage differences between male and female offers grows from \$2.49 in the *Baseline* to \$3.08 in the *Quota Treatment*, implying a 24.34% widening of the gap (p-value = 0.027). When restricting the analysis to participants who did not make any errors in the instructions, the gap becomes even larger, with the difference between the two treatments increasing to \$1.23 (p-value = 0.001). Thus, at least in the first decision, we find strong evidence in support of the quota backlash.

Figure 2: Mean Offers in the First Decision

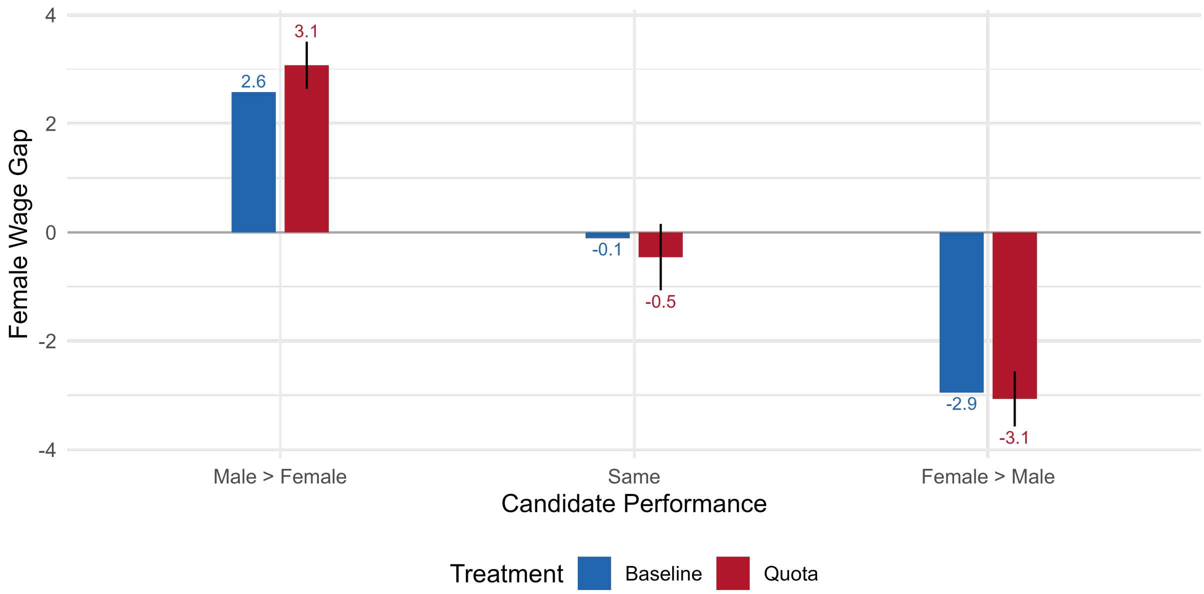


Notes: Error bars correspond to the 95% confidence interval of the difference in means after controlling for demographic characteristics and errors in the instructions. Controls include dummies for whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. See Appendix Table XX for details.

We look next at the wage gap in all twelve decisions from Part I involving a female and male candidate, separated by relative performance (Figure 3). As expected, candidates with higher performance than their counterparts receive better offers, showing participants pay attention to the performance measure. However, the gap is affected by the presence of the quota: when the male candidate outperforms the female one, they receive an offer \$2.58 higher in the *Baseline*. This difference grows larger in the *Quota Treatment*, where the outperforming male now receives an offer

$3.07 higher. In percentage terms, this $0.49 difference represents a 19% increase in the pay gap caused by the presence of quotas (p-value = 0.026).[4]

Figure 3: Relative Performance Wage Gap



Notes: Error bars correspond to the 95% confidence interval of the difference in means after controlling for demographic characteristics and errors in the instructions. Point estimates include the same controls: dummies for whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. We only use decisions where the entire budget was allocated and when the candidates are of different genders. See Appendix Table XX for details.

Overall, these results demonstrate that quota backlash exists, but its impacts are performance-specific. When women candidates outperform men or perform equally, quotas do not create a backlash for other women. However, in situations where women underperform their male peers, we observe that quotas exacerbate wage differences, resulting in reduced pay for female candidates. Hence, it is as if the presence of a quota raises the bar by which participants judge other women. So far, these results examine average offers, making it difficult to determine whether the intensive margin or the extensive margin matters most. Hence, we look an individual offers next.

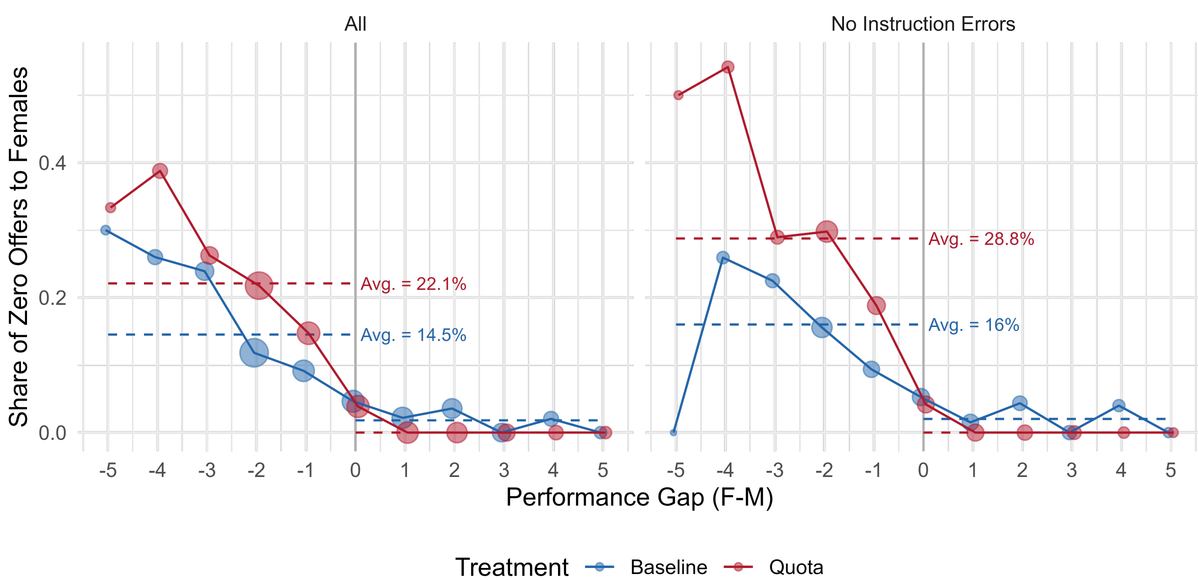
Looking at the distributions of offers (Appendix Figure A.9) reveals that the extensive margin is the primary driver of the quota backlash. Women are 7.6pp (52%) more likely to receive zero offers in the *Quota Treatment* (p-value = 0.025) and are also less likely to receive $2 offers (p-value = 0.000) when they underperform compared to the male candidate, suggesting a shift between both choices. Motivated by this finding, we study in detail how the likelihood of receiving no offer depends on the performance difference between the candidates. Figure 4 illustrates the share of zero offers based on the performance gap, with positive values indicating that the female candidate

[4]This difference is robust to focusing on participants with no instruction errors. The gap widens, with males in *Quota Treatment* receiving a $0.66 higher premium with respect to females, compared to decisions in the *Baseline* (p-value = 0.03).

outperforms the male candidate, and vice versa. A clear pattern emerges: better female candidates do not experience any backlash in the *Quota Treatment* while underperforming ones do. On average, a woman with a lower performance than a man has a 14.5% chance of receiving no offers in the *Baseline*. This probability increases to 22.1% once we introduce quota candidates, representing a 52 percent increase (p-value = 0.031). Once again, focusing on participants with no instruction errors exacerbates this result, with the quota effect widening to 12.8 percentage points, an 80% increase (p-value = 0.014).

Same-gender decisions reveal that women are penalized even when there is no other male candidate to compare them with. Figure A.10 depicts our previous analysis but now using the performance gap within same-gender decisions. We find again that the likelihood of underperforming female candidates receiving no offer increases in the *Quota Treatment* by 10.7 percentage points (p-value = 0.008), while there is no significant effect among underperforming males (p=0.719). That low-performance female candidates are penalized not only when they are compared to male candidates shows the quota backlash is not driven by quotas changing beliefs of female performance relative to males.

Figure 4: Likelihood of No Offer by Female Performance Gap



Notes: Circle size represent sample size. Performance gap refers to the difference in performance between female and male candidates in decisions where candidates are of different genders. Horizontal lines depict the average likelihood of receiving zero offers when the male candidate outperforms the female candidate. The left panel only uses decisions where the entire budget was allocated and when the candidates are of different genders. The right panel also restricts to participants who answer all understanding questions correctly. See Appendix Table XX for details.

Finally, to test whether the quota backlash is driven by gender bias, we implement the *Quota Male* treatment, where the designated quota candidate is a male. If preferences for diversification or a general aversion to quota policies were causing the quota backlash irrespective of gender, there would also be backlash in the *Quota Male* treatment. However, we find no evidence of quota backlash against males. Figure A.11 shows that male quotas don't affect wage gaps between male and female candidates, and Figure A.12 reveals no quota backlash against males at the extensive margin either. These results demonstrate quota backlash is specific against women, which suggests it arises from gender bias.

3.1 The Role of Competition

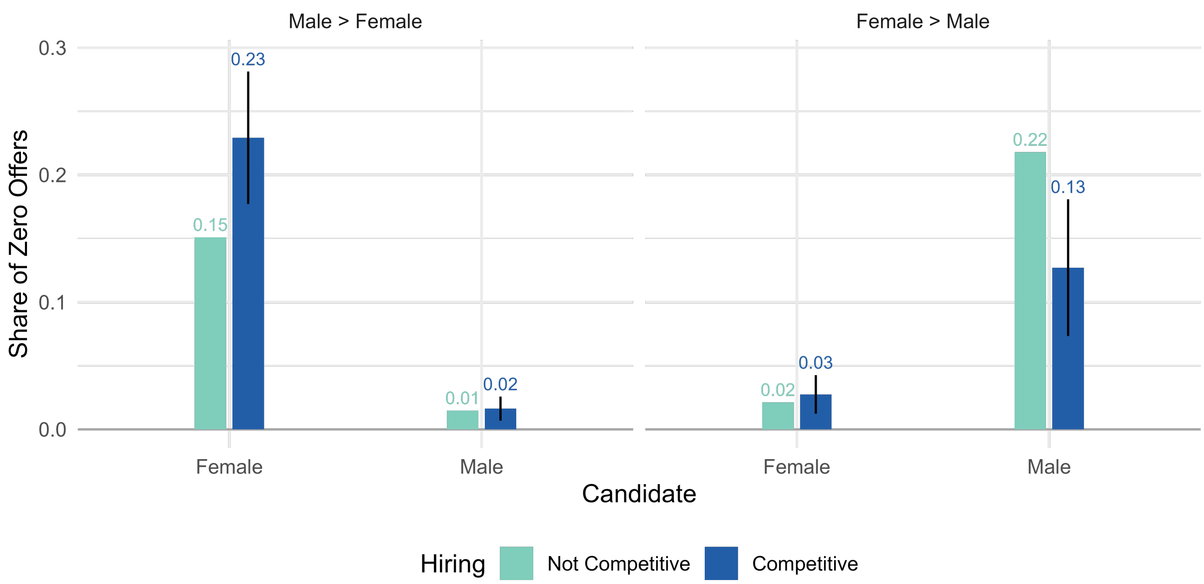
We begin by analyzing the direct effects of competition on discrimination. A key hypothesis, originally proposed by Becker (2010), suggests that competition reduces discrimination in labor markets. The rationale is that discriminatory employers incur higher costs by passing over equally productive but lower-paid workers from marginalized groups, leading competitive forces to eliminate discrimination over time. However, competition could also intensify discrimination if employers bid competitively for their preferred candidates, amplifying their own biases. Our experimental findings show that competition significantly increases discrimination. In the *Baseline* treatment, Figure 5 illustrates that competitive pressure prompts participants to increase the wage premium for high-performance male candidates by \$0.42 or 16.3% (p-value = 0.011). In contrast, the impact of competition on high-performance female candidates is negative: their wage premium decreases by 0.63\$ or 21.3% in the *Baseline* condition (p-value = 0.000). Taken together, these results imply that when competitive pressure is present, recruiters tend to favor male candidates, irrespective of their relative performance.

Figure 5: Relative Performance Wage Gap By Treatment

Notes: Error bars correspond to the 95% confidence interval of the difference in means after controlling for demographic characteristics and errors in the instructions. Point estimates include the same controls: whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. We only use decisions where the entire budget was allocated and when the candidates are of different genders. See Appendix Table XX for details.

Competition also amplifies discrimination at the extensive margin of hiring decisions. Figure 6 shows that competition increases gender disparities in the likelihood of receiving no offer. Among underperforming candidates, competition increases this likelihood for female candidates by 7.8 percentage points ($p-$value $= 0.003$), but reduces it by 9.1 percentage points for male candidates ($p-$value < 0.001). Overall, competition makes women more likely to be overlooked for offers, and makes men more likely to receive an offer even when they are worse candidates.

Figure 6: Probability of No Offer by Competition and Candidate Performance



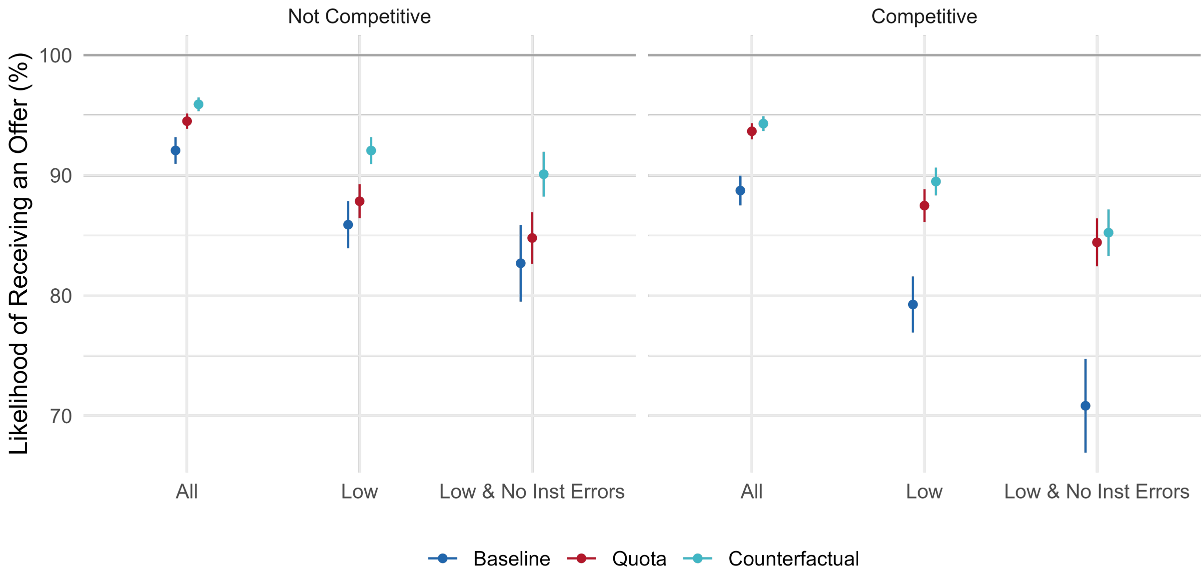
Notes: Error bars correspond to the 95% confidence interval of the difference in means after controlling for demographic characteristics and errors in the instructions. Point estimates include the same controls: whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. This figure focuses on decisions in the *Baseline* treatment. We only use decisions where the entire budget was allocated and when the candidates are of different genders. See Appendix Table XX for details.

3.2 The Efficiency of Gender Quotas

Next, we examine whether competition mitigates quota backlash. Figure 5 shows that competition reduces the quota backlash at the wage margin: when hiring is competitive, quotas don't penalize female candidates as much. To better visualize the quota backlash, we construct a counterfactual where quotas have no backlash on non-targeted candidates but only add a \$5 offer to a female candidate in each decision [5]. Thus, the difference between the counterfactual and the outcomes in the *Quota Treatment* captures the backlash against non-targeted candidates, revealing how backlash undermines the effectiveness of quotas.

[5] In particular, we calculate the counterfactual by adding the mandated \$5 offer to each hiring decision in the *Baseline* treatment.

Figure 7: Likelihood of Female Candidates Receiving an Offer



Notes: Error bars correspond to ± SE of the mean, controlling for whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. The left panel shows the share of positive offers made to female candidates when hiring is not competitive, the right panel when hiring is competitive. Column *All* shows all decisions, Column *Low* restricts the sample to underperforming female candidates, and *Low & No Inst Errors* further restricts it to participants who made no understanding errors during the instructions. The likelihood of receiving an offer in the *Quota* and *Counterfactual* environments is calculated by artificially adding a $5 offer to an additional female candidate for each decision in the *Quota* and *Baseline* treatment, respectively. The counterfactual should be interpreted as the likelihood of receiving an offer if the introduction of quotas had no backlash on non-targeted candidates. Both panels only use decisions where the entire budget was allocated.

Figure 7 shows that in non-competitive hiring decisions, gender backlash limits the positive effect of quotas on female hires, reducing the overall effectiveness of quotas by 36%. Among underperforming candidates, quota backlash reduces the effectiveness of quotas by 68%, nearly nullifying the quota's intended gains. In contrast, when hiring is competitive, the negative impact of quotas on non-targeted female candidates is substantially lower. Gender backlash only reduces by 11% the effectiveness of quotas for female candidates, and only by 5% for underperforming candidates among those participants who get the instructions right. Because outcomes for female candidates are already adversely affected by competition, there is limited scope for quotas to further worsen their position. These are the scenarios where quotas seem to be most effective.

Ale: downstream profits are the same since reservations wages are more than 2, where most. Avg. reservation of female is XX and for males is XX. Consistent with evidence from Nielsen et al. on gap in overconfidence.

4 Robustness and alternative explanations

In this section, we rule out alternative explanations for our results and present a robustness analysis using a treatment with male quota candidates.

Are our results driven by confusion?

When interpreting our findings, we rely on the assumption that participants understand the experimental decision problem equally well in both treatments. To minimize differential confusion across treatments, we limited differences between instructions and included a comprehensive set of understanding questions. Almost 47% of participants answered all questions correctly, and 75% make no more than one mistake, with understanding errors in the instructions balanced across treatments[6]. Notably, all our results remain or become even more pronounced when restricting the sample to those who answer all questions correctly.

Are our results driven by statistical discrimination?

Baseline statistical discrimination by itself doesn't imply backlash: backlash requires quotas amplifying statistical discrimination, as discussed in the conceptual framework. This could be a mechanism of backlash if quotas implicitly convey information about non-quota candidates that makes participants update their beliefs over productivity or reservation wages. For example, a recruiter could infer that, if quotas target female candidates, female candidates must be less productive than males. While plausible, we find no evidence supporting this mechanism.

Regarding performance beliefs, note that participants observe individual signals that are highly informative for both male and female candidates, so it's unclear why quotas would override these signals. Anyway, if changes in performance beliefs are causing the backlash, quotas should relatively increase expectations about male performance. However, we observe the exact opposite: quotas *increase* the perceived performance of female candidates more than that of males, improving accuracy and reversing the gap in expectations [7]. This effect is even more pronounced for those who answer the instructions correctly. Finally, our results do not vary across participants with different beliefs of performance, and including perceived performance in the analysis doesn't change the impact of quotas. These results are consistent with Dianat et al. (2022), who find that affirmative action does not have persistent effects on beliefs of performance and statistical discrimination.

Although recruiters might believe that if quotas target female candidates they might have lower reservation wages, it's theoretically unclear whether this would generate backlash. Lower reservation wages have two opposing effects on optimal offers: candidates with lower wages become at the same time more easily hired at a given offer (which depresses offers) and less costly in the margin (which

[6]In our main treatments, the average number of mistakes in the instructions is 1.12 for the *Baseline* and 1.08 for the *Quota Treatment* (p-value = 0.826).

[7]Predicted performance of female candidates is 5.55 in the *Baseline* and 5.89 in the *Quota Treatment*. In comparison, average predictions for males are 5.66 and 5.84, respectively.

incentivizes higher offers). Importantly, we find no evidence of participants considering reservation wages when making offers. When asked what strategy they followed, virtually all participants report making offers according to performance, and only one participant mentions reservation wages at all. Decisions in the *Baseline Treatment* back the strategies reported by participants: offers mostly vary by performance and not at all by gender, despite female candidates having significantly lower reservation wages. The lack of a wage gap in the *Baseline Treatment* shows participants are not sophisticated about reservation wages.

Are our results driven by demand effects?

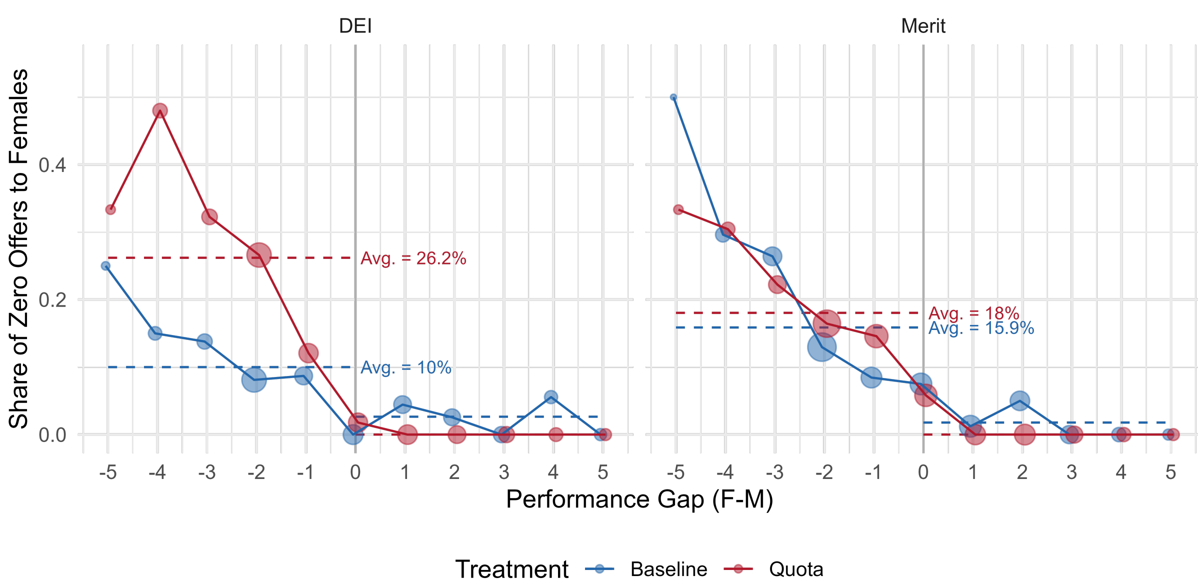
The quota policy in the *Quota treatment* is intentionally salient. A potential explanation for our results is that quotas make participants realize this is an experiment about gender, and they respond to experimenter demands by following what they perceive as social norms. For this mechanism to generate a backlash, participants driving the results should be those reporting views against quotas and gender diversity. However, we observe the opposite: the participants who exhibit more quota backlash are those who agree with affirmative action to improve diversity in hiring and college admissions.

5 Discussion

Our experimental results show that quota backlash exists and is driven by taste rather than statistical discrimination. To further explore the mechanism of quota backlash, we briefly discuss heterogeneous treatment effects across participants' characteristics and types of decisions. As the extensive margin is the main driver of quota backlash, we focus on the effect of quotas on the likelihood of receiving no offers.

We begin by examining how participants' policy views relate to their revealed quota backlash. At the end of the experiment, participants report their opinion on 5 policy issues using a 1 to 5 Likert scale. Figure 8 shows the quota backlash at the extensive margin when dividing the sample by views on affirmative action in college admissions. Counterintuitively, we find that quota backlash is most pronounced among those participants who support affirmative action in college admissions (16pp, $p-$value $= 0.001$), and doesn't arise among participants who believe applicants should be admitted solely on the basis of merit. Figure A.13 extends this analysis to other policy views, showing that quota backlash is higher among liberals and those who agree with gender quotas on the boards of firms or that the country has not gone far enough in giving women equal opportunities with men. Overall, quota backlash is driven by those participants who support affirmative action and quota policies, suggesting the backlash is caused by implicit bias.

Figure 8: Heterogeneity by Policy Views - College Admissions



Notes: Circle size represent sample size. Performance gap refers to the difference in performance between female and male candidates in decisions where candidates are of different genders. Positive values reflect female candidates that outperform the male candidate. Horizontal lines depict the average likelihood of receiving zero offers. The panels divide the sample according to their views on diversity in college admissions: the left panel shows decisions by participants who believe "applicants' racial and ethnic background should be considered to help promote diversity", while the right panel shows those who believe "applicants should be admitted solely on the basis of merit". We only use decisions where the entire budget was allocated and when the candidates are of different genders. See Appendix Table XX for details.

Notably, quota backlash also depends on the performance of the mandated quota candidate relative to the non-targeted female. Figure A.13 shows that when the quota candidate has a lower performance, quota backlash reduces the share of non-targeted females that receive an offer by 9.2 percentage points ($p-$value $= 0.019$). In contrast, there is no backlash when quota candidates have a higher performance ($p-$value=0.218). This difference occurs despite a lower-perfomance quota candidate implying a higher-performance non-targeted female, and viceversa. When controlling for the performance of the non-quota female candidate, backlash rises to 11.3 percentage points ($p-$value < 0.001) in decisions where the quota candidate has a lower performance. Consistent with this heterogeneous effects, we find more pronounced backlash among participants who believe that candidates hired due to diversity policies are less qualified. This evidence, when taken together with the null effects of quotas on the perceived performance of non-targeted females, suggests that quota backlash is driven by recruiters punishing underperforming females in response to having to hire low productivity quota candidates. Importantly, this is a taste-based gendered backlash, as there is no backlash against men in response to enforcing male quotas.

6 Conclusion

This paper proposes and provides experimental evidence for the existence of backlash to gender quotas. We find that mandating the selection of a female candidate significantly reduces the likelihood that other women, those not targeted by the quota, receive offers. This backlash is concentrated among lower-performing candidates, suggesting that it is performance-contingent.

A key advantage of our experimental design is the use of objective and transparent performance metrics, which allow us to precisely measure how the effects of quotas vary with candidate quality. In contrast, performance is often noisy or unobserved in real-world settings, creating scope for distorted beliefs. As shown in Bohren et al. (2025), such misperceptions can play a central role in discriminatory behavior. Our results suggest that when women's abilities are underestimated, backlash is more likely to occur. This poses a significant challenge for quota policies, as it is precisely in environments where women are underestimated that a policy-maker might be more inclined to implement quotas. While our study does not directly test information interventions, our findings imply that correcting biased beliefs may be an effective strategy for mitigating backlash.

While the controlled laboratory setting is well-suited to identifying the existence and underlying mechanisms of quota backlash, we acknowledge the importance of assessing its relevance in real-world contexts. Although the effects we document are sizable—quotas increase the likelihood that non-beneficiary women receive no offer by 50 percent—the extent to which these results generalize remains an open question. We view this study as a starting point, motivating further investigation into the external validity of quota backlash. Similar trajectories have been observed in prior experimental work. For example, the gender gap in competitiveness first documented in laboratory experiments by Niederle and Vesterlund (2007) was later shown to predict educational and occupational choices in Buser et al. (2014). Likewise, Babcock et al. (2017) identify gender differences in volunteering for low-promotability tasks in the lab and validate their findings using survey data from university departments. We view the establishment of quota backlash in field settings as a promising avenue for future research.

Finally, our results document a new retaliatory response of quotas that can help rationalize why quotas are, at times, ineffective. One of the most well-known examples is Bertrand et al. (2019), which looks at the impact of gender quotas on corporate boards in Norway. Despite newly appointed female board members being more qualified than their predecessors, they find no improvements for lower-ranked women or at the firm level, pointing to a backlash that prevents aggregate improvements. One example where exposure to high-performing women eventually eroded previous barriers is Lee and Zanella (2024), which finds that political parties in South Korea react to gender quotas by nominating fewer women to positions unaffected by the policy. However, parties gradually reverse this response as they get exposed to competent women. Therefore, the reaction to gender quotas appears to hinge critically on the perception of women's competence[8]. To design better policies to promote equity, it is crucial to understand all their consequences.

References

Arcidiacono, P. and Lovenheim, M. (2016), 'Affirmative action and the quality–fit trade-off', *Journal of Economic Literature* **54**(1), 3–51.

Babcock, L., Recalde, M. P., Vesterlund, L. and Weingart, L. (2017), 'Gender differences in accepting and receiving requests for tasks with low promotability', *American Economic Review* **107**(3), 714–747.

Beaman, L., Duflo, E., Pande, R. and Topalova, P. (2012), 'Female leadership raises aspirations and educational attainment for girls: A policy experiment in india', *science* **335**(6068), 582–586.

Becker, G. S. (2010), *The economics of discrimination*, University of Chicago press.

Bertrand, M., Black, S. E., Jensen, S. and Lleras-Muney, A. (2019), 'Breaking the glass ceiling? the effect of board quotas on female labour market outcomes in norway', *The Review of Economic Studies* **86**(1), 191–239.

Bleemer, Z. (2022), 'Affirmative action, mismatch, and economic mobility after california's proposition 209', *The Quarterly Journal of Economics* **137**(1), 115–160.

Bohren, J. A., Haggag, K., Imas, A. and Pope, D. G. (2025), 'Inaccurate statistical discrimination: An identification problem', *Review of Economics and Statistics* pp. 1–16.

Bohren, J. A., Hull, P. and Imas, A. (2022), Systemic discrimination: Theory and measurement, Technical report, National Bureau of Economic Research.

Bohren, J. A., Imas, A. and Rosenberg, M. (2019), 'The dynamics of discrimination: Theory and evidence', *American economic review* **109**(10), 3395–3436.

Buser, T., Niederle, M. and Oosterbeek, H. (2014), 'Gender, competitiveness, and career choices', *The quarterly journal of economics* **129**(3), 1409–1447.

[8]Bohren et al. (2019) is another example where female discrimination reverses after receiving a sequence of positive evaluations in a field experiment on a large online platform.

Cullen, Z. B. and Pakzad-Hurson, B. (2023), ‘Equilibrium effects of pay transparency’, *Econometrica* **91**(3), 765–802.

De Sousa, J. and Niederle, M. (2022), Trickle-down effects of affirmative action: A case study in france, Technical report, National Bureau of Economic Research.

Dianat, A., Echenique, F. and Yariv, L. (2022), ‘Statistical discrimination and affirmative action in the lab’, *Games and Economic Behavior* **132**, 41–58.

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

Exley, C. L. and Nielsen, K. (2024), ‘The gender gap in confidence: Expected but not accounted for’, *American Economic Review* **114**(3), 851–885.

Lee, J. E. and Zanella, M. (2024), ‘Learning about women’s competence: the dynamic response of political parties to gender quotas in south korea’.

Niederle, M. and Vesterlund, L. (2007), ‘Do women shy away from competition? do men compete too much?’, *The quarterly journal of economics* **122**(3), 1067–1101.

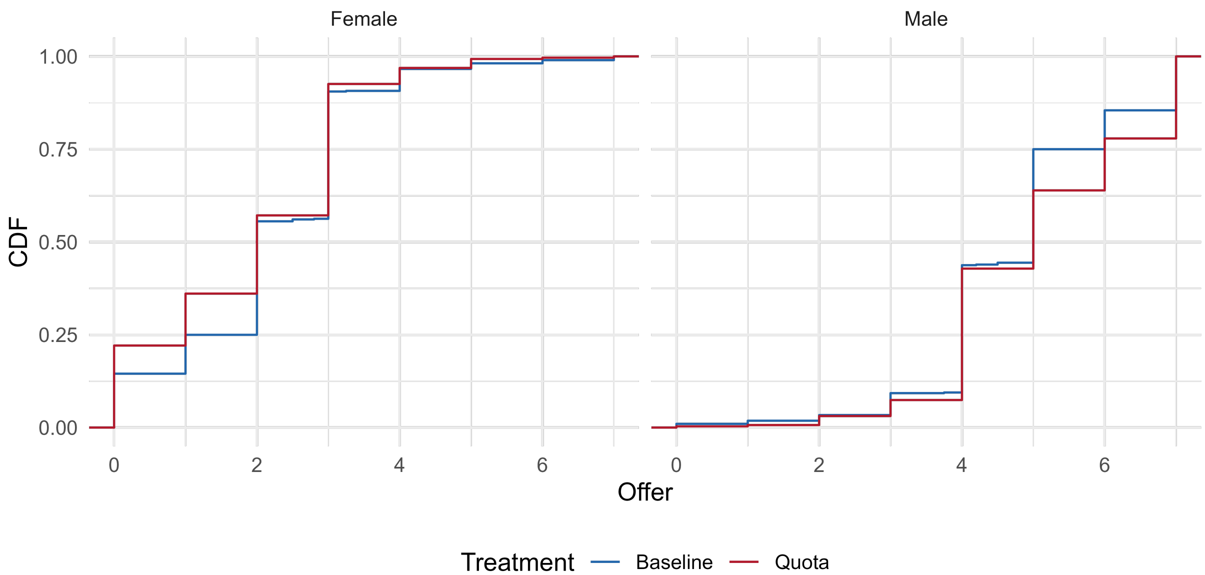
Otero, S., Barahona, N. and Dobbin, C. (2021), ‘Affirmative action in centralized college admission systems: Evidence from brazil’, *Unpublished manuscript* .

Passaro, D. G., Kojima, F. and Pakzad-Hurson, B. (2023), ‘Equal pay for similar work’, *arXiv preprint arXiv:2306.17111* .

A Appendix

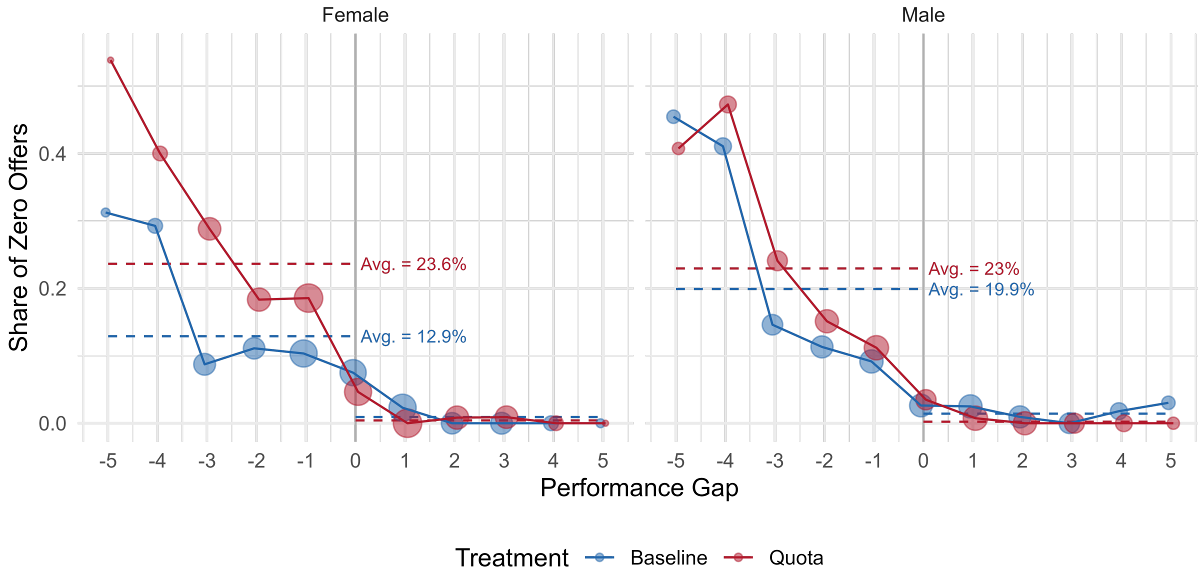
A.1 Additional Figures and Tables

Figure A.9: Offer Distribution when Male outperforms Female



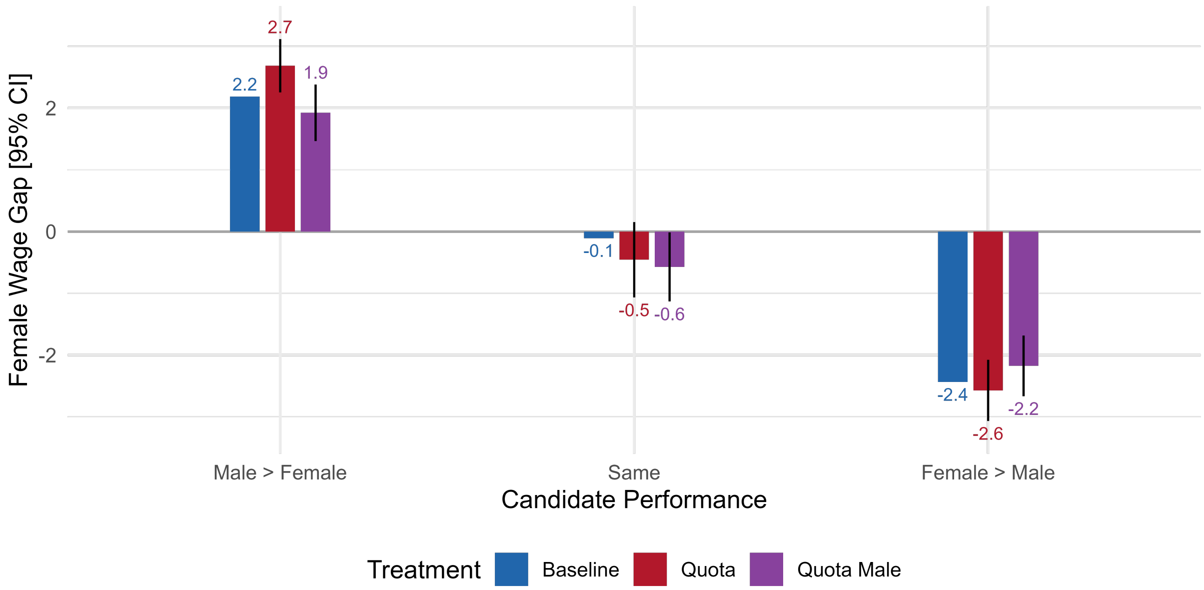
Notes: Add notes. We only use decisions where the entire budget was allocated and when the candidates are of different genders. See Appendix Table XX for details.

Figure A.10: Likelihood of No Offer by Performance Gap - Same Sex Pairs



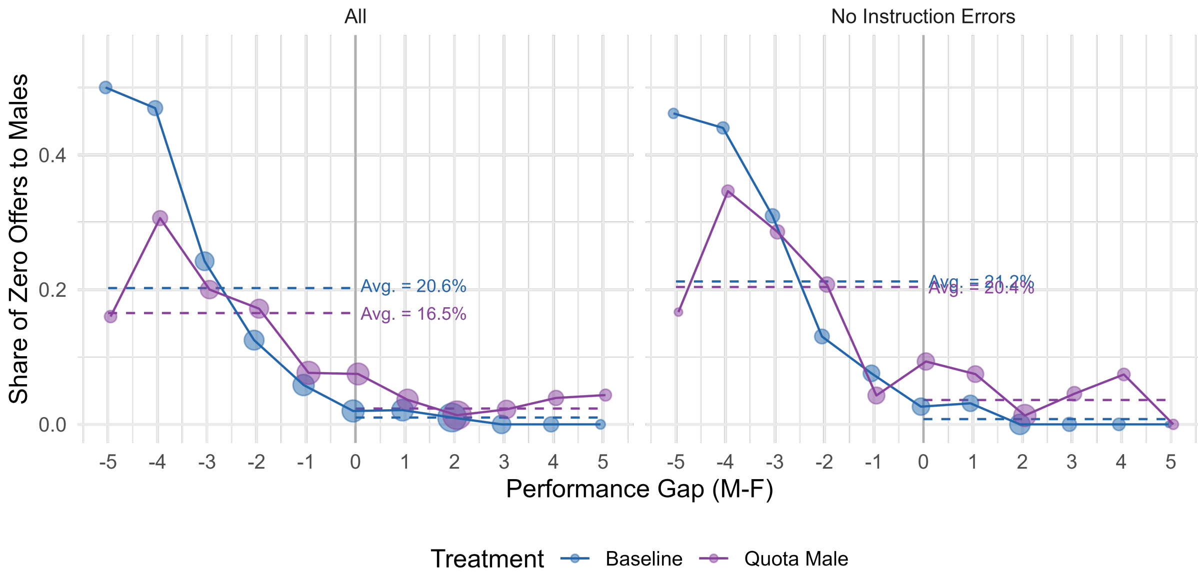
Notes: Circle size represent sample size. Performance gap refers to the difference in performance between candidates in decisions where both have the same gender. Positive values reflect candidates with better performance than the other candidate in the pair. Horizontal lines depict the average likelihood of receiving zero offers. Both panels use only decisions where the entire budget was allocated.

Figure A.11: Relative Performance Wage Gap – All treatments



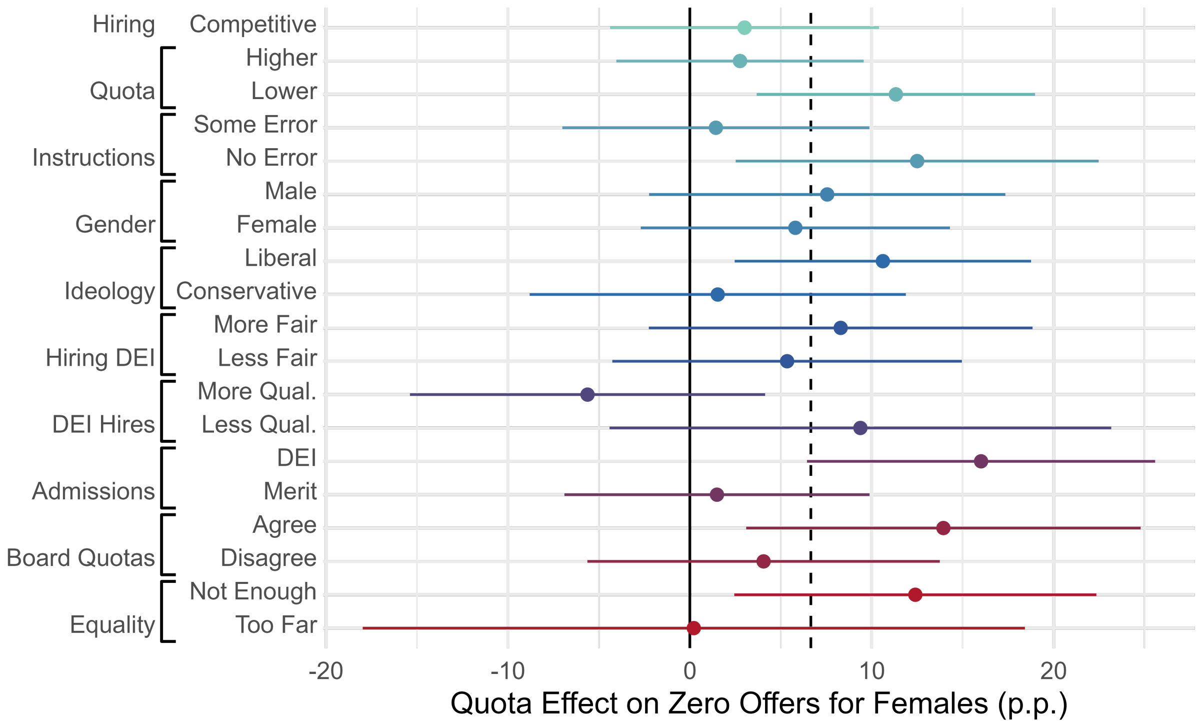
Notes: Error bars correspond to the 95% confidence interval of the difference in means after controlling for demographic characteristics and errors in the instructions. Point estimates include the same controls: dummies for whether participants did not make any errors in the instructions, female gender, age above the median, white ethnicity, and no college degree. We only use decisions where the entire budget was allocated and when the candidates are of different genders.

Figure A.12: Likelihood of No Offer by Performance Gap - *Quota Male* Treatment



Notes: Circle size represent sample size. Performance gap refers to the difference in performance between male candidates and female candidates, in decisions where candidates have different gender. Positive values reflect male candidates with better performance than the female candidate. Horizontal lines depict the average likelihood of receiving zero offers. Both panels only use decisions where the entire budget was allocated.

Figure A.13: Heterogeneity of *Quota Treatment* Effect on Zero Offers



Notes: The vertical dashed line shows the Average Treatment Effect in Part I. Each point corresponds to the coefficient of *Quota Treatment* when regressed on whether a female candidate receives a zero offer, including the usual controls: whether the participant committed any mistake in the instructions, female gender, age above the median, white ethnicity, and no college degree. Error bars correspond to 95% confidence interval of the coefficient. Each regression is estimated by interacting the *Quota Treatment* with a different variable. We first explore different decisions: those with *Competitive Hiring*, while *Quota* interacts the treatment effect with whether the quota candidate has a higher or lower performance than the non-quota female one. Then we focus on heterogeneity by participant characteristics: errors in the *Instructions*, and participant's *Gender* and *Ideology*. Finally, we show heterogeneity by policy views: whether gender diversity makes hiring processes more or less fair (*Hiring DEI*) and the candidates more or less qualified (*DEI Hires*); whether *Admissions* should only consider merit or also DEI, whether there should be gender quotas on the board of directors of firms (*Board Quotas*); and whether the country has gone too far or not far enough in giving women equal opportunities (*Equality*).

Table A.2: Balance Table

Variable	Baseline	Quota	Quota Male
No Instruction Errors, P1	0.44	0.50	0.47
		[0.20]	[0.55]
No Instruction Errors, P2	0.51	0.44	0.45
		[0.20]	[0.24]
All Budget Allocated	0.90	0.92	0.88
		[0.31]	[0.63]
Age	39.01	40.76	39.31
		[0.15]	[0.81]
Male	0.52	0.48	0.48
		[0.43]	[0.49]
White	0.75	0.79	0.62
		[0.38]	[0.01]
No College Education	0.28	0.34	0.11
		[0.18]	[0.00]
Ideology	3.99	3.62	4.01
		[0.07]	[0.91]

Notes: Column *Baseline* shows the mean of each variable. Columns *Quota* and *Quota Male* show, for each variable, its mean in the first row, and in the second row and in brackets the p-value of a t-test of difference in means with the *Baseline* treatment. *No Instruction Errors, P1* and *No Instruction Errors, P2* show the share of participants who made no understanding errors in the instructions for Part 1 and Part 2, respectively. *All Budget Allocated* shows the share of decisions that used the entire available budget to make offers. *Male* and *Age* reflect the demographic composition of the sample, *White* and *No College Education* display the share of participants who identify as racially White and have no college education, respectively. *Ideology* shows the average political alignment of the participants, where 1 reflects very liberal and 7 very conservative views.

A.2 Policy Views Questions

All questions are selected from Gallup polls.

Q1-2. Gender in Hiring Decisions. *When companies consider gender as a factor in hiring decisions in order to increase the gender diversity of the workplace, do you think...*

1. *This makes the overall admissions process of these companies...* [More-Less Fair]
2. *That the candidates who are hired to these companies are...* [More-Less Qualified]

Q3. DEI in college admissions. *Which comes closer to your view about evaluating students for admission into a college or university...*

a. Applicants should be admitted solely on the basis of merit, even if that results in few minority students being admitted.

b. An applicant's racial and ethnic background should be considered to help promote diversity on college campuses, even if that means admitting some minority students who otherwise would not be admitted

Q4. Gender quotas in boards of directors. *Should companies impose gender quotas to increase diversity on their board of directors?* [Agree - Disagree]

Q5. Equal opportunities for women. *When it comes to giving women equal opportunities with men, do you think our country...* [Has gone too far - Has not gone far enough]