Can Temporary Affirmative Action Improve Representation?

Neeraja Gupta
University of Richmond*

July 9, 2024

Abstract

This study examines whether temporary affirmative action can reduce employer biases and improve representation after the policy ends. I conducted an experiment analyzing employer beliefs about employee performance and hiring decisions in two treatments: one with and one without affirmative action. In the control treatment, biases against women were evident in both beliefs and hiring outcomes. However, under temporary affirmative action, I observed lasting improvements in women's representation and employer beliefs about their performance. Employers with stronger initial biases showed the most significant reduction in bias which in-turn explained increased hiring of women.

JEL Codes: D83, D90, J23, J70, J71, J78, M51

Keywords: Learning, Biased Beliefs, Gender Discrimination, Affirmative Action, Hiring Decisions

*Email: neeraja.gupta@richmond.edu. I am deeply grateful to Lise Vesterlund and Richard Van Weelden for their invaluable guidance throughout this project. Additionally, I extend my thanks to Mackenzie Alston, David Huffman, Alex Imas, Ernesto Reuben, Luca Rigotti, Stephanie Wang, Alistair Wilson, Mallory Avery, Juan Campanario, Kelly Hyde, and Marissa Leppers. I am also thankful to seminar participants at New York University Abu Dhabi, Fairfield University, University of Richmond, Skidmore College, Middlebury College, and Haverford College as well as conference participants at Economic Science Association (ESA) 2022 North American Meeting, ESA 2022 Special Meeting, Stanford Institute for Theoretical Economics (Experimental Economics Session), Rising Scholars Conference at Chicago Booth, Behavioral Research in Economics Workshop, and the 3rd International Behavioral Public Policy Conference. Financial support for this project was generously provided by the Behavioral Economics Design Initiative (BEDI) and the Robins School of Business at the University of Richmond. All errors are my own.

1 Introduction

Employers with biased beliefs against a group of workers are less likely to hire candidates from that group (Rooth, 2010; Reuben et al., 2014; Coffman et al., 2021). Since employers only receive feedback about the performance of workers they hire and not the entire pool of applicants, these biased beliefs may be sustained due to hiring decisions that limit belief updating. Consequently, biased beliefs can trigger a self-perpetuating and endogenous under-representation of the group of workers against whom the employers are biased (Lepage, 2021; Coffman et al., 2024). External interventions, such as affirmative action, may help break this pattern. Increased exposure to diverse workers can provide employers with opportunities to correct their biased beliefs and potentially improve long-term representation.

In this paper, I examine an environment where employers hold biased beliefs about women's performance. I investigate whether implementing a temporary affirmative action quota can lead to lasting improvements in women's representation. I analyze how beliefs about women's performance evolve and whether temporary quotas reduce biases after they are removed. Additionally, I explore how changes in beliefs about performance may explain shifts in hiring decisions.

The under-representation of women is a well-documented phenomenon, manifesting in both horizontal and vertical segregation. Women are underrepresented in traditionally male-dominated fields (Beede et al., 2011; Mann and DiPrete, 2013; Savigny, 2014; Goldin, 2015; Leslie et al., 2015; Teele et al., 2018) as well as in managerial positions (Matsa and Miller, 2011; Rhode, 2017; Buckles, 2019). This under-representation can be attributed to various factors, including discriminatory hiring practices.

The economics literature defines three distinct types of discrimination. Taste-based discrimination (Becker, 1957) occurs when employers' prejudices or personal preferences influence their hiring decisions. A model of statistical discrimination (Aigner and Cain, 1977; Phelps, 1972; Arrow, 1971) instead predicts under-representation if a group is expected to perform worse on average. Recently, scholars have introduced the concept of inaccurate statistical discrimination (Fershtman and Gneezy, 2001; Bohren et al., 2019). This model also bases hiring decisions on expected performance but involves employers holding inaccurate beliefs about a group's abilities. Discrimination arises when employers inaccurately perceive one group as less capable than another. For example, if men and women are equally skilled at math but an employer incorrectly believes men are better, women will be less likely to be hired for math-intensive positions.

While affirmative action mechanically improves representation regardless of the underlying discrimination's nature, it is of interest to investigate whether we observe sustained improvements in representation even after the policy ends. For instance, if employers harbor taste-

based discrimination against women, exposure mandated by affirmative action might mitigate distaste against women and lead to lasting improvements in their representation post-policy. However, this mandated exposure could also potentially increase resentment towards women instead. Alternatively, if employers hold inaccurate beliefs affecting hiring decisions, mandatory hiring under affirmative action provides feedback on women's actual performance. This feedback can lead to corrections in inaccurate beliefs, thereby improving representation even after the quota expires.[1]

Through a series of three online experiments, I explore a setting characterized by biased beliefs against women, focusing on the dynamics of hiring decisions and beliefs about performance. The experimental design addresses several critical challenges. Firstly, it enables a causal analysis of how temporary affirmative action influences representation by randomly assigning participants to either environments with no restrictions or those under a temporary affirmative action policy for women. Secondly, the experiments explore a setting where there are no gender differences in actual performance, yet a belief persists that men outperform women. Lastly, the study utilizes reliable measures of beliefs that are consistent over time, essential for understanding how learning mechanisms operate in this context. This carefully crafted approach aims to overcome analytical hurdles and contribute nuanced findings regarding the impact of temporary affirmative action policies.

In the first experiment, participants take on the role of workers and answer trivia questions from several male-type trivia categories (Coffman, 2014; Bordalo et al., 2016; Coffman et al., 2019). Experiment 2 elicits beliefs on gender difference in performance of participants in the previous experiment. Of key interest here is finding a trivia category where evaluators hold biased beliefs on the gender gap in performance. The first two experiments identify sports as the trivia category where there are no gender differences in average performance but where men are believed to outperform women. In the third and final experiment of the study, the performance of workers in the sports trivia quiz forms the foundation for employer beliefs and subsequent hiring decisions.

Participants in the employer experiment review four randomly selected resumes from a gender-balanced group of workers from experiment 1. Employers' beliefs are elicited about expected performance of all presented worker resumes on the sport trivia quiz. Subsequently, they are tasked with hiring two out of the four presented workers for their performance on the sports trivia quiz. After making their hiring decisions, employers receive feedback on the actual performance of both hired employees. This process repeats for six rounds, each with different sets of workers, allowing us to observe how employer beliefs and hiring decisions

[1] A temporary affirmative action policy can also have lasting effects under accurate statistical discrimination by creating economic incentives for skill enhancement among underrepresented groups, contingent on specific initial parametric conditions (Coate and Loury, 1993).

evolve with feedback. The experiment employs two between-subject treatments: a control treatment where employers have no hiring restrictions, and a temporary affirmative action treatment. In the temporary affirmative action treatment, the first three rounds enforce a quota policy for women, requiring at least one hired workers to be a woman. This policy is lifted in the last three rounds, allowing for unrestricted hiring decisions.

In the absence of affirmative action, women were 28 percentage points less likely to be hired than men, a disparity that persisted and stood at 25 percentage points in the final three rounds. By design, affirmative action improves representation of women while in effect. Importantly, this positive effect continued after the policy ended. In the last three rounds of the temporary affirmative action treatment, women were only 16 percentage points less likely to be hired than men, a substantial reduction compared to the male advantage observed in the control condition. These findings highlight a lasting and significant impact of the temporary affirmative action policy on improving the representation of women in hiring decisions.

Next, we examine beliefs about performance and find that in the first round men are perceived to have an 18 percentage point advantage in likelihood of success in the control condition, which declines to 11 percentage points by the final three rounds. In contrast, in the temporary affirmative action treatment, employers initially exhibit a similar 18 percentage point perceived advantage for men in the first round, which decreases to 9 percentage points by the last three rounds. The reduction in gender bias in beliefs progresses statistically significantly faster due to a temporary affirmative action.

To link observed changes in beliefs with changes in hiring decisions, we delve deeper into the data. Interestingly, within our sample, there exists a large subset of employers who exhibit less biased beliefs and are inherently less inclined to discriminate against women from the outset. For these employers, quotas are not constraining, and their exposure to women workers is not restricted. Consequently, we anticipate no distinct impact of the temporary affirmative action treatment on shifts in beliefs or how beliefs correlate with changes in hiring decisions. Indeed, our findings confirm this expectation.

On the other hand, we find that employers who initially exhibit higher levels of bias and are more inclined to discriminate against women experience a significantly larger reduction in gender bias in beliefs due to the temporary affirmative action treatment, even after quotas are lifted. Within this subset of employers, we also observe a positive and statistically significant effect of the treatment on the representation of women in hiring. Furthermore, this positive effect on women's hiring is partly explained by changes in beliefs regarding their performance. This subgroup of employers, characterized by higher initial bias and discrimination tendencies, provides a stark comparison to explore the dynamics of inaccurate statistical discrimination under a temporary affirmative action policy. The study's findings suggest that exposure under

a temporary quota can lead to a reduction in gender bias in beliefs, subsequently fostering a lasting improvement in the representation of women.

The rest of the paper is organized as follows: Section 2 reviews the relevant literature and highlights the contributions of this study. Section 3 provides a detailed overview of the experiment design and outlines the primary hypotheses derived from the employer experiment. Section 4 presents a summary of the main findings. Lastly, Section 5 presents conclusions and discusses the policy implications of the study's results.

2 Related Literature and Contributions

Findings from this paper contribute to the broad literature on lasting effects of temporary affirmative action. Existing economic theory (Coate and Loury, 1993) and experimental research (Dianat et al., 2022) on long-term effects of temporary affirmative action have focused exclusively on studying accurate statistical discrimination wherein employer beliefs are accurate. This study introduces a novel perspective by examining the impact of temporary affirmative action in a context marked by inaccurate beliefs. Specifically, it explores employer responses using a real-effort task where gender does not affect actual performance, yet women are perceived as less capable compared to men.

Previous studies have faced challenges in establishing whether the actual performance of over- and underrepresented groups is identical while only beliefs are biased. This ambiguity can complicate efforts to attribute changes in beliefs about performance to lasting changes in representation. For instance, Beaman et al. (2009), in a field experiment, found that quotas ultimately increase the likelihood of women participating in and winning political elections even after they are removed. However, they also observed that survey measures of leader effectiveness initially worsened after exposure to a female leader but improved over time. They argue that the underlying discrimination may have been accurately statistical, creating initial conditions conducive for less effective female leaders to invest in skill enhancement and improve their effectiveness over time, while the belief assessments of participants remained consistent throughout the experiment.[2] Similarly, Paola et al. (2010) also observe a lasting positive effect on the representation of women in politics due to temporary quotas but could not pinpoint learning about ability as the underlying mechanism. In contrast, Bhavnani (2017) finds that the positive effects of an affirmative action policy diminish once the policy is lifted and attributes this to the lower possibly ability of the underrepresented group. To overcome this challenges, the controlled environment of an online experiment in this study was chosen

[2]Through an additional experiment using hypothetical leaders and vignettes, they demonstrate that beliefs can be malleable, and attitudes can change through exposure to female leaders.

to mitigate the potential for accurate statistical discrimination to arise.

Many other notable studies also document longer term effects of temporary affirmative action across diverse contexts, yielding mixed results (Dickson, 2006; Miller and Segal, 2012; Antonovics and Sander, 2013; Antonovics and Backes, 2014; Deininger et al., 2015; Yagan, 2016; Hill, 2017; Miller, 2017; Banerjee et al., 2018). This paper contributes to this literature by identifying beliefs as the mechanism driving changes in hiring decisions. Understanding how biased beliefs evolve and influence hiring outcomes is essential for designing effective policies that promote equitable representation.

3 Experiment Design

The study comprises three experiments. The first experiment, detailed in Section 3.1, gathers performance data on several trivia quizzes and demographic characteristics of participants. The second experiment, detailed in Section 3.2, elicits beliefs about the performance of workers with different demographic characteristics across the different trivia quiz categories. These experiments help identify a trivia quiz category where, despite no actual gender difference in performance, men are perceived to be better performers than women, thus establishing a setting with biased beliefs. The third and primary experiment, detailed in Section 3.3, examines employer beliefs and hiring decisions across multiple rounds using randomly selected workers from Experiment 1.

It's important to note that the worker's performance measure in this study is simplified to a binary indicator of success or failure. A worker is deemed successful if their score in a trivia category is greater than or equal to the the median score within that category. For every belief elicitation, performance is measured as the likelihood of a worker being successful in a particular trivia category. This binary approach streamlines the belief elicitation process into a single probability measure, abstracting away from distributional complexities. All experiments are programmed using Qualtrics and run on Prolific.[3]

3.1 Experiment 1: Workers

The objective of this initial experiment is to gather performance data on a real-effort task where men and women have an equal likelihood of success. The task selected here is a trivia quiz from three male-type trivia quiz categories: sports, business, and video games. Previous

[3]In a comparative study among physical laboratories, Amazon's Mechanical Turk, Prolific, and online laboratories, Prolific has demonstrated superiority in terms of eliciting data on the extensive margin, especially in scenarios not involving social tension among participants, due to its minimal noise and significantly lower data collection costs (Gupta et al., 2021).

research suggests that trivia quizzes in male-typed categories can trigger biased beliefs due to stereotypes linked to these domains (Coffman, 2014; Bordalo et al., 2016; Coffman et al., 2019).

In this experiment, a gender-balanced pool of 50 participants is given 5 minutes to complete 30 multiple-choice questions, with 10 questions from each trivia category.[4] Upon completion, participants proceed to a demographic survey that collects data on 12 personal attributes, which are used in subsequent experiments to identify them anonymously.[5] Participants receive a $1.50 completion fee along with a $0.10 bonus payment for each correct answer on the quiz.[6,7]

We find that the gender difference in performance is small and statistically insignificant for the trivia categories of sports and business.[8] While the gender difference in the video games category is also statistically insignificant, we exclude it due to its relatively larger magnitude (see Appendix 5 for details).[9] Therefore, the next experiment proceeds with the trivia categories of sports and business.[10]

3.2 Experiment 2: Evaluators

To understand how perceptions of success likelihood change with demographic characteristics for each trivia quiz category, we recruited a new set of participants on Prolific to serve as evaluators. Participants were independently recruited for evaluations related to the business and sports trivia quizzes.[11]

For these evaluations, the 12 demographic characteristics were divided into two groups,

[4]All questions appear on the same page, and their order, as well as the order of the options within each question, is randomized.

[5]The demographic survey includes attributes commonly found on resumes, such as age, years of schooling, employment status, gender, geographical location (deduced from the time zone of residence), volunteer work in the past 5 years, number of languages spoken, and arbitrary characteristics including height, voter registration status in the United States, smoking status, preference for cats or dogs, and the operating system of their phones.

[6]Detailed instructions, sample trivia quiz questions, and screenshots of the experiment can be found in the online appendix: https://tinyurl.com/4yaumhjw

[7]The worker experiment was conducted on Prolific during the summer of 2021, with approval from the University of Pittsburgh IRB (STUDY19020366).

[8]Performance is measured as the likelihood of a worker being successful in a particular trivia category where success is defined as getting a score greater than or equal to the median. The median scores are 6 for business, 9 for sports, and 7 for video games.

[9]Given the levels and standard deviations, the gender difference for video games is likely to become statistically significant with a small increase in the sample size. A simple power calculation in STATA for a test of means confirms this conjecture.

[10]The likelihood of success in business varied significantly by age, number of languages spoken, smoking behavior, and phone operating system. In sports, the likelihood of success varied by age, employment status, and pet preference (cats or dogs). This indicates that both common resume attributes and arbitrary characteristics can explain the likelihood of success in these trivia categories.

[11]The evaluator experiment was also conducted during the summer of 2021, with approval from the University of Pittsburgh IRB (STUDY19020366).

Table 1: Demographic characteristic groups for evaluator experiment

Demographic Characteristic	Group 1 (1)	Group 2 (2)
Age	27 yrs or below	Above 27 yrs
Gender	Male	Female
Height	5 ft 6 in or less	More than 5 ft 6 in
Time Zone	Eastern	Not eastern
Education as Years of Schooling	13.5 yrs or below	Above 13.5 yrs
Employment Status	Employed	Not employed
Volunteer Work	Yes	No
Voter Registration	Yes	No
Smoker	Yes	No
No. of Spoken Languages	1	More than 1
Cat/Dog Person	Cat	Dog
Phone OS	Android	iOS

Notes - This table presents the set of 24 demographic characteristic groups for which beliefs were elicited in the evaluator experiment. For each demographic characteristic the two groups are defined based on a median worker's response from the worker experiment.

generating 24 demographic groups, as shown in Table 1. For example, categorical variables like time zone of residence were divided into "eastern" vs. "not eastern," while continuous variables such as age were divided based on the median response of the participants.

Evaluator beliefs about performance are elicited for a randomly selected worker from the first experiment within each demographic group. For each elicitation, evaluators indicate their perception of the likelihood that a randomly selected worker from the respective demographic group in the first experiment was successful. As before, success is defined as achieving a score equal to or greater than the median score in the specific trivia quiz.[12] For example, in the sports trivia category and gender elicitation, evaluators assess the likelihood that a randomly selected man (or woman) achieved a score of 9 or higher.

Evaluators receive a $2 completion fee and have the opportunity to earn a bonus payment of up to $2. For the bonus payment, two of their submitted guesses are randomly selected, and they can earn $1 for each for the accuracy of their guess incentivized using the minimum information binarized scoring rule (Hossain and Okui, 2013; Danz et al., 2022). Evaluators are advised that truthful reporting maximizes their chances of earning the bonus. The experiment

[12]The order of presentation of the 12 characteristics to evaluators is randomized, and within each characteristic, the order of the two groups is also randomized.

concludes with a brief exit survey and payment information provided to the evaluators.[13,14]

A gender-balanced pool of 80 evaluators for the business trivia category and 61 for the sports trivia category was recruited.[15] Two main findings emerge from this experiment. Firstly, there is no gender difference in perceived likelihood of success for the business trivia category, suggesting it does not invoke biased beliefs against women and is therefore eliminated at this stage. Secondly, a significant gender difference in perceived likelihood of success exists for the sports trivia category, where men are perceived to outperform women (details in Appendix A).[16] Therefore, the employer experiment proceeds with the sports trivia quiz as the task for worker performance, as there are no gender differences in average success likelihood, yet men are on average perceived to perform better on this task.

3.3 Experiment 3: Employers

Experiment 3 focuses on eliciting performance evaluations and hiring decisions, serving as the primary experiment of interest. Participants from the first experiment constitute the pool of workers for whom hiring decisions are solicited. Resumes are created based on workers' demographic characteristics: years of schooling (education level), employment status, geographical location (time zone of residence), number of spoken languages, and gender (male or female). Employer decisions are based on workers' performance in the sports trivia quiz. The criterion of interest remains the probability of achieving a score of 9 or higher, which corresponds to the median score from experiment 1. The employer experiment design and hypothesis (as described in the next section 3.4) are pre-registered with the AEA RCT Registry: AEARCTR-0013391.

The employer experiment comprises six rounds, each featuring two decision stages: evaluation and hiring. Each round concludes with feedback based on the employer decisions. Participants receive a completion fee of $2 and have the opportunity to earn a bonus payment of up to $3 from one randomly selected decision stage.

In each round, employers are presented four randomly selected worker resumes – two men and two women.[17] Employers first evaluate four worker resumes, rating on a scale of 0 to 100

[13]Attention checks are included in this study. A post hoc assessment verifies response data from these checks and identifies any outliers in time spent on the instructions page. No participants are excluded from analysis based on this assessment.

[14]Detailed instructions and screenshots of the experiment can be found in the online appendix: https://tinyurl.com/4yaumhjw

[15]Power calculations are parameterized using Coffman et al. (2019) and available upon request.

[16]Perceived likelihood of success in sports varies across all demographic characteristics except time zone and number of spoken languages, indicating both common resume attributes and arbitrary characteristics explain perceived success likelihood. These factors need control in subsequent analyses.

[17]To randomize resumes, 24 out of 50 (12 men, 12 women) are selected randomly and split into 6 groups of 2 men and 2 women each. These groups are assigned sequentially across 6 rounds for employer evaluations: the

the likelihood that each worker scored 9 or higher in the sports trivia quiz.[18] If this stage is chosen for bonus payment, one worker resume is randomly selected, and employers receive a $3 bonus for the accuracy of their estimate for that worker. The evaluation process incentivizes accurate reporting using the minimum information binarized scoring rule, emphasizing that truthful reporting enhances their chances of receiving the bonus (Hossain and Okui, 2013; Danz et al., 2022). Participants seeking further details can access a non-mathematical explanation of the payment rule at the experiment's end (Wilson and Vespa, 2018).[19]

Employers then proceed to make hiring decisions from the same pool of four workers. They must select two workers and rank them as their 1st and 2nd choices for hiring.[20] If this stage is chosen for bonus payment, employers earn $2 if their 1st choice is successful on the sports trivia quiz, $1 if their 2nd choice is successful, and $0 otherwise. The round concludes with feedback, informing employers whether their two hired workers were successful or not. This design mirrors real-world scenarios where employers receive performance updates only for the workers they hire, not for those they do not select.

The employer experiment includes two between-subject treatments: a control treatment and a temporary affirmative action treatment. In the affirmative action treatment, a soft quota requires employers to hire at least one woman in the first three rounds.[21] The quota is lifted in the last three rounds, and a comprehension check in round 4 ensures participants in this treatment are aware of the policy change. In contrast, employers in the control treatment do not face any affirmative action policies regarding women throughout the experiment, nor are they informed about the possibility of such policies.[22]

A gender-balanced pool of 997 participants from Prolific were recruited for this experiment.[23] Participants were randomly assigned to one of two treatments upon entering the study. To ensure data quality, participants faced multiple comprehension and attention checks throughout the study. They were allowed multiple attempts for comprehension checks, and the

first group starts in round 1, the second in round 2, and so on. From this selection, six employer assignments are created to ensure balance across rounds. This involves permuting the sequence of groups for each employer: Group 1 starts as the first sequence for Employer 1, then shifts to the second sequence for Employer 2, and so forth. This process is repeated nine times to generate 54 unique sets of 24 resumes. These assignments are structured into a matrix and uploaded to Qualtrics using JavaScript. The same set of 54 assignments is used consistently across all employers' experimental sessions.

[18] Employers are aware that correct answers range from 2 to 10.

[19] Participants with further queries could contact me via email; no such requests were received.

[20] Their evaluation ratings for each worker are displayed alongside the resumes to aid in the hiring decision.

[21] This policy is enforced such that employers cannot proceed if they do not hire at least one woman.

[22] Following Niederle et al. (2013) the knowledge of affirmative action is considered to be a part of the treatment.

[23] The employer experiment was pre-registered and conducted in summer 2024, with approval from the University of Richmond IRB (URIRB240416). A pilot was conducted in summer 2022; results from which were presented and circulated as the author's job market paper and are available upon request.

Table 2: Employer experiment summary statistics

		(1) Control	(2) TempAA	(3) Mean Diff.
Panel A: Employer Demographic Characteristics				
Age (in years)		38.86	39.41	-0.55
Gender	Men	0.49	0.47	0.02
	Women	0.50	0.51	-0.01
	Other	0.01	0.02	-0.01
Years of Schooling	16 years or less	0.82	0.83	-0.01
(level of education)	More than 16 years	0.18	0.17	0.01
Employment status	Currently Employed	0.71	0.73	-0.01
	Not Currently Employed	0.29	0.27	0.01
Sample Size		**498**	**499**	
Panel B: Presented Resume Characteristics				
Gender	Men	0.50	0.05	0.00
	Women	0.5	0.5	0.00
Years of Schooling	13.5 years or less	0.47	0.47	0.00
(level of education)	More than 13.5 years	0.53	0.53	0.00
Employment status	Currently Employed	0.66	0.66	0.00
	Not Currently Employed	0.34	0.34	0.00
Time Zone of Residence	Eastern	0.45	0.45	0.00
(Geographical Location)	Central/Mountain/Pacific	0.55	0.55	0.00
Number of Spoken Languages	1	0.66	0.66	0.00
	More than 1	0.34	0.34	0.00
Average Proportion Successful	Men	0.49	0.49	0.00
	Women	0.52	0.52	0.00
	Overall	0.50	0.50	0.00

Notes - *** p<0.01, ** p<0.05, * p<0.1, for a two-tailed test of mean difference between the control and the temporary affirmative action treatment values against a null of 0. This table presents the summary statistics from the employer experiment. Panel A shows the employer characteristics and Panel B shows the resume characteristics which were shown to the employers. Years of schooling is indicative of the level of education and the categories are split by the median value . Currently employed subjects were either employed full-time or part-time or self-employed. Not currently employed subjects were either unemployed or students or retired or homemakers or unable to work.

number of attempts was monitored to identify outliers.[24] Participants had only one attempt for attention checks, and failure resulted in immediate dismissal.[25]

[24]No participants were identified to be excluded based on their number of attempts for comprehension checks.

[25]Detailed instructions and screenshots of the experiment can be found in the online appendix: https://tinyurl.com/4yaumhjw

The summary statistics for the employer sample are detailed in Table 2. Panel A displays the demographic characteristics of employers in both the control and temporary affirmative action treatments. The sample consists of 499 employers in the temporary affirmative action treatment and 498 in the control, with balanced employer characteristics across treatments. Panel B summarizes the characteristics of the resumes presented to employers. Since the same randomly selected resumes are used across treatments, resume characteristics are naturally balanced. Interestingly, within the set of resumes, 52% of women and 49% of men are successful in the sports trivia quiz.

3.4 Hypotheses

In the employer experimental setup, gender discrimination in hiring is defined as a systematic bias against hiring women. We can analyze the nature of this discrimination based on the decision-making process. For instance, if hiring decisions are primarily influenced by perceived likelihood of success, then any observed discrimination is likely statistical and rooted in beliefs. However, if decisions are influenced by an inherent bias against women independent of beliefs, the discrimination is taste-based. Given the experimental design where women are perceived, on average, to perform less well than men, observed gender discrimination in hiring can either be statistically inaccurate, or taste-based, or a combination of both.

If the nature of underlying discrimination is in parts inaccurately statistical, then in the absence of affirmative action for women, employers are likely to have limited exposure to women workers. With feedback provided at the end of each round about hired employees, beliefs about men may converge towards their true value over time. However, the same cannot be said for gender differences in beliefs due to limited exposure to women workers, leading to potential persistence of discrimination against women in the control treatment of the experiment.

In contrast, a temporary affirmative action intervention exposes employers to more women workers. This exposure allows employers to learn not only about men but also about women, potentially leading to a convergence of gender differences in beliefs towards a true value, set at 0 by design. The impact of affirmative action for women may extend beyond the policy instance, potentially producing lasting improvements in the representation of women based on beliefs. These arguments lead to the three hypotheses for this experiment.

In environments where beliefs about performance are biased against women relative to men:

H1: Temporary affirmative action for women improves representation of women even after affirmative action is lifted.

H2: Temporary affirmative action for women reduces the bias in beliefs about performance of

women relative to men.

H3: Beliefs about performance help explain the lasting improvement in representation of women due to a temporary affirmative action policy.

An auxiliary hypothesis is a possibility of heterogeneous effects based on the initial bias in hiring and beliefs. For employers who are less biased at the outset, quotas are not restrictive, and their exposure to women workers is not limited. Therefore, we expect that the temporary affirmative action treatment will not significantly influence shifts in their beliefs or how these beliefs correlate with changes in hiring decisions. In contrast, employers who initially demonstrate higher levels of bias and are more likely to discriminate against women are anticipated to experience a greater reduction in gender bias in beliefs as a result of the temporary affirmative action treatment. Moreover, we hypothesize that shifts in beliefs will explain changes in hiring decisions among this subgroup of employers.

4 Results

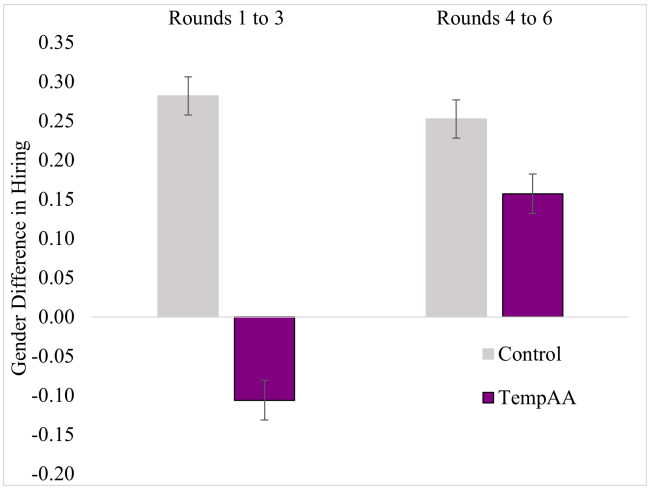
This section presents findings from the employer experiment. I begin by examining the hiring decisions of employers in the temporary affirmative action treatment compared to the control treatment in Section 4.1. Next, I analyze the dynamics of beliefs regarding performance in both experimental groups in Section 4.2. In Section 4.3, I discuss beliefs as the mechanism influencing changes in hiring decisions. Additional analysis in Section 4.4 explores gender differences in the distribution of hired employees as 1st and 2nd preferred candidates for hiring.

4.1 Evolution of Hiring Decisions

Results from the hiring decisions are summarized in Figure 1. In the control treatment without affirmative action, men are hired 28 percentage points more often than women. This difference reduces slightly to 25 percentage points in the last three rounds, although the difference is not statistically significant. In the temporary affirmative action treatment, women are hired 11 percentage points more often than men during the first three rounds when quotas are enforced. When the quotas are removed in the last three rounds, men still have an advantage, but at 16 percentage points, it is considerably smaller than in the control treatment ($p < 0.001$). These findings provide preliminary evidence supporting the first hypothesis that temporary affirmative action enhances the representation of women even after the policy is discontinued.

To confirm these findings we use a regression framework and model the hiring decisions as

Figure 1: Difference in Proportions of Men Hired vs. Women Hired



Notes - This figure shows mean gender difference in proportions of men hired vs. women hired. The left panel shows the first three rounds where there is a quota for women in effect under the temporary affirmative action treatment, and the right panel compares the two experimental treatments for the last three rounds when affirmative action is removed. Error bars correspond to 95% confidence intervals.

follows

$$\pi_{ijr} = x_{ir}\beta + u_{ir} + \epsilon_{ijr} \tag{1}$$

where π_{ijr} represents employer i's profit from hiring worker j in round r, with $j \in J = \{1, 2, 3, 4\}$ and $r \in \{1, ..., 6\}$. Further, x_{ijr} represents the row vector of covariates and β is the column vector of coefficients. The u_{ir} represents the group level heterogeneity where a group represents a single round where an employer makes decisions over 4 resumes, and ϵ_{ijr} is the observation-level error term. Employer i will hire the worker j in any given round if hiring that worker maximizes their profit:

$$Pr(y_{ikr} = 1) = Pr\{max(\pi_{ikr}) = \pi_{ijr}\} \forall k \in J \tag{2}$$

Assuming a standard Type I extreme value for ϵ_{ijr}, this gives rise to the following choice model where $j = 0$ is the baseline outcome:

$$Pr(y_{ijr}|x_{ijr}, \beta) = \begin{cases} \frac{1}{1+exp(x_{ijr}\beta)} & \text{if } j = 0 \\ \frac{exp(x_{ijr}\beta)}{1+exp(x_{ijr}\beta)} & \text{if } j = 1 \end{cases} \tag{3}$$

Table 3 presents the results from a conditional fixed effects logit model (eq. 3) using data

Table 3: Hiring - conditional fixed effects Logit regression

Dependent Variable:	Rounds 4 to 6	
	I(Hired)	
	(1)	(2)
I(Female)	-0.758***	-0.728***
	(0.0450)	(0.0500)
I(TempAA)*I(Female)	0.287***	0.310***
	(0.0635)	(0.0658)
N	11964	11964
Worker controls		Yes

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents conditional fixed logit regression results from hiring decisions where a group comprises a single round where an employer makes decisions over 4 resumes. Dependent variable is an indicator variable =1 when a worker is hired. Worker controls include demographic characteristics presented on their resume - employment status, education, number of spoken languages, and time zone of residence. All options within each resume characteristic are aggregated in two groups characterized by the median worker's characteristic from the worker experiment. Standard errors are shown in parentheses.

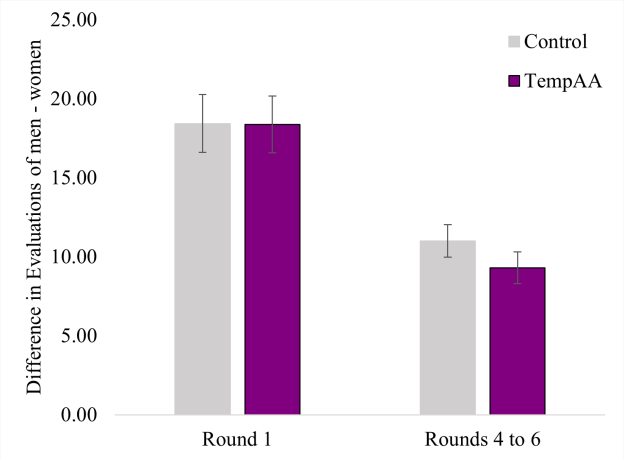
from rounds 4 to 6, after the quotas for women were lifted in the temporary affirmative action treatment. The main explanatory variables in this model include indicator variables for female workers ($I(Female)$) and its interaction with the indicator for the temporary affirmative action treatment ($I(TempAA)$).[26]

In the control treatment without affirmative action, the log odds of women being hired (vs. not) are expected to be 0.758 lower than men in rounds 4 to 6 ($p < 0.001$). In the temporary affirmative action treatment, after quotas for women are removed, the log odds of women being hired remain 0.471 lower than men. However, the temporary affirmative action treatment shows a significant improvement, increasing the log odds of women being hired by 0.287 ($p < 0.001$). These results hold robustly even after controlling for worker characteristics presented on resumes.

In summary, this analysis reveals that women are significantly less likely to be hired relative to men in the absence of affirmative action. Quotas effectively enhance the representation of women, and these positive effects persist even after the policy is terminated. Temporary affirmative action thus continues to promote the hiring of women even after its removal.

[26]The experiment does not involve clustered sampling, and there is no evidence of randomization failure between the control and temporary affirmative action treatment. Thus, clustering of standard errors is considered unnecessary as per recommendations by Abadie et al. (2022).

Figure 2: Difference in Evaluations of Men and Women



Notes - This figure shows mean gender difference in evaluations of men and women. The left panel represents round 1 evaluations, and the right panel compares the two experimental treatments for the last three rounds where quota is removed under the temporary affirmative action treatment. Error bars correspond to 95% confidence intervals.

4.2 Evolution of Beliefs About Performance

With the ability to choose which candidate to hire and therefore receive feedback about, we find that in rounds 1, 2, and 3 of the control treatment, approximately 36% of observed signals are about women employees. In contrast, in the temporary affirmative action treatment, about 55% of signals are about women employees during the same rounds. This represents a 53% increase in signals about women due to affirmative action. The critical question then arises: does this increased exposure to women lead employers to recognize and correct their inaccurate beliefs, resulting in belief updating?

Figure 2 illustrates the raw mean gender difference in evaluations, separated by treatment conditions— initially for round 1 alone and then collectively for rounds 4 to 6. At the outset in round 1 of the control treatment, men are perceived to be 18 percentage points more likely to succeed than women. This gender bias decreases to 11 percentage points in the final three rounds. Similarly, in the temporary affirmative action treatment, employers initially believe men are 18 percentage points more likely to succeed than women. This similarity between treatments is expected, as the round 1 belief elicitation occurs before participants are informed about the hiring environments. Once quotas are removed in the temporary affirmative action treatment, the gender bias diminishes to 9 percentage points, a statistically significant difference from the belief held in the last three rounds of the control treatment ($p = 0.021$). These findings provide initial evidence supporting the second hypothesis: temporary affirmative action reduces gender bias in perceptions of women's performance relative to men's.

Table 4: Evaluations - OLS with fixed effects

Dependent Variable:	Rounds 4 to 6	
	Evaluation	
	(1)	(2)
I(Female)	-11.01***	-10.58***
	(0.388)	(0.411)
I(TempAA)*I(Female)	1.701***	1.712***
	(0.549)	(0.541)
Constant	66.08***	61.55***
	(0.194)	(0.394)
N	11964	11964
Worker controls		Yes

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents OLS regression results on evolution of beliefs about performance elicited as evaluations on a scale of 0 to 100. The estimation uses a fixed effects model where each group comprises a single round where an employer makes decisions over 4 resumes. Worker controls include demographic characteristics presented on their resume - employment status, education, number of spoken languages, and time zone of residence. All options within each resume characteristic are aggregated in two groups characterized by the median worker's characteristic from the worker experiment. Standard errors are shown in parentheses.

To confirm these preliminary findings, the evaluations are modeled for regression analysis as follows:

$$Evaluation_{ijr} = x_{ir}\beta + u_{ijr} + \epsilon_{ijr} \tag{4}$$

where $Evaluation_{ijr}$ represents employer i's evaluation of jth worker in round r, with $j \in \{1, ..., 4\}$ and $r \in \{1, ..., 6\}$. The variable x_{ijr} denotes the row vector of covariates, β is the column vector of coefficients, u_{ir} is the group level heterogeneity (where a group comprises a single round in which an employer makes evaluation decisions over 4 resumes), and ϵ_{ijr} is the observation-level error term. The main regressors include indicator variables for a female worker and their interaction with the temporary affirmative action treatment indicator. Estimation results are shown in Table 4 using data from rounds 4, 5, and 6 and, employing a fixed effects OLS model.

In the control treatment, where employers are free to make hiring choices without any affirmative action constraints, women are perceived to be 11 percentage points less likely to succeed than men. However, when temporary affirmative action policies are implemented, this bias in beliefs is reduced. Specifically, under the temporary affirmative action treatment, women are believed to be about 9 percentage points less likely to succeed than men. The exposure to more women workers due to the temporary affirmative action policy results in a

reduction of the gender bias in beliefs by 1.7 percentage points. This reduction is statistically significant, with a p-value of 0.002. These findings suggest that temporary affirmative action can effectively reduce the gender bias in beliefs about performance, thereby supporting the second hypothesis that increased exposure to women workers can lead to more accurate and less biased evaluations.

4.2.1 Heterogeneous Effects on Evolution of Beliefs

Interestingly, many employers in the sample do not hold biased beliefs and hire women even without any hiring restrictions. The conjecture is that affirmative action, by increasing exposure to underrepresented women, compels employers to gather more feedback about female employees, thereby reducing gender bias in beliefs. To test this hypothesis in the starkest possible comparison, I divide the sample into two subgroups and explore heterogeneous effects: one that is more biased and discriminatory against women, and another that is less biased and discriminatory.

Under the temporary affirmative action treatment, the requirement to hire at least one woman begins from the first round. Therefore, it is challenging to classify employers in groups as more or less likely to discriminate against women initially. To address this, I use the gender difference in evaluations as a measure to predict the probability of hiring two men in round 1, since belief elicitation in both treatments occurs before any differences in hiring environments are introduced.

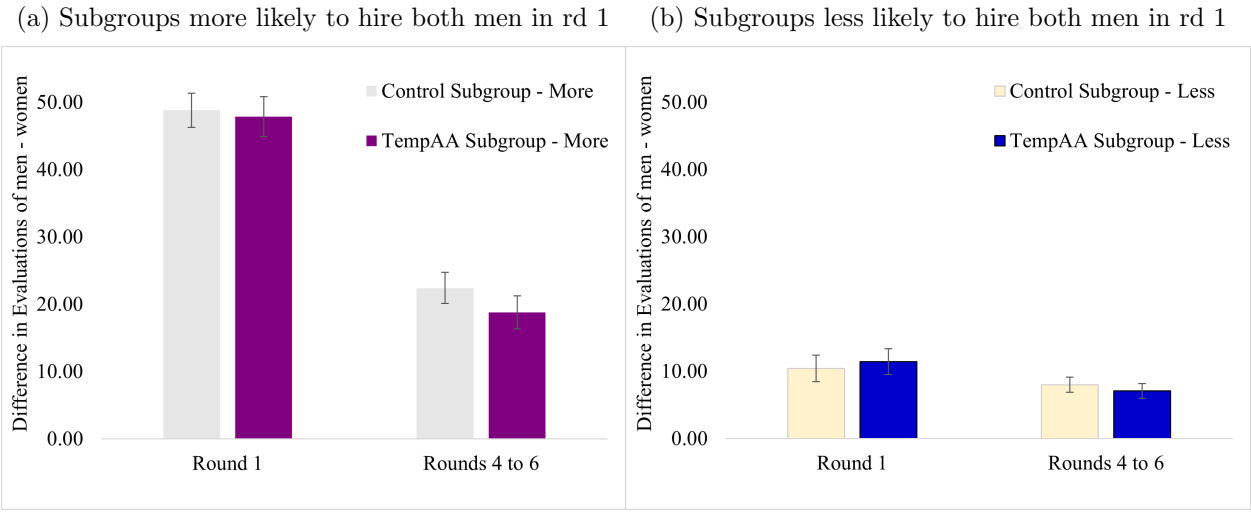
I begin by employing data from the control treatment to estimate a logit regression, where the dependent variable indicates if two men were hired in round 1, and the explanatory variable is the gender difference in beliefs about worker performance in round 1. Using this model, I predict the probability of hiring two men for the control group of employers and establish a cutoff based on the top 20% of employers most likely to hire two men in round 1 (0.576).[27] This threshold is then used to classify two subgroups within the control group: "More" (those with a predicted probability greater than 0.576) and"" Less" (those with a predicted probability lesser).

To achieve a similar classification for employers in the temporary affirmative action treatment, I utilize the previously estimated coefficients to predict the probability of hiring two men in round 1. Subsequently, I classify the two subgroups within the temporary affirmative action treatment based on the predicted probability of hiring two men in round 1 around the cutoff point of 0.576.

Figure 3 Panel (a) illustrates employers from the "More" subgroup, while panel (b) repre-

[27] This cutoff is chosen because approximately 20% of employers in the control group hired two men in round 1, but I demonstrate that all subsequent results are robust to variations in this cutoff.

Figure 3: Difference in Evaluations of Men and Women - Heterogeneous Effects



Notes - This figure shows mean gender difference in evaluations of men and women. Sub-figure (a) shows the gender difference by experimental conditions for the subgroup of employers more likely to hire both men in round 1 and sub-figure (b) show this difference for employers less likely to hire both men. The cutoff to divide the two subgroups is at 58% which leaves about 20% of the sample for sub-figure (a). The left panel in both figures represents round 1 evaluations, and the right panel compares the two experimental treatments for the last three rounds where quota is removed under the temporary affirmative action treatment. Both figures exclude employers who disregard information about performance of workers as identified by the final stage of the experiment to minimize noise. Error bars correspond to 95% confidence intervals.

sents the "Less" subgroup. Among employers more inclined to hire both men—indicating a higher likelihood of discrimination against women—we observe that in the control treatment, men are perceived to be 49 percentage points more likely to succeed than women initially. This bias decreases to 22 percentage points in rounds 4 to 6. In contrast, under the temporary affirmative action treatment, beliefs evolve significantly faster, starting with a gender bias of 48 percentage points and decreasing to 19 percentage points in rounds 4 to 6 after quotas are removed. Conversely, in the "Less" subgroup (panel b), where employers are less inclined to discriminate against women, there is initially a small gender bias that remains relatively stable throughout the experiment in both treatments.

Regression estimates confirming the observed patterns are presented in Table 5. Columns (1)-(2) examine the effect of the temporary affirmative action treatment on beliefs within the "More" subgroup of employers, while columns (3)-(4) focus on the "Less" subgroup.

In the "More" subgroup, without affirmative action in the control treatment, women are perceived to be 22 percentage points less likely to succeed than men in rounds 4 to 6. However, there is an approximately 4 percentage points improvement in beliefs about success of women due to a temporary affirmative action policy ($p = 0.013$). Conversely, in the "Less" subgroup

Table 5: Heterogeneous effects in evaluations- OLS with fixed effects

Dependent Variable:	Rounds 4 to 6			
	Evaluation			
	Subgroup - More		Subgroup - Less	
	(1)	(2)	(3)	(4)
I(Female)	-22.42***	-22.64***	-7.994***	-7.428***
	(1.011)	(1.080)	(0.401)	(0.422)
I(TempAA)*I(Female)	3.638**	3.427**	0.917	0.960*
	(1.464)	(1.452)	(0.564)	(0.554)
Constant	66.05***	62.56***	66.08***	61.19***
	(0.517)	(1.065)	(0.199)	(0.402)
N	2388	2388	9576	9576
Worker controls		Yes		Yes

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents OLS regression results on evolution of beliefs about performance elicited as evaluations on a scale of 0 to 100. The estimation uses a fixed effects model where each group comprises a single round where an employer makes decisions over 4 resumes. Subgroup - More (Less) represents the subgroups of employers within the two experimental treatments who are more (less) than 57.6% likely to hire both men in round 1. Worker controls include demographic characteristics presented on their resume - employment status, education, number of spoken languages, and time zone of residence. All options within each resume characteristic are aggregated in two groups characterized by the median worker's characteristic from the worker experiment. Standard errors are shown in parentheses.

(column 3), there is a smaller but statistically significant gender bias against women in the control treatment, which does not significantly change due to the temporary affirmative action treatment (coefficient 0.917, $p = 0.104$). Controlling for worker characteristics in column (4) marginally enhances the treatment effect in the "Less" subgroup, although the magnitude remains significantly smaller than in the "More" subgroup ($p = 0.063$).

This provides evidence supporting our auxiliary hypothesis that employers more biased and discriminatory against women show a positive effect of the temporary affirmative treatment on reducing gender bias in beliefs. In contrast, employers less biased and discriminatory against women demonstrate a weaker treatment belief debiasing. This finding remains robust across various cutoff points used for subgroup classification, underscoring that as we select employers more predisposed to discriminate against women, the positive impact of temporary affirmative action on reducing gender bias in beliefs consistently strengthens (see Appendix Table B2).[28]

[28] Another auxiliary hypothesis was drawn from a surprise final stage of the experiment, designed to possibly identify employers likely to disregard any information given during the experiment. Excluding such employers was conjectured to strengthen the overall belief results. However, we find no evidence to support this in the data (see Appendix C for details).

Table 6: Heterogeneous effects in evaluations as mechanism - conditional fixed effects Logit regression

Dependent Variable:	Rounds 4 to 6			
	I(Hired)			
	Subgroup - More		Subgroup - Less	
	(1)	(2)	(3)	(4)
I(Female)	-1.214***	-0.0523	-0.640***	-0.179***
	(0.100)	(0.132)	(0.0505)	(0.0577)
I(TempAA)*I(Female)	0.466***	0.316*	0.234***	0.217***
	(0.144)	(0.168)	(0.0709)	(0.0791)
Evaluation (Normalized)		6.565***		7.770***
		(0.381)		(0.248)
N	2388	2388	9576	9576

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents conditional fixed logit regression results from hiring decisions where a group comprises a single round where an employer makes decisions over 4 resumes. Subgroup - More (Less) represents the subgroups of employers within the two experimental treatments who are more (less) than 57.6% likely to hire both men in round 1. Dependent variable is an indicator variable =1 when a worker is hired. Scale of evaluations has been normalized to 0 to 1 (from 0 to 100). Standard errors are shown in parentheses.

4.3 Mechanisms

Given the evolution of hiring choices and beliefs, the final step is to determine how beliefs influence changes in hiring decisions. Considering the heterogeneous effects observed in changes in beliefs, we do not anticipate significant inaccurate statistical discrimination within the "Less" subgroup of employers, as they exhibit lower bias and are less likely to discriminate against women without policy intervention. Conversely, the "More" subgroup displays higher bias and propensity to discriminate against women, potentially fostering conditions conducive to inaccurate statistical discrimination. Having established that temporary affirmative action reduces gender bias in beliefs among the "More" subgroup, this section investigates whether this reduction in bias translates into improved representation of women even after quotas are lifted.

Table 6 explores the hiring dynamics within the "More" subgroup (columns (1)-(2)) and the "Less" subgroup (columns (3)-(4)). This estimation decomposes the effect of the temporary affirmative action treatment between the effect driven by beliefs about performance and any residual effect not explained by beliefs.

In the "More" subgroup under the control treatment, the log odds of hiring a woman are 1.214 less than those for men. The temporary affirmative action treatment increases this by 0.466 ($p = 0.001$) units in log odds. Beliefs influence hiring decisions in two ways: first,

after controlling for beliefs, the baseline gender discrimination in hiring disappears (log odds of 0.0523, $p = 0.691$); second, beliefs partially mitigate the effect of temporary affirmative action, reducing both the size and explanatory effect of the interaction term to 0.316 log odds ($p = 0.508$).

Conversely, in the "Less" subgroup, gender discrimination against women in the control treatment is less pronounced, with log odds of hiring a woman being 0.640 less than for men. When controlling for beliefs, this effect diminishes to a 0.179 log odds against women, although the reduction is less than that observed in the "More" subgroup. Nonetheless, the temporary affirmative action treatment has a notable effect in this subgroup (0.234 log odds, $p = 0.001$), which persists even after controlling for beliefs (0.217 log odds, $p = 0.006$).

Evidence supports hypothesis 3 wherein for employers more biased and inclined to discriminate against women, temporary affirmative action leads to a lasting increase in women's representation. Beliefs about performance partially account for this improvement, correcting inaccurate statistical discrimination within the "More" subgroup of employers. On the other hand, employers less biased and less likely to discriminate against women do not experience belief changes due to temporary affirmative action. Nonetheless, we observe a positive and significant effect of temporary affirmative action on women's representation in this subgroup, which operates independently of beliefs.

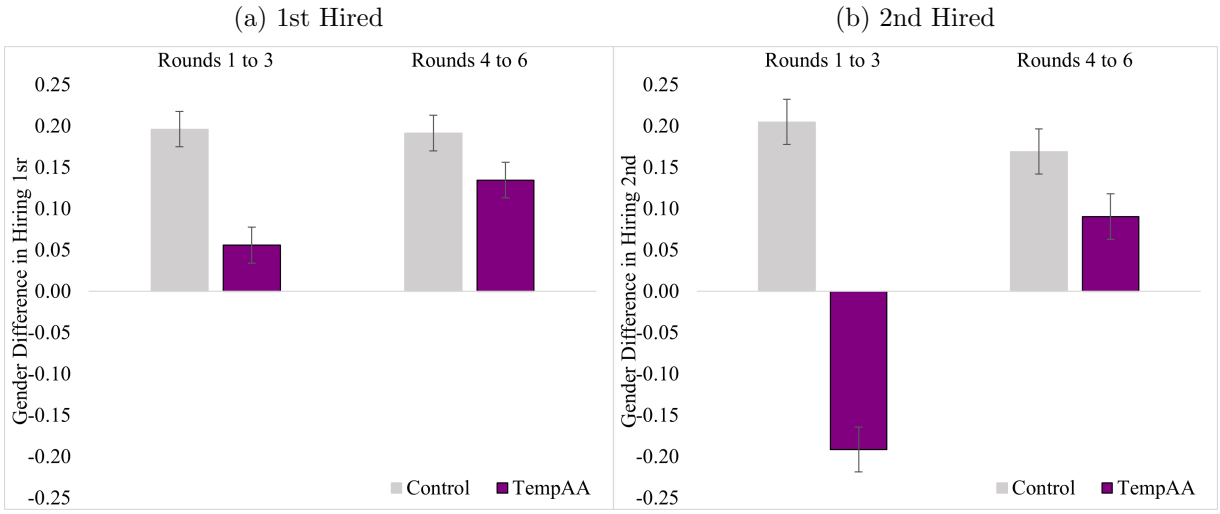
4.4 Additional Analysis - Distribution within Hiring Choices

Recall that during the submission of hiring decisions, employers not only select two candidates for hiring but also rank their choices as 1st and 2nd. This approach enables us to investigate whether the temporary affirmative action treatment can induce substantial effects, potentially leading women to be not only included as second candidates but also elevated to the position of the first choice for hiring by employers. This analysis seeks to uncover if the policy intervention can effectively influence hiring preferences towards greater gender equity in employment decisions.

Figure 4 panel (a), illustrates the gender disparity in first-choice hires. In the control treatment without affirmative action, men are 20 percentage points more likely to be hired than women. This disparity remains significant at 19 percentage points in the final three rounds of the experiment. Under the temporary affirmative action treatment, hiring becomes more balanced, with men only 6 percentage points more likely to be hired. Once quotas are lifted in the last three rounds, the male advantage persists at 13 percentage points, albeit reduced compared to the control treatment.

Panel (b) of Figure 4 shows the gender difference in second-choice hires. In the control treatment, men have a pronounced advantage, with women 20 percentage points less likely to

Figure 4: Difference in Proportions of Men Hired 1st and 2nd vs. Women Hired 1st and 2nd



Notes - This figure shows mean gender difference in proportions of men hired vs. women hired. Sub-figure (a) shows this gender difference for the candidates who are hired first and sub-figure (b) shows this difference for the candidates who are hired second excluding those hired first. The left panel in both figures show the first three rounds where there is a quota for women in effect under the temporary affirmative action treatment, and the right panel compares the two experimental treatments for the last three rounds when affirmative action is removed. Error bars correspond to 95% confidence intervals.

be hired in rounds 1 to 3. This gap narrows to 17 percentage points in rounds 4 to 6. Under the temporary affirmative action treatment, women are favored, being hired 19 percentage points more often than men when quotas are in effect. However, in rounds 4 to 6, after affirmative action is lifted, the male advantage returns, but it diminishes to 9 percentage points, significantly smaller than in the control treatment.

Overall, we observe a positive and statistically significant effect of the treatment on both 1st and 2nd hired candidates. Regarding the representation of women, they are not only given a seat at the table but are also being preferred 1st for hiring, with both effects sizes being comparable. These findings are further substantiated through regression analysis (see appendix table B1).

In line with the overall hiring results, we find that among the employers more biased and more prone to discriminate against women, improvement in representation of women as both the 1st and 2nd candidate for hiring is almost entirely driven by changes in beliefs. Conversely, for the employers who are less biased and less likely to discriminate against women, we still observe a positive and significant effect of temporary affirmative action on women's representation both as the 1st and 2nd preferred candidates.However, this effect occurs independently of any changes in beliefs (see appendix table B3).

5 Discussion and Conclusions

In a setting where employers' beliefs and hiring choices are biased against women, this paper uses a series of experiments to examine whether a temporary affirmative action policy can correct these biases and lead to a lasting improvement in the representation of women. The data reveals that a temporary quota significantly improves the representation of women and reduces gender bias in beliefs even after the policy is lifted. Notably, among employers who were initially more biased and more likely to discriminate against women, we observe a significant reduction in gender bias and a lasting improvement in women's representation due to the temporary affirmative action treatment. Within this subgroup, the changes in hiring choices favoring women are explained by shifts in beliefs. This supports the conjecture that increased exposure to women under a temporary quota provides an opportunity to correct biased beliefs, thereby improving the representation of women even after the policy ends.

This paper documents the presence of inaccurate statistical discrimination in hiring against women and presents evidence that it can be remedied through a temporary affirmative action policy. From a policy perspective, the results from this paper motivate the use of temporary affirmative action quotas to correct biased beliefs and break a pattern of self-perpetuating under-representation. Even advocates of affirmative action policies often view them as temporary fixes for social inequalities (Sowell, 2008), while our findings suggest that such policies can create fundamental changes in society that address social inequalities even beyond the policy's duration. Reflecting on the unpopularity of affirmative action as a policy measure, it might become more acceptable to the public if quotas are understood to be temporary.[29] Given the high likelihood that affirmative action will be banned in many parts of the world, this paper represents a small step forward in understanding what to expect in a post-affirmative action world.

While this study provides evidence of inaccurate statistical discrimination and demonstrates that it can be remedied through exposure using a temporary quota, the dynamics of hiring in the real world may vary based on the nature of the underlying discrimination. Future research should focus on developing methods to distinctly identify inaccurate statistical discrimination in real-world settings, differentiating it from other types of discrimination, such as those based on personal prejudice or systemic biases. By accurately diagnosing the type of discrimination present, policymakers can design more effective interventions. Temporary quotas may work well for correcting biased beliefs, but other measures like diversity training, structural reforms, or long-term affirmative action policies might be necessary to address other forms of discrimination.

[29]For instance, a 2016 Gallup poll showed that 65% of Americans disagreed with the Supreme Court's decision to allow race to be a factor in college admissions.

Another notable finding from this study is that among employers who initially held less biased beliefs and were less likely to discriminate against women, the temporary affirmative action policy also results in a sustained improvement in the representation of women. In a context where biased beliefs prevail, this outcome suggests a reduction in taste-based discrimination against women. Exploratory analysis, based on a survey question measuring attitudes towards hiring a woman when the qualifications of male and female candidates are equal, indicates that temporary affirmative action policies can cause shift in attitudes towards favoring women (see Appendix D for details). Further investigation into identifying and understanding the mechanisms behind this reduction in discriminatory preferences remains a crucial area for future research.

Moreover, longitudinal studies could offer valuable insights into the enduring impacts of temporary affirmative action policies. These studies could examine whether the initial correction of biased beliefs leads to sustained changes in hiring practices over time. Additionally, exploring the effectiveness of different types of temporary quotas across diverse industries and cultural settings could aid in customizing policies to suit specific environments and maximize their impact.

Overall, advancing our understanding of these dynamics will be pivotal in crafting informed policies that promote fairness and inclusivity in the workforce. This study highlights the potential for temporary affirmative action policies to create lasting changes in employer beliefs and practices, suggesting that even short-term interventions can have enduring positive effects on the representation of women in the workplace.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge (2022) 'When should you adjust standard errors for clustering?' *The Quarterly Journal of Economics* 138(1), 1–35

Aigner, Dennis J, and Glen G Cain (1977) 'Statistical theories of discrimination in labor markets.' *Ilr Review* 30, 175–187

Antonovics, Kate, and Ben Backes (2014) 'The effect of banning affirmative action on college admissions policies and student quality.' *Journal of Human Resources* 49, 295–322

Antonovics, Kate L, and Richard H Sander (2013) 'Affirmative action bans and the "chilling effect".' *American law and economics review* 15, 252–299

Arrow, Kenneth (1971) 'The theory of discrimination'

Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval (2018) 'The spillover effects of affirmative action on competitiveness and unethical behavior.' *European Economic Review* 101, 567–604

Becker, Gary Stanley (1957) *The economics of discrimination: an economic view of racial discrimination* (University of Chicago)

Beede, David N, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms (2011) 'Women in stem: A gender gap to innovation.' *Economics and Statistics Administration Issue Brief*

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope (2019) 'Inaccurate statistical discrimination: An identification problem'

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016) 'Stereotypes.' *The Quarterly Journal of Economics* 131, 1753–1794

Buckles, Kasey (2019) 'Fixing the leaky pipeline: Strategies for making economics work for women at every stage.' *Journal of economic perspectives* 33, 43–60

Coate, Stephen, and Glenn C Loury (1993) 'Will affirmative-action policies eliminate negative stereotypes?' *The American Economic Review* pp. 1220–1240

Coffman, Katherine B, Christine L Exley, and Muriel Niederle (2021) 'The role of beliefs in driving gender discrimination.' *Management Science* 67(6), 3551–3569

Coffman, Katherine Baldiga (2014) 'Evidence on self-stereotyping and the contribution of ideas.' *The Quarterly Journal of Economics* 129, 1625–1660

Coffman, Katherine, Manuela Collis, and Leena Kulkarni (2019) *Stereotypes and belief updating* (Harvard Business School)

Coffman, Katherine, Scott Kostyshak, and Perihan Saygin (2024) 'Choosing and using information in evaluation decisions'

Danz, David, Lise Vesterlund, and Alistair J Wilson (2022) 'Belief elicitation and behavioral incentive compatibility.' *American Economic Review*

Deininger, Klaus, Songqing Jin, Hari K Nagarajan, and Fang Xia (2015) 'Does female reservation affect long-term political outcomes? evidence from rural india.' *The Journal of Development Studies* 51, 32–49

Dianat, Ahrash, Federico Echenique, and Leeat Yariv (2022) 'Statistical discrimination and affirmative action in the lab.' *Games and Economic Behavior* 132, 41–58

Dickson, Lisa M (2006) 'Does ending affirmative action in college admissions lower the percent of minority students applying to college?' *Economics of Education Review* 25, 109–119

Fershtman, Chaim, and Uri Gneezy (2001) 'Discrimination in a segmented society: An experimental approach.' *The Quarterly Journal of Economics* 116, 351–377

Goldin, Claudia (2015) 'Gender and the undergraduate economics major: Notes on the undergraduate economics major at a highly selective liberal arts college.' *manuscript, April*

Gupta, Neeraja, Luca Rigott, and Alistair Wilson (2021) 'The experimenters' dilemma: Inferential preferences over populations.' *arXiv preprint arXiv:2107.05064*

Hill, Andrew J (2017) 'State affirmative action bans and stem degree completions.' *Economics of Education Review* 57, 31–40

Hossain, Tanjim, and Ryo Okui (2013) 'The binarized scoring rule.' *Review of Economic Studies* 80, 984–1001

Lepage, Louis Pierre (2021) 'Endogenous learning, persistent employer biases, and discrimination.' *Persistent Employer Biases, and Discrimination (March 2, 2021)*

Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland (2015) 'Expectations of brilliance underlie gender distributions across academic disciplines.' *Science* 347, 262–265

Mann, Allison, and Thomas A DiPrete (2013) 'Trends in gender segregation in the choice of science and engineering majors.' *Social science research* 42, 1519–1541

Matsa, David A, and Amalia R Miller (2011) 'Chipping away at the glass ceiling: Gender spillovers in corporate leadership.' *American Economic Review* 101, 635–639

Miller, Amalia R, and Carmit Segal (2012) 'Does temporary affirmative action produce persistent effects? a study of black and female employment in law enforcement.' *Review of Economics and Statistics* 94, 1107–1125

Miller, Conrad (2017) 'The persistent effect of temporary affirmative action.' *American Economic Journal: Applied Economics* 9, 152–190

Niederle, Muriel, Carmit Segal, and Lise Vesterlund (2013) 'How costly is diversity? affirmative action in light of gender differences in competitiveness.' *Management Science* 59(1), 1–16

Paola, Maria De, Vincenzo Scoppa, and Rosetta Lombardo (2010) 'Can gender quotas break down negative stereotypes? evidence from changes in electoral rules.' *Journal of Public Economics* 94, 344–353

Phelps, Edmund S (1972) 'The statistical theory of racism and sexism.' *The american economic review* 62, 659–661

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014) 'How stereotypes impair women's careers in science.' *Proceedings of the National Academy of Sciences* 111, 4403–4408

Rhode, Deborah L (2017) *Women and leadership* (Oxford University Press)

Rooth, Dan-Olof (2010) 'Automatic associations and discrimination in hiring: Real world evidence.' *Labour Economics* 17, 523–534

Savigny, Heather (2014) 'Women, know your limits: Cultural sexism in academia.' *Gender and education* 26, 794–809

Sowell, Thomas (2008) *Affirmative action around the world* (Yale University Press)

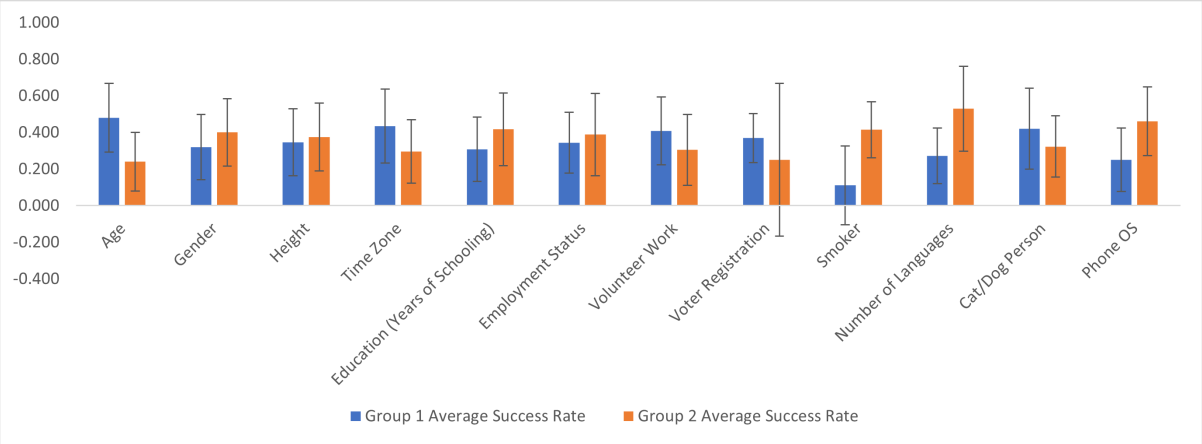
Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth (2018) 'The ties that double bind: social roles and women's underrepresentation in politics.' *American Political Science Review* 112, 525–541

Wilson, Alistair, and Emanuel Vespa (2018) 'Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language'

Yagan, Danny (2016) 'Supply vs. demand under an affirmative action ban: Estimates from uc law schools.' *Journal of Public Economics* 137, 38–50

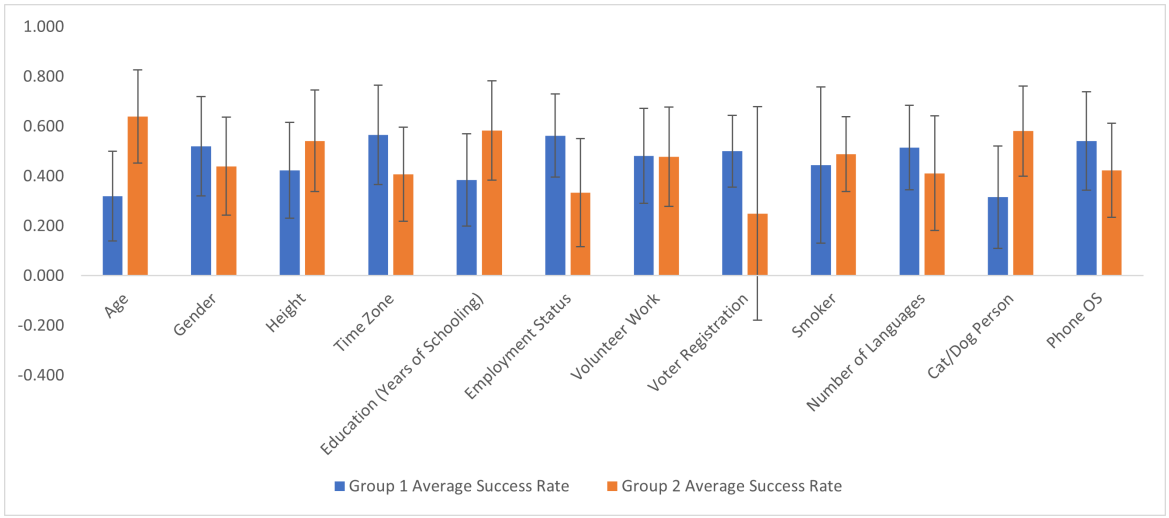
APPENDIX A: Worker and Evaluator Results

Figure A1: Average rate of success by worker demographic characteristics: Business



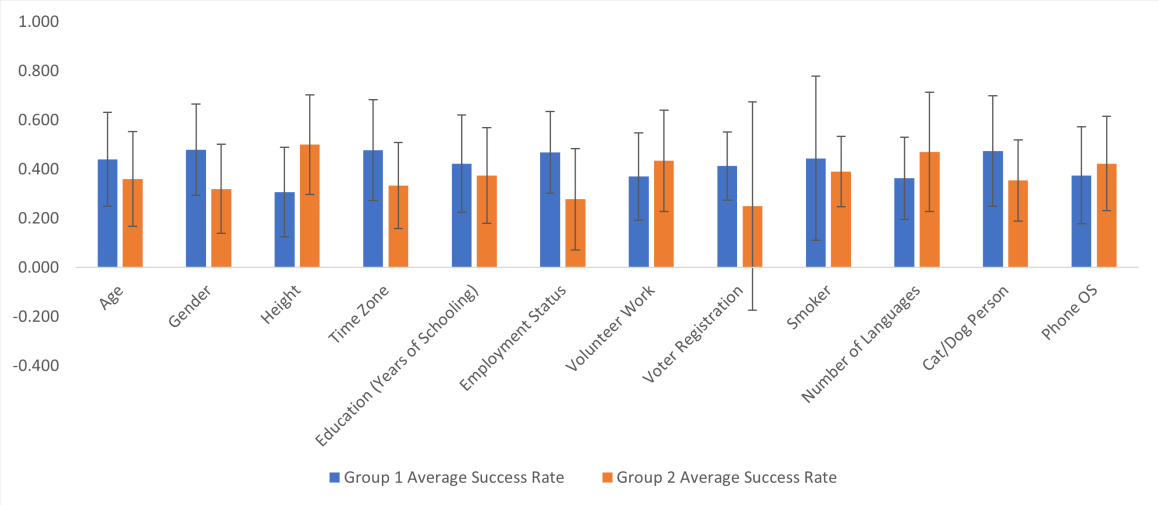
Notes - This figure shows average success rate by worker characteristics. Groups are divided based on demographic characteristics as shown in Table 1. Error bars represent bootstrapped 95% confidence intervals.

Figure A2: Average rate of success by worker demographic characteristics: Sports



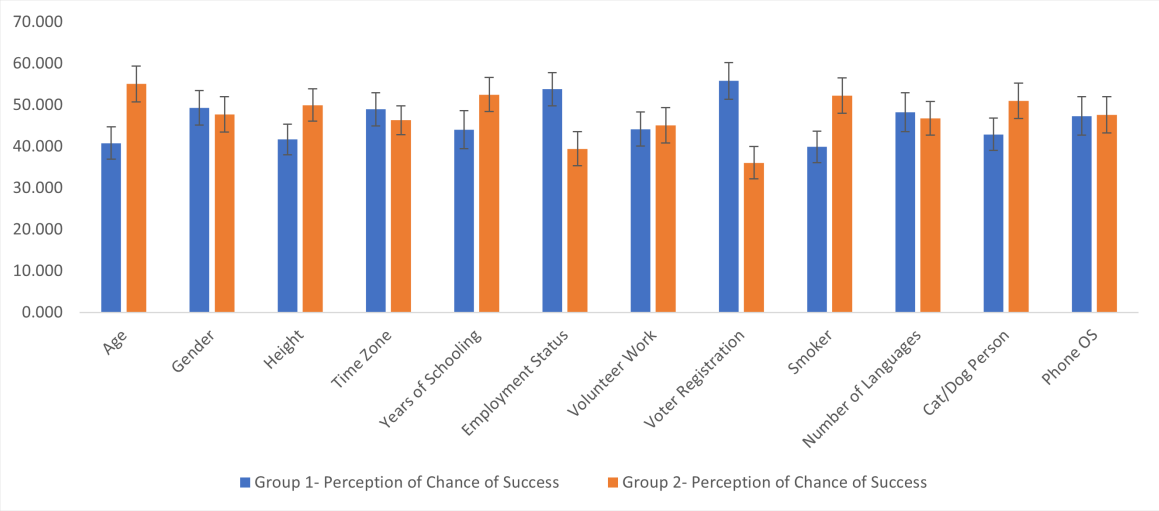
Notes - This figure shows average success rate by worker characteristics. Groups are divided based on demographic characteristics as shown in Table 1. Error bars represent bootstrapped 95% confidence intervals.

Figure A3: Average rate of success by worker demographic characteristics: Video Games



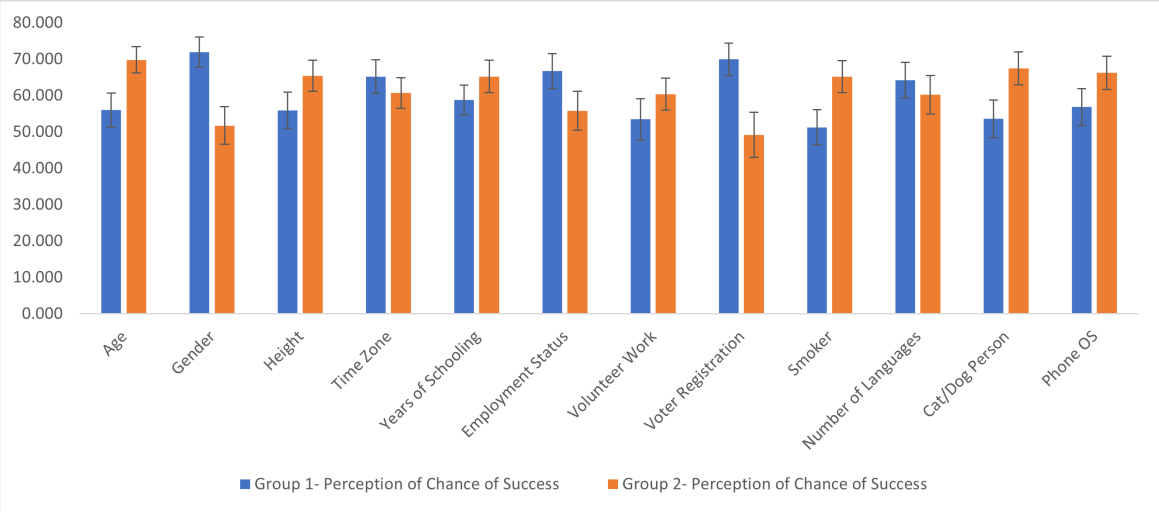
Notes - This figure shows average success rate by worker characteristics. Groups are divided based on demographic characteristics as shown in Table 1. Error bars represent bootstrapped 95% confidence intervals.

Figure A4: Average perceived likelihood of success from experiment 2 split by worker demographic characteristics: Business



Notes - This figure shows average perceived likelihood of success from experiment 2 split by worker characteristics. Groups are divided based on demographic characteristics as shown in Table 1. Error bars represent bootstrapped 95% confidence intervals.

Figure A5: Average perceived likelihood of success from experiment 2 split by worker demographic characteristics: Sports



Notes - This figure shows average perceived likelihood of success from experiment 2 split by worker characteristics. Groups are divided based on demographic characteristics as shown in Table 1. Error bars represent bootstrapped 95% confidence intervals.

APPENDIX B: Additional Tables - Employer

Table B1: 1st and 2nd hiring results - conditional fixed effects Logit regression

Dependent Variable:	Rounds 4 to 6			
	I(1st Hired)		I(2nd Hired)	
	(1)	(2)	(3)	(4)
I(Female)	-0.807***	-0.803***	-0.552***	-0.491***
	(0.0560)	(0.0611)	(0.0545)	(0.0599)
I(TempAA)*I(Female)	0.255***	0.272***	0.252***	0.258***
	(0.0776)	(0.0800)	(0.0771)	(0.0788)
N	11964	11964	8973	8973
Worker controls		Yes		Yes

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents conditional fixed logit regression results from hiring decisions where a group comprises a single round where an employer makes decisions over 4 resumes. Dependent variable for specifications 1 and 2 is an indicator variable =1 when a worker is hired 1st and, for specifications 3 and 4 it is an indicator variable =1 when worker is hired 2nd after excluding those hired 1st. Worker controls include demographic characteristics presented on their resume - employment status, education, number of spoken languages, and time zone of residence. All options within each resume characteristic are aggregated in two groups characterized by the median worker's characteristic from the worker experiment. Standard errors are shown in parentheses.

Table B2: Heterogeneous effects in evaluations with moving cutoff - OLS with fixed effects

Dependent Variable:	Rounds 4 to 6						
	Evaluations						
	> 0	> 0.10	> 0.20	> 0.30	> 0.40	> 0.50	> 0.60
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Female)	-11.01***	-11.07***	-12.92***	-15.03***	-17.49***	-19.94***	-22.60***
	(0.388)	(0.391)	(0.445)	(0.517)	(0.667)	(0.838)	(1.081)
I(TempAA)*I(Female)	1.701***	1.621***	1.855***	1.928***	2.370**	2.663**	2.750*
	(0.549)	(0.552)	(0.626)	(0.729)	(0.935)	(1.178)	(1.607)
Constant	66.08***	66.10***	65.89***	65.63***	66.30***	66.06***	66.39***
	(0.194)	(0.195)	(0.221)	(0.258)	(0.330)	(0.416)	(0.566)
N	11964	11640	9372	7308	5040	3480	2040

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents OLS regression results on evolution of beliefs about performance elicited as evaluations on a scale of 0 to 100. The estimation uses a fixed effects model where each group comprises a single round where an employer makes decisions over 4 resumes. Moving from columns (1) to (7) the sample restriction based on probability of hiring both men in round 1 increases, for e.g. column (4) shows the results when probability of hiring both men in round 1 is greater than 30%. Standard errors are shown in parentheses.

Table B3: Heterogeneous effects in evaluations as mechanism for 1st and 2nd hiring choices - conditional fixed effects Logit regression

Dependent Variable:	Rounds 4 to 6							
	Evaluations							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I(Female)	-1.204***	-0.184	-0.712***	-0.287***	-0.970***	0.0239	-0.441***	-0.0796
	(0.134)	(0.163)	(0.0619)	(0.0691)	(0.121)	(0.154)	(0.0612)	(0.0681)
I(TempAA)*I(Female)	0.348*	0.183	0.228***	0.199**	0.463***	0.315	0.190**	0.215**
	(0.187)	(0.213)	(0.0856)	(0.0940)	(0.173)	(0.198)	(0.0862)	(0.0936)
Evaluation		6.233***		7.527***		5.906***		6.862***
		(0.459)		(0.294)		(0.460)		(0.302)
N	2388	2388	9576	9576	1791	1791	7182	7182

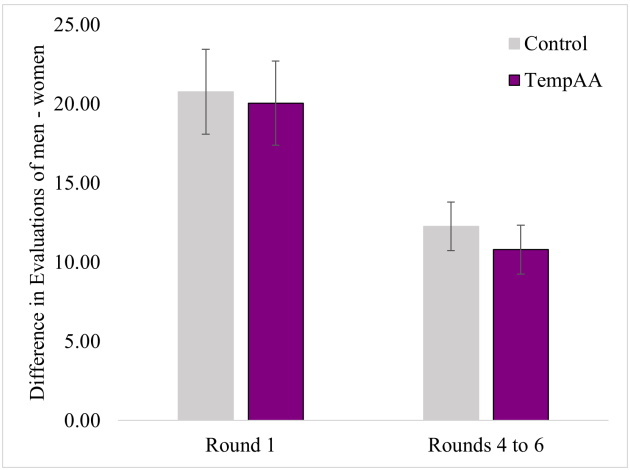
Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents OLS regression results on evolution of beliefs about performance elicited as evaluations on a scale of 0 to 100. The estimation uses a fixed effects model where each group comprises a single round where an employer makes decisions over 4 resumes. Moving from columns (1) to (7) the sample restriction based on probability of hiring both men in round 1 increases, for e.g. column (4) shows the results when probability of hiring both men in round 1 is greater than 30%. Standard errors are shown in parentheses.

APPENDIX C: Additional Stage of Employer Experiment

A surprise final decision stage in the experiment helps identify employers likely to disregard information about employee performance. After submitting their hiring choices in the last round and before learning about the outcomes of their hired employees, the employer participants enter this final decision stage. Here, they can costlessly request information about a worker's success by pressing a button for as many worker resumes as they wish. On the next page, they see the actual performance of the workers they chose to review and have the opportunity to revise their hiring choices accordingly. If they do not request this information for any worker, they skip to the end of the round.

Employers who do not opt for this costless information for any worker are characterized as likely to disregard employee performance information during the experiment. We identified a total of 543 such employers out of 997: 274 in the control treatment and 269 in the temporary affirmative action treatment. This classification of employers was pre-registered with the conjecture that their presence in the sample could potentially suppress belief dynamics as these employers are likely to ignore the feedback about hired employees' performance provided at the end of each round. However, as evident from Figure C1, excluding these employers from the sample does not strengthen the belief results as conjectured.

Figure C1: Difference in Evaluations of Men and Women



Notes - This figure shows mean gender difference in evaluations of men and women. The left panel represents round 1 evaluations, and the right panel compares the two experimental treatments for the last three rounds where quota is removed under the temporary affirmative action treatment. Employers likely to disregard information information as identified by the surpirse final stage of the experiment are excluded in this figure. Error bars correspond to 95% confidence intervals.

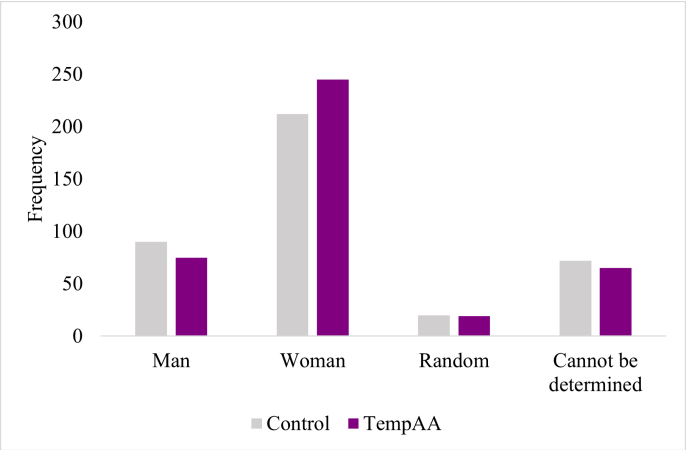
APPENDIX D: Attitude of Employers in "Less" Subgroup

A question designed to measure changes in attitudes towards hiring women was included in the exit survey of the employer experiment. Note that this question and the analysis described in this section were not pre-registered. The question asked:

"Suppose that there are two candidates for a job, one identifying as a man and the other as a woman. Both are equal in all characteristics. When only one can be hired, who, in your opinion, should be hired: the man or the woman?"

It was an open-response question, and the answers were later categorized into four groups: (1) Man, (2) Woman, (3) Random, and (4) Cannot be determined.[30] The raw frequency of these responses within the "Less" subgroup is shown in Figure D1 below. Employers under the temporary affirmative action program exhibit a stronger preference for hiring the female candidate when the qualifications of the male and female candidates are equal.

Figure D1: Frequency of responses to attitudes question within the "Less" subgroup



Notes - This figure shows frequency of response to the survey question measuring attitudes towards hiring a man vs. a woman by the two treatment conditions within the "Less" subgroup.

The causal impact of the temporary affirmative action on attitude towards hiring women is confirmed in a simple regression model using a logit model:

$$I(Attitude - Woman)_i = I(TempAA)_i\beta + \gamma_i + \epsilon_i \tag{5}$$

[30]The responses were independently coded by two research assistants, with any conflicts resolved by the author.

Table D1: Attitude of Employers in "Less" Subgroup - Logit regression

Dependent Variable:	I(Attitude-Woman)	
	(1)	(2)
I(TempAA)	0.280*	0.282*
	(0.144)	(0.146)
Constant	0.153	-0.347
	(0.101)	(1.552)
N	798	798
Employer controls		Yes

Notes - *** p<0.01, ** p<0.05, * p<0.1; This table presents Logit regression results for attitude of employers in "Less" subgroup towards hiring women. Dependent variable is an indicator variable =1 when the employer is from the temporary affirmative action treatment condition. Employer controls include age, education, and employment status. Standard errors are clustered at employer level and shown in parentheses.

where $I(Attitude-Woman)_i$ represents an indicator variable for whem employer i response to the attitude question points towards a woman. The variable $I(TempAA)_i$ denotes an indicator for the temporary affirmative action treatment condition,, γ_i represent controls for employer demographics, and ϵ_i is the error term. The standard errors are clustered at the employer level. The output from this regression model is shown in Table D1 where we see a significant effect of the temporary affirmative action treatment on attitude towards hiring women.