Fair Play in a Male Dominated Diversity Setting[*]

Puneet Arora
Management Development Institute Gurgaon

Abstract

This study explores whether male leaders in male-dominated environments make biased decisions against female employees when work assessments are subjective. We explore two potential explanations: resorting to stereotypes due to ambiguity about individual performance (statistical bias) and lack of personal rewards to male leaders for retaining females. On the employee side, we examine whether in such male-dominated work settings, female employees underestimate their performance compared to their male counterparts, whether they have an optimism bias regarding how a male leader would assess them relative to their male counterparts, and whether they would prefer to work in a male-dominated work environment characterized with lesser ambiguity around individual performance. These questions are studied through a controlled online experiment involving a cricket prediction game, a sport traditionally dominated by males.

Keywords: gender, discrimination, statistical bias, inclusion, fairness, diversity
JEL Codes: C91, D91, J01, J16, J71

[*] Corresponding Author: Management Development Institute Gurgaon, Haryana, India; E-mail: *puneet.arora@mdi.ac.in.* The study received IRB approval and financial support from MDI Gurgaon. Declarations of interest: none.

Introduction

As management increasingly prioritizes workforce diversity, driven by both internal motivations and external pressures like the pursuit of higher Environmental, Social, and Governance (ESG) scores, a critical question arises: does diversity within a team ensure fair treatment for all its members? This study primarily aims to explore this question within the context of team-based work dynamics, where evaluating individual performance is inherently complex. Although team leaders' performance reviews are crucial for career decisions, the subjective nature of these assessments can introduce bias. In a predominantly male-dominated environment, where females represent the diversity group, this study investigates whether biases rooted in gender stereotypes affect the decisions of male team leaders concerning female team members, especially when individual performance data is unclear.

We will conduct a controlled online experiment using an incentivized game to simulate team-based decision-making processes. Each team will consist of three members: one male leader and two "predictors", who will forecast the outcomes of T20 World Cup 2024 cricket matches (excluding matches involving India) on several match-related measures. The study will be conducted on four match days (four rounds) during the Super 8 stages. We exclude India matches from our experiment to avoid desirability-driven optimism bias that could confound predictions and our findings (Massey et al., 2011). Monetary rewards will be allocated to the top three teams with the highest scores on each match day, with male team leaders receiving a larger share (50%) of the reward and the predictors evenly splitting the remaining reward (25% each), regardless of their individual performances. Each predictor will be assigned to two leaders from two different treatments, increasing their potential earnings.

Team leaders will have the discretion to replace one or both of their predictors before each match day, with a penalty of 100 points deducted from their score on the following day for each replacement decision. While one predictor will be male and the other female, we will not explicitly reveal their gender to the team leader. Instead, each predictor will be assigned a randomly generated male or female avatar, which will be visible only to the team leader without any explicit mention of gender. Consequently, female predictors may be represented by either male or female avatars to the team leader, and the same applies to male predictors. The team leader will always be assigned a male avatar that matches their actual gender. This approach aims to create the impression among team leaders that the avatars accurately reflect the true gender of the predictors. Since team members never physically interact, they remain unaware of the actual gender of any team player.

A team leader's decision to retain or replace a predictor theoretically hinges on both team performance and individual predictor performance. However, lacking individual-level performance data, leaders might attribute poor team performance to a stereotypically disadvantaged group, such as females in cricket. This bias may result in a higher likelihood of replacing players with female avatars during instances of poor team performance, a phenomenon known as statistical discrimination, which is rooted in group-based stereotypes (Phelps, 1972). Alternatively, leaders might replace female avatars more frequently despite believing they perform as well as or better than male avatars, driven by a preference for male predictors due to lower female potential (Benson et al., 2024). This is referred to as taste-based discrimination (Becker, 2010).

To differentiate between these two forms of bias, we conduct an incentivized retrospective survey with team leaders at the experiment's end, asking them to predict each predictor's performance relative to the other. Comparing their performance predictions with their decisions to replace or retain will help us determine whether their decisions are influenced by performance predictions or gender preferences, irrespective of performance. Additionally, we implement a treatment where team leaders have partial visibility into individual performance, reducing ambiguity. If bias against female predictors decreases in this information treatment, it would indicate statistical bias; if it persists, it would suggest taste-based bias.

The primary motivation behind a team leader's potential statistical bias against females may arise from a desire to improve their chances of ranking in the top three and securing associated rewards. This scenario mirrors many corporate environments, where top management advocates for diversity but rewards managers based solely on team performance metrics, with little consideration for team diversity. As a result, ethical considerations may become less prominent, and leaders might prioritize actions that maximize their rewards, such as unfairly replacing a player to increase their likelihood of success. However, if top management were to implement a reward system for managers with diverse teams, the focus could shift from merely winning through replacement to winning through retention and fairness. To test this hypothesis, we introduce another treatment arm where team leaders receive additional personal compensation (Rs 50) for retaining their existing team for the subsequent round. While this reward is for retention, with no explicit mention of diversity, we believe that this additional incentive will reduce discriminatory behavior against female predictors compared to the baseline treatment, where no personal reward is provided for retaining predictors. This approach aims to encourage diversity and promote fairness in decision-making concerning diverse groups.

The secondary objective of this study is to understand the beliefs of employees from diverse backgrounds, specifically female employees in our context. While their primary role is to predict match outcomes on each matchday, we will also ask additional incentivized survey questions. These questions will have them rate their expected performance relative to the other predictor to gauge their confidence in a male-dominated environment. We hypothesize that females will be underconfident and rate themselves lower than males, while males will be overconfident and rate themselves higher than females.

Additionally, participants will predict the leader's decisions to retain or replace each predictor based on their team's ranking, using a strategy method. This will help us understand their beliefs about how leaders will assess them. We expect that female predictors will exhibit optimism bias, expecting leaders to retain them as likely as their male counterparts, even though leaders may be more inclined to replace females, conditional on performance. Such optimism bias suggests that female employees in male-dominated environments might not be adequately prepared for potential biases, potentially resulting in less effort to signal their quality and commitment to leaders. Finally, we will assess their preference for moving to a less ambiguous male-dominated work environment, expecting that underconfident and optimistic females may not prefer this shift.

This study contributes to the literature by investigating how gender stereotypes and ambiguous performance information affect team leader decisions. We also examine the effectiveness of policies enhancing transparency and rewarding leaders for retaining diversity as strategies to mitigate biases. These policies could inform interventions to promote fairness and reduce biases in performance evaluations and promotions. Additionally, understanding the beliefs of diverse groups regarding their performance and leader assessments will provide insights into the efforts females may exert to move to a more transparent work environment. However, a potential limitation is the use of a simulated task, which may not fully reflect real-world scenarios. Future research should explore the applicability of these findings in actual work environments.

Literature Review

Gender bias can affect women at different stages of their careers, such as during recruitment (Goldin and Rouse, 2000). For instance, studies such as Bertrand and Mullainathan (2004), Carlsson (2011) and Banerjee et al. (2009) varied applicant names to examine discrimination in callback rates for real vacancies. Similarly, Moss-Racusin et al. (2012) manipulated names to examine bias against female candidates in science faculties. Additionally, gender bias can impact women's promotion prospects (Huang et al., 2024; Sarsons, 2017).

Our study focuses on gender bias in managers' performance evaluation and promotion of their team members. The most closely related study is by Benson et al. (2024), which finds that women are assessed to have lower potential despite having the same performance ratings as men. However, our study examines this bias in a male-dominated setting, where the issue becomes more critical due to policy efforts to increase diversity. Women often face the stigma of perceived incompetence, particularly when recruited under affirmative action policies (Heilman et al., 1992; Kravitz and Platania, 1993). This can lead to a potential DEI *backfire*, where diversity initiatives may unintentionally result in unfair treatment of the diverse population (Leslie, 2019).[1]

Several studies have tested solutions to reduce gender bias. For example, the role of deliberation among committee members has been examined, but Mengel (2021) finds that gender bias persists even after deliberation. Anonymizing resumes during hiring has been suggested and adopted (Krause et al., 2012), but this approach can have unintended consequences as recruiters might use implicit signals and cues to infer gender identities, thus continuing gender bias (Foley and Williamson, 2018; Behaghel et al., 2015). Deciding jointly vs. separately also tends to base decisions on performance over stereotypes (Bohnet et al., 2016). Organizations can incorporate diversity-based criteria into their performance and promotion policies, which are more likely to succeed in retaining and promoting diverse talent (Cleveland et al., 2000).

Our study contributes to the literature by examining two specific policies: enhancing transparency in performance measurement (Information Treatment) and rewarding leaders for retaining and fostering diversity (Reward Treatment). In the Information Treatment, we provide partial information on team members' performance. Although this information might not perfectly predict future performance, it can help leaders make more equitable decisions. According to the Peter Principle, managers often prioritize current performance in promotion decisions (in our case, retention) over potential future performance (Benson et al., 2019; Waldman, 2003; Peter et al., 1969). We hypothesize that the application of the Peter Principle may differ based on the predictor's gender, with females being more likely to be replaced due to perceived lower future potential despite higher current performance,

[1] Such unfair treatment may be more pronounced in female leaders' decision-making in male-dominated environments, as women often face penalties for valuing diversity (Hekman et al., 2017). As a result, women may internalize the notion of not prioritizing diversity. To ensure adequate statistical power, we focus our study exclusively on male leadership, which is a significant area of concern.

while males might be retained strictly based on current performance. Additionally, we test the findings reported by Egan et al. (2022), which indicate that females are punished more harshly than males for low performance. In our Information Treatment, we will also investigate whether the likelihood of replacing females is higher than that of males, especially when the team's performance on the scoreboard is poor, thereby replicating the findings from Egan et al. (2022).

In the Reward Treatment, we offer personal rewards to team leaders who retain their team members for the next round. We hypothesize that this policy will shift the focus from maximizing team performance to retaining team members, thereby incentivizing leaders to maintain diverse teams that include both male and female avatars. While this reward will benefit both male and female avatars, we believe it will particularly help to reduce bias against female avatars, which might be more pronounced in the absence of such rewards.

Hypotheses

In a profit-driven environment, management should prioritize recruiting based on merit, regardless of candidate's identity. Once hired, team leaders should evaluate employee performance and contributions, rewarding strong performers and addressing underperformance. However, the growing emphasis on ESG scores, including social responsibility, necessitates affirmative action for private companies in India. This involves predetermined quotas for diverse candidates, specifically focusing on increasing female representation in traditionally male-dominated fields, during recruitment, without compromising on the quality of the selected candidates. The goal is to achieve a balanced workforce with a high average employee quality to maximize profits while fostering gender diversity. Team members' performance is assessed by team managers, and with complete information, rewards and penalties can be fairly determined for each member, regardless of their identity. However, in work scenarios where individual performance is difficult to quantify and has a high ambiguous component, team leaders may rely on expected individual contributions. If leaders are unbiased, then on average, there will be no gender-based bias in their assessment decisions.

Hypothesis 1*: Team leaders evaluate employees based on expected performance, unrelated to gender, assuming recruitment maintained quality.*

However, subjective evaluations in ambiguous performance scenarios may introduce bias based on gender-related stereotypes, potentially disadvantaging females in traditionally male-dominated fields.

Hypothesis 2*: Conditional on equal actual but unobservable performance within a male-dominated environment, team leaders are likely to exhibit bias against female employees.*

This stereotype-based bias is expected to decrease with greater transparency of individual team member performance. Reduced bias would indicate a statistical bias based on group attributes, which is likely to diminish as more information about individual performance becomes available. Bias that persists even after individual performance information is revealed indicates taste-based discrimination.

Hypothesis 3*: Conditional on equal actual and (partially) observable performance within a male-dominated environment, team leaders will exhibit lesser bias (relative to unobservable performance scenario) against female employees.*

While top management advocates for diversity, team leaders are often informed that profit maximization is the company's priority. This creates an ethical dilemma for team leaders who must balance increasing profits with managing diversity within the organization. In corporations where team performance and profits are more salient than retaining diversity, an ethical blind spot may arise, leading team leaders to unfairly judge females as less productive in male-dominated fields. However, a shift in corporate focus that promotes and rewards team leaders for retaining and managing diversity (e.g., through additional compensation or bonuses for retaining diverse teams) could shift managers' focus towards supporting diversity, while still striving for higher profits but with diversity as a more salient goal.

Hypothesis 4*: Conditional on equal actual but unobservable performance within a male-dominated environment, when team leaders are rewarded for retaining their teams, they are lesser likely to exhibit bias against female employees.*

Methods

Participants

The study aims to recruit 900 participants through social media posts to participate in the experiment.[2] Of these, 450 male participants will be randomly designated as Team Leaders, responsible for overseeing a team consisting of two predictors or Group Members. To ensure diversity within the teams, each set of predictors will include one male and one female participant. We will enroll an additional 225 male participants and 225 female participants as predictors. Each predictor will be randomly assigned to two leaders, while each leader will be assigned to only one team.

The 450 male leaders will be recruited from master's programs or those who have already graduated, while the 450 predictors will be drawn from undergraduate programs across various fields of study. This setup ensures that predictors are more comparable to each other in skill set and fosters the perception of the leader as a senior member relative to the predictors. This distinct recruitment process for male leaders from master's (or higher) programs and predictors from undergraduate programs also serves to eliminate any potential interaction between leaders and predictors outside the experiment.

Since the leader may choose to replace their assigned predictor(s), we will also have a waitlist of 400 predictors (200 male and 200 female) and 50 team leaders (all male) ready before the start of the experiment. This allows for replacements based on the leaders' decisions and accounts for any non-responses from the recruited participants, ensuring we

[2] Note that 900 is a planned sample size, however, the actual sample size will depend on the number of registrants who show interest in participating in the study.

have sufficient participants for each match day.

Treatment

Each team is composed of three members: one male team leader and two predictors, one male and one female. However, the gender or any other demographic information of any team member is not explicitly disclosed to anyone within the team. Instead, participants are assigned random numbers and avatars. These avatars are randomly assigned as either male or female to the two predictors in each team. Consequently, one-fourth of the time, predictors are allocated avatars corresponding to their actual gender, while in two-fourth of cases, one predictor is assigned an avatar of the incorrect gender. In the remaining one-fourth of instances, both predictors receive avatars representing the incorrect gender. These avatars are visible solely to the team leader and remain concealed from the predictors, who can only see the random number assigned to the team members and to the team. The team leader, however, is consistently assigned a male icon, aligning with their actual gender. This approach suggests to the leader that the assigned avatar gender of the predictors corresponds to their true gender.

The random allocation of gendered avatars serves as our primary experimental intervention, allowing us to assess whether leaders evaluate individuals based on these assigned genders inferred through the avatars. We refer to this as our Baseline treatment (Treatment T1), which helps us understand whether team leaders exhibit bias in their replacement decisions against female predictors when individual predictor performances are not visible. This experimental manipulation is similar, in spirit, to several other studies that examine the presence of gender bias by varying names on resumes or in emails (MacNell et al., 2015; Milkman et al., 2012; Moss-Racusin et al., 2012; Bertrand and Mullainathan, 2004), and has emerged as a response to challenges in causally identifying discrimination using naturally occurring data (Charles and Guryan, 2011). In our case, we vary avatars that appear male or female but do not use any names. Use of avatars to signal gender has been used in several prior studies to induce a particular gender (Crone and Kallen, 2022; Lopez et al., 2019; Chang et al., 2019). This intervention helps us understand how a random assignment of gender, controlling for actual performance and team performance, influences team leaders' decisions, providing clearer evidence of biases based on stereotypes and personal tastes.

The experiment will include two additional treatments: Discrimination Type Treatment (T2) and Reward for Retention Treatment (T3). Treatment T2 aims to determine if providing team leaders with partial visibility into individual predictor performance reduces bias in their replacement decisions, helping to identify whether the discrimination is statistical or taste-based. Leaders will receive performance data for 60% of the predicted outcomes on a match day, with the remaining 40% still ambiguous. Treatment T3 investigates whether incentivizing team leaders to retain their team members promotes fairer treatment of employees and reduces biased decision-making against female avatars, even when individual predictor performance is not visible.

Across all treatments, the primary outcome measured is the team leader's decision to replace or retain existing predictors. The scoring rules will remain consistent across all treatments and teams; however, the visibility of individual performance varies between T1 and T2, and the personal reward for the leader to retain the existing team varies between T1 and T3. Team leaders (or teams) will be randomly assigned to the three treatments (T1, T2, and T3) in a between-subject design, where each team will participate in only one of the three treatments across their four rounds (match days) of surveys. This setup ensures that each team leader is assigned to a single treatment, while each predictor will be part of two treatments. Predictors will not predict match outcomes separately for each of their team leaders, who belong to different treatment arms. They will not be informed about the different treatments, as the treatments only affect the decision-making of the team leaders. Each predictor will be informed about one treatment and will predict the outcomes for the team assigned to that treatment, unaware that the other team they are part of is assigned to a different treatment.

Validation Check

We will conduct a validation check with 10% of the leaders, asking them to choose avatars for themselves from two options: one male and one female. Our goal is to determine whether the majority can correctly identify and select the avatar that matches their own gender.

Using the survey with predictors (described later), we will validate the perception of cricket as a male-dominated sport, especially regarding prediction skills. Specifically, at the end of each round, we will ask participants to evaluate their performance relative to the other predictor in their team. If females and males believe that males are going to perform better than females on the prediction task, this will help us verify whether both gender believe cricket to be a male-dominated context.

Cricket as a Decision-Making Setting

The T20 Men's World Cup 2024 spans nine days, featuring seven days of Super Eight matches with two cricket matches per day, except for two days with one match each. Additionally, there is one day for the semi-finals, consisting of two matches, and one day for the final match. Each cricket match involves a maximum of 20 overs per side and typically lasts around four hours. A cricket team consists of 11 players, including expert batters, bowlers, and all-rounders proficient in both batting and bowling. The outcome of a match is determined by a win or loss, with tie-breaker rules in place. Furthermore, a man-of-the-match title is awarded to the best performer of each match. We conduct our prediction surveys and leader's decision-making tasks only on four match days during the Super 8 stages, when the matches do not involve team India.

The context of cricket is well suited to our research question, which investigates how male team leaders make decisions about male and female team members in an environment of ambiguous performance information within a traditionally male-dominated setting. While women play and watch cricket, it has historically been a male-dominated sport and continues to be perceived as such. Generally, people—both men and women—believe that men are more knowledgeable about cricket. This makes our prediction game about cricket match outcomes a reflection of a male-dominated work environment, where males are expected to be better performers than females, paralleling traditional male-dominated fields like STEM, where males are often perceived as better performers than females.

Using our experimental prediction game, we can test whether team leaders, handling two predictors—one male and one female—decide differently on the consequences to follow (replace or retain) for each gender, especially when they do not know the actual performance outcomes of the two genders. These aspects of our experimental game, in collaboration with the ongoing T20 Men's Cricket World Cup, create a male-dominated work environment where we have introduced diversity into the roles of team members or predictors. The team leader, striving for better team performance and greater rewards, faces the challenge of making decisions about the fate of these predictors as their unobservable decision-making unfolds for the team and the leader.

Procedure

In our experimental setup, we randomly assign 900 participants to form 450 teams, each with three members. These teams are randomly assigned to one of three treatments: T1, T2, and T3. In each treatment, participants designated as predictors will forecast match outcomes before each match day. Predictions will include the winning team (200 points), toss winners (100 points), batting or bowling preferences (100 points), score predictions (100 points), top run-scorers (100 x 2 points), top wicket-taker (100 x 2 points), and man of the match (100 points). Predictors can score a maximum of 1000 points individually and 2000 points collectively as a team. Before each match day, predictors will receive an email with a survey link to predict the match outcomes and answer other survey questions, taking about 3-5 minutes to complete.

After each match day, teams will be scored and ranked based on their performance, with the top three teams receiving monetary rewards. In case of ties within the top three ranks, the reward will be shared equally among all tied teams. For example, if two teams tie for 1st place, they will share the combined reward for the 1st and 2nd ranks equally. Similarly, the distribution will follow suit for other rank ties. The tournament rewards the top-performing teams after each match day, with 1st place receiving Rs. 7000, 2nd place Rs. 5000, and 3rd place Rs. 3000. If there's a tie for 1st place, the combined Rs. 10000 (1st + 2nd place) will be divided between the two teams, awarding Rs. 5000 each. This logic applies to all ties within the top three ranks. For example, with three teams tied for 1st place, the total Rs. 15000 (1st, 2nd, and 3rd place) gets split three ways, giving each team Rs. 5000. The same principle applies for ties in 2nd and 3rd place, ensuring a fair distribution of rewards for exceptional performance, even in the case of a tie.

Team leaders, although not participating in outcome predictions, are entitled to a 50% share of the reward if their team ranks in the top three, with the remaining 50% split equally between the two predictors. After each match day, team leaders must decide whether to retain or replace predictors. Replacing a predictor costs the team 100 points, deducted from their next match day score. The team leader has no control over the new predictor, who will be assigned by the experimenter and will be of the same actual gender (unknown to the team leader) but with a randomly assigned avatar visible to the team leader. Each predictor is part of two teams; if replaced by one team leader, the predictor continues to predict for the other team until replaced by that leader as well. The new predictor starts with one team until another leader in a different team replaces a predictor of the same actual gender as the new predictor.

After each match day, team leaders receive an email with a survey link detailing their team's scores (and in T3, additional information on individual performance for some prediction questions), rank, and are asked to decide whether to replace or retain each predictor. This survey takes about 2-3 minutes to complete. Once team leaders have made their decisions, predictors are emailed their team's score, rank, and the leader's decisions. If retained, they receive a prediction survey for the next match day. If replaced, their participation in that team ends, although they may continue in the other team they were assigned to if the other team's leader retains them. Once both leaders have replaced a predictor, their participation ends, and they receive their experimental earnings.

In case of a predictor's non-response on any match day during the experiment, the team score will be based on the remaining participant's responses, with zero scores if neither predictor responds. The non-responding predictor(s) will be replaced by players from a predictors' waitlist for the next match day, maintaining the gender match of the original predictor but with randomly assigned avatars. If a team leader fails to respond on a match day, they are replaced by another male player from the leaders' waitlist. In such instances, if the team ranks in the top three, the reward is evenly distributed among the predictors. All participants will receive detailed game rules and provide informed consent before participation. Additionally, demographic information will be collected to ensure balanced groups across treatments.

Survey with Leaders: Performance Evaluation

At the experiment's conclusion, team leaders will also be provided with the team scores across all four rounds (whenever not zero) and asked to predict which predictor scored more (or whether they scored equally) in each round. Correct predictions will earn them an additional Rs. 25 per match. Ideally, this prediction survey should occur after each round, but we believe it could influence their decision-making process by exposing them to potential contradictions where they perceive equal expected performance but were planning to replace only females. While this may be a good intervention to reduce biases, the real-world context of an ambiguous work environment does not make team leaders explicitly predict the expected performance of each team member; the process is rather more implicit and therefore likely to be subject to bias based on stereotypes. Thus, asking these predictions at the end of each round could introduce an experimenter's demand effect on our primary outcome variable of replace or retain, which may not reflect how they'd react in true real-world circumstances. Hence, we opt to conduct it retrospectively at the experiment's end to avoid such influence.

Survey with Predictors: Performance Evaluation, Retention, and Willingness to Signal Performance

In all four rounds, we survey the predictors concurrently while they predict the match outcomes. Female participants are informed that their team consists of a male predictor (or a female if the participant is male) and a male leader. We ask predictors about their beliefs regarding who they think will perform better in each round. Each correct prediction earns them Rs. 25. We expect that females, in general, may believe male predictors will perform better than themselves, indicating a lack of confidence in their own performance relative to males. We will contrast their beliefs with their actual performances to see whether these beliefs are misinformed or accurately informed.

Next, we use a strategy method to survey the predictors. We ask them to imagine the team leader's likely actions toward them and the other predictor—whether they would be retained or replaced—across different team rank scenarios: 201-450, 101-200, 51-100, 11-50, 4-10, and the top 3. Specifically, we inquire what they believe the leader will decide for themselves and for the other predictor in each scenario, for only one of the two treatment groups to which they have been assigned. To incentivize accuracy, we offer an additional Rs. 25 for correct predictions about both predictors in any round. Despite expecting themselves to perform lower than male predictors in the performance evaluation survey, we believe female predictors will expect to be retained or replaced at the same rate as male predictors by the male team leaders, after controlling for their team's rank. This would indicate an optimism bias (discrepancy between expectations and reality), where female predictors expect fair treatment similar to their male counterparts, which may not align with actual outcomes. Overall, findings from these surveys will help us understand if diversity groups possess an optimism bias.

We believe that female employees' optimism bias is influenced by the adopted work policy (e.g., transparency in performance measurement or reward for retention). Specifically, we expect that, relative to the baseline treatment T1, this optimism bias will decrease significantly in the information treatment T2, where performance will be partially visible to the leader. Female predictors would then expect leaders to base decisions on observed performance, which they may expect to favor male predictors in a male-dominated work environment. However, in the reward treatment T3, we believe this optimism bias will increase significantly relative to the baseline treatment T1, due to females' belief that performance is not visible and that the organization supports and fosters diversity through its policy to reward managers for retention.

Lastly, we will examine whether their optimism bias influences their actions at work. Specifically, we examine whether, when given an opportunity to show their individual performance partially to the leader, female predictors wish to do so, and how this willingness varies across treatments T1 and T3, where the performance is ambiguously observed. We ask predictors a hypothetical question in the second round about whether they'd be willing to pay Rs. 25 from their experimental earnings to reveal their 60% performance to the team leader.[3] We ask this question only in treatments T1 and T3 where performance is ambiguous. We expect that their belief of being the lower performer of the two predictors and optimism bias that they'll be treated the same or better than male predictors under ambiguous performance conditions would make female predictors, relative to male predictors, less likely to reveal such performance information to the leaders. In treatment T3, where there is already a push for diversity through an additional bonus for the leader, they believe the leaders to be even less biased than in treatment T1, and thus, their tendency to pay Rs. 25 to reveal their performance to the experimenter would decrease further in T3.

In treatment T2, where their performance is already 60% visible, we reverse the question and ask them their willingness to pay Rs. 25 to make it completely ambiguous to the team

[3] We will ask this in the 2nd round so that they have a feel for how the game works from the first round.

leader. We believe females would be much more likely to pay Rs. 25 to make performance measurement ambiguous relative to male participants, due to the belief that males are going to be better performers than females.

These findings will create a dilemma for organizations. On one hand, they want to promote diversity-fostering policies, such as increasing visibility of individual performance to leaders or rewarding team leaders to retain diversity, to ensure fair decisions for diverse employees. On the other hand, diversity groups, expecting themselves to be worse performers than male predictors, may oppose policies like treatment T2 that enhance the visibility of their performance, and may even be willing to pay to keep it ambiguous if there is partial visibility as in treatment T2. It will be interesting to see whether their aversion to greater transparency in performance measurement is backed by actual evidence of females performing lower than males, i.e., whether their negative performance evaluation of themselves relative to males is just a pessimistic belief or also supported by their actual performance. In any case, our expected findings, if they hold true, would indicate two problems: one of optimism bias, which may lead them to be underprepared for the real biased decisions of the leaders, and the other of their illintentions to keep performance measurement ambiguous, expecting that less ambiguity will harm them. This may also indicate that females may choose to work in more ambiguous male-dominated work environments, which may potentially harm their prospects due to the expected greater bias by male leaders in such settings.

Statistical Analysis

We first analyze the data as a cross-sectional pooled sample, treating each of the four decisions as independent across all team leaders. We use the following model to explore the impact of the randomly assigned gender of a predictor on the team leader's decision to replace the predictor:

$$Y_{ij} = \alpha_0 + \alpha_1 F_i + \alpha_2 X_i + \alpha_3 L_j + \epsilon_{ij}^0 \tag{1}$$

Here, Y_{ij} represents the “replace” variable, taking a value of 1 if predictor i assessed by team leader j is replaced, and 0 if retained for the next matchday. F_i is the treatment dummy, equaling 1 if the randomly assigned gender of the avatar is female for predictor i and 0 if male. X_i is a vector of predictor-level characteristics, including actual gender, individual performance on the current matchday, and team performance on the current matchday. These controls enable us to isolate the effect of the predictor's assigned gender among other identical individuals with the same actual gender, individual performance, and team performance. L_j includes team leader j's attributes, such as age, education, state of residence, mother's education, father's education, prior STEM/non-STEM background, and family income. ϵ_{ij} denotes the idiosyncratic error term. For robustness, we also estimate this model by clustering the standard errors at the team leader level.

α_1 represents the parameter of interest, estimating the average bias in the leader's assessment of a predictor assigned a female avatar. A positive value of α_1 suggests that leaders are more inclined to replace predictors with a female avatar relative to those with a male avatar, all else being equal, indicating unfair treatment against female avatar predictors.

We will also examine if the expected biases exhibited by male team leaders are influenced by their team's position on the scoreboard and the rewards at stake. We hypothesize that bias will be strongest among teams in the bottom ranks and weakest among teams in the top ranks. This suggests that high-performing teams may not face significant fairness issues against diverse groups, while such issues may be more pronounced in lower-performing teams struggling with profitability.

To further explore the nature of such bias, we incorporate treatment T2, which provides partial visibility into individual performance information to the team leaders, and treatment T3, which offers a personal reward to the leader for retaining their team:

$$Y_{ij} = \beta_0 + \beta_1 F_i + \beta_2 X_i + \beta_3 L_j + \beta_4 T_2 + \beta_5 F_i * T_2 + \beta_6 T_3 + \beta_7 F_i * T_3 + \epsilon_{ij}^1 \qquad (2)$$

In model (2), β_5 estimates how sharing individual performance information influences gender bias, with an expected negative value suggesting a reduction in bias against females as more information is provided about individual performance. Similarly, β_7 estimates how additional personal rewards to the leader for retaining diversity affect bias against females, with an expected negative value indicating the effectiveness of the mitigation strategy. We will estimate all main models using both Ordinary Least Squares (OLS) and Logit models. However, for simplicity, we will present findings from the OLS model in the main text, while results from the Logit model will be provided in the Appendix for reference.

To further understand the underlying mechanism, we use survey responses from team leaders conducted at the end of the experiment where they rated the expected performance of each predictor in their team in each round. We will estimate the following model:

$$Y_{ij} = \gamma_0 + \gamma_1 F_i + \gamma_2 X_i + \gamma_3 L_j + \gamma_4 T_2 + \gamma_5 F_i * T_2 + \gamma_6 T_3 + \gamma_7 F_i * T_3 + \epsilon_{ij}^2 \qquad (3)$$

Here, Y_{ij} is an indicator variable "predictor i rated worse than other predictor" by leader j. A positive value of γ_1 would indicate that, even among predictors with identical actual individual and team performance and the same actual gender, the likelihood of predictor i being rated worse than the other predictor by leader j increases if i has a female avatar. This perceived performance bias could be the reason behind expected biased replacement decisions, and will indicate the existence of statistical discrimination where they replaced female avatars due to the belief that females are low performers in the team. Negative values of γ_5 and γ_7 would indicate whether that the two mitigation strategies reduced the perceived bias against female avatars in their expected performance.

While the above analysis will suggest a general tendency to perceive women avatars as lower performer than male avatars, which may be the cause behind discriminating statistically against the female avatar participants, next we test whether they still replaced a female avatar player even if they believed female player to be same or better performer than the male player, which would be the evidence of taste-based discrimination. To do that, we create a dummy: $D1 =$ "predictor i same or better than other predictor", which takes a value of 1 or 0 depending on how the team leader predicted in the survey. We estimate the following model:

$$Y_{ij} = \delta_0 + \delta_1 F_i + \delta_2 D1 + \delta_3 F_i * D1 + \delta_4 L_j + \epsilon_{ij}^3 \tag{4}$$

Here, Y_{ij} represents the "replace" variable, taking a value of 1 if predictor i assessed by team leader j is replaced, and 0 if retained for the next matchday. We are interested in the estimate of coefficient δ_3. A positive coefficient for δ_3 would suggest that despite believing that predictor i has same or better performance than the other predictor, leaders replaced the one with a female avatar. This would indicate taste-based bias. An significant estimate of δ_3 would suggest that observed bias is statistical in nature, meaning that the leader does not make biased decisions against female avatars when he believes them to be the same or better performers than the other predictor. We will also examine how such coefficient δ_3 changes across different treatments, with an expectation that it will be lowest in the personal reward for retention treatment, where an economic reward may lower the tendency to discriminate on the taste-basis, while the coefficients in the other two treatments (baseline and information) would not be different from each other.

Finally, we will investigate the presence of optimism bias in predictors based on their survey responses during each round using the following model:

$$Y_{ij} = \kappa_0 + \kappa_1 Fem_i + \kappa_2 X_i + \kappa_2 Rank_j + \kappa_4 T_2 + \kappa_5 Fem_i * T_2 + \kappa_6 T_3 + \kappa_7 Fem_i * T_3 + \epsilon_{ij}^4 \tag{5}$$

In this model, Y_{ij} denotes the "replace" variable, which equals 1 if predictor i predicts being replaced by team leader j, and 0 if they predict being retained for the next match day. Fem_i represents the actual gender of predictor i. We are particularly interested in the coefficient κ_1; a non-significant or negative value would indicate optimism bias among females, suggesting they believe their likelihood of being retained is equal to or greater than that of their male counterparts, even after controlling for team rank. This would confirm optimism bias if models (1) and (2) show a bias against females in leaders' replacement decisions. Additionally, positive values of κ_5 and κ_7 would suggest that this optimism bias increases when predictors believe the organization is promoting diversity and inclusion through performance transparency or by rewarding leaders for retention. For robustness, standard errors ϵ_{ij}^4 will be clustered at the predictor i level.

Discussion

Our study aims to provide evidence on the presence of gender biases in team leader decision-making within a simulated team-based environment. We specifically test whether male team leaders are more likely to replace predictors with female avatars than those with male avatars, particularly when performance information is ambiguous. In the baseline scenario, where individual performance details are absent, any bias against female avatars would suggest reliance on gender stereotypes. This would indicate that team leaders might underestimate the expected performance of female employees compared to males in male-dominated settings.

Introducing partial performance information (treatment T2) is expected to reduce bias against female avatars, indicating the presence of statistical discrimination. However, if some bias remains, it would suggest that while statistical discrimination based on performance stereotypes is a factor, taste-based discrimination might also be at play. Any reduction in bias with the introduction of performance data would imply that increased transparency in performance metrics can mitigate discriminatory practices, particularly those driven by statistical bias.

In treatment T3, where a reward for retaining team members is introduced, we expect a further reduction in bias, affecting both statistical and taste-based decisions. In the baseline setting, the cost of replacement is borne by the entire team, similar to organizational structures. However, in T3, team leaders are personally incentivized to retain team members while maintaining team performance. This makes retention more salient compared to simply optimizing team performance and profits, the primary focus in the baseline scenario. This retention incentive is likely to benefit both male and female predictors, with females potentially benefiting more. These findings would suggest that well-structured incentives can promote fair decision-making. By offering additional rewards for retention (and diversity), organizations can encourage team leaders to consider both performance and equity, ultimately reducing the likelihood of biased decisions against female team members.

We also examine whether female predictors perceive themselves as lesser performers compared to males and how these perceptions align with their actual performance on the prediction task. Additionally, we investigate whether females exhibit optimism bias, expecting fair treatment from male leaders—i.e., being replaced or retained at similar rates to male predictors, controlling for individual and team performance. Finally, we explore whether females would support or oppose policies promoting transparency in work measurement. These findings will help us understand the psychology of diversity groups and their decision-making in male-dominated work contexts.

Our study design, featuring random avatar assignment and a controlled environment, allows us to isolate the effects of gender bias from other external factors. More broadly, our study highlights the impact of gender stereotypes in team-based performance evaluations and demonstrates the potential of performance transparency and incentive-based interventions to promote diversity and equity. These measures can mitigate the adverse effects of gender biases and contribute to more effective and inclusive team dynamics. However, the simulated nature of the task may limit the direct applicability of our findings to real-world settings. Future research should investigate similar dynamics in actual workplaces to validate and expand upon our conclusions.

References

Banerjee, A., Bertrand, M., Datta, S., and Mullainathan, S. (2009). Labor market discrimination in delhi: Evidence from a field experiment. *Journal of comparative Economics*, 37(1):14–27.

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

Behaghel, L., Crépon, B., and Le Barbanchon, T. (2015). Unintended effects of anonymous resumes. *American Economic Journal: Applied Economics*, 7(3):1–27.

Benson, A., Li, D., and Shue, K. (2019). Promotions and the peter principle. *The Quarterly Journal of Economics*, 134(4):2085–2134.

Benson, A., Li, D., and Shue, K. (2024). Potential and the gender promotions gap. *Available at SSRN*.

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

Bohnet, I., Van Geen, A., and Bazerman, M. (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science*, 62(5):1225–1234.

Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the swedish labor market? *Feminist Economics*, 17(3):71–102.

Chang, F., Luo, M., Walton, G., Aguilar, L., and Bailenson, J. (2019). Stereotype threat in virtual learning environments: Effects of avatar gender and sexist behavior on women's math learning outcomes. *Cyberpsychology, Behavior, and Social Networking*, 22(10):634–640.

Charles, K. K. and Guryan, J. (2011). Studying discrimination: Fundamental challenges and recent progress. *Annu. Rev. Econ.*, 3(1):479–511.

Cleveland, J. N., Stockdale, M., Murphy, K. R., and Gutek, B. A. (2000). *Women and men in organizations: Sex and gender issues at work*. Psychology Press.

Crone, C. L. and Kallen, R. W. (2022). Interview with an avatar: Comparing online and virtual reality perspective taking for gender bias in stem hiring decisions. *PloS one*, 17(6):e0269430.

Egan, M., Matvos, G., and Seru, A. (2022). When harry fired sally: The double standard in punishing misconduct. *Journal of Political Economy*, 130(5):1184–1248.

Foley, M. and Williamson, S. (2018). Does anonymising job applications reduce gender bias? understanding managers' perspectives. *Gender in Management: An International Journal*, 33(8):623–635.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American economic review*, 90(4):715–741.

Heilman, M. E., Block, C. J., and Lucas, J. A. (1992). Presumed incompetent? stigmatization and affirmative action efforts. *Journal of applied psychology*, 77(4):536.

Hekman, D. R., Johnson, S. K., Foo, M.-D., and Yang, W. (2017). Does diversity-valuing behavior result in diminished performance ratings for non-white and female leaders? *Academy of Management Journal*, 60(2):771–797.

Huang, R., Mayer, E. J., and Miller, D. P. (2024). Gender bias in promotions: Evidence from financial institutions. *The Review of Financial Studies*, 37(5):1685–1728.

Krause, A., Rinne, U., and Zimmermann, K. F. (2012). Anonymous job applications in europe. *IZA Journal of European Labor Studies*, 1:1–20.

Kravitz, D. A. and Platania, J. (1993). Attitudes and beliefs about affirmative action: Effects of target and of respondent sex and ethnicity. *Journal of applied psychology*, 78(6):928.

Leslie, L. M. (2019). Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review*, 44(3):538–563.

Lopez, S., Yang, Y., Beltran, K., Kim, S. J., Cruz Hernandez, J., Simran, C., Yang, B., and Yuksel, B. F. (2019). Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*, pages 1–12.

MacNell, L., Driscoll, A., and Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291–303.

Massey, C., Simmons, J. P., and Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, 22(2):274–281.

Mengel, F. (2021). Gender bias in opinion aggregation. *International Economic Review*, 62(3):1055–1080.

Milkman, K. L., Akinola, M., and Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological science*, 23(7):710–717.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the national academy of sciences*, 109(41):16474–16479.

Peter, L. J., Hull, R., et al. (1969). *The peter principle*, volume 4. Souvenir Press London.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review*, 62(4):659–661.

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, 107(5):141–145.

Waldman, M. (2003). Ex ante versus ex post optimal promotion rules: The case of internal promotion. *Economic Inquiry*, 41(1):27–41.