Persistence of discrimination: Theory and experimental evidence

Nishtha Sharma¹²

September 2021

Recent studies on the dynamics of discrimination find that the bias against unfavored groups declines after they succeed against the favored groups. Yet, the possibility that discrimination can be overcome may not go unnoticed by favored group members. In this paper, I model dynamic competition between favored and unfavored group members to show that in a dynamic setting, if discrimination can be overcome, then the favored group’s members have an incentive to increase their efforts to prevent a loss of privilege. My theory and experimental evidence reveal that dynamic bias will generally serve to exacerbate discrimination rather than eradicate it. Finally, I propose taxrebates as a policy solution to mitigate discrimination when bias can be overcome.

¹University of California, Irvine. Contact: nishths@uci.edu
²†This paper has benefited from comments by Jean-Paul Carvalho, John Duffy, Meera Mahadevan, Michael McBride, Stergios Skaperdas, and Stephan Wolton and the conference audience at the Econometric Society’s Winter Conference in Delhi School of Economics 2020, Southern Economic Association 2020, Economic Science Association Global Conference 2020, and Contests: Theory and Evidence Conference 2020. I am thankful for generous grants from the Center for Global Peace and Conflict Studies, the Initiative to End Family Violence Fellowship, Kugelman Citizen Peacebuilding Research Fellowship, and the Department of Economics Graduate Student Fellowship at the University of California, Irvine
1. Introduction

Identity-based discrimination is known to put unfavored groups such as women, people of color and immigrants at a persistent or cumulative disadvantage at all the levels of economic and social participation (Blank (2005); Stolzenberg et al. (2013); Gandy (2016)). While discrimination is still prevalent, recent studies have found that unfavored groups can overcome bias. Bohren et al. (2019), Mengel et al. (2019), Fryer Jr (2007), Groot & Van Den Brink (1996) and Lewis (1986) show that when people from unfavored groups succeed initially, decision-makers update their beliefs about unfavored types’ ability, and bias against them declines for future evaluations. The initial success of unfavored groups also grants exposure to them which changes biased attitudes and causes a subsequent decline in bias (Beaman et al. (2009), Boisjoly et al. (2006)). This paper takes the next logical step in understanding the dynamics of discrimination by shifting the focus from decision-makers to the people competing in biased settings. Using theory and experimental evidence, it analyzes how the dynamics of bias create new incentives that lead members of favored groups to prevent the reduction of bias.

Under biased competition, the unfavored type has a smaller chance of winning when efforts and abilities are equal. When the bias is constant, it remains the same across periods irrespective of the initial winner. But when bias is dynamic, unfavored groups can overcome the bias if they initially win. The possibility of overcoming bias creates strategic considerations for members of both types who choose how much effort to exert while competing against each other. Whether the unfavored type’s opportunity to gain fairness outweighs the favored type’s threat of losing favor ultimately determines if the disparity in chances of winning declines or further increases due to dynamic bias. The theoretical model predicts that each type exerts a higher effort under dynamic bias than constant bias. Moreover, the favored type’s increase in effort due to dynamic bias is higher than the unfavored type’s increase in effort. Hence, the favored type has a higher probability of winning in the first period when bias is dynamic and discrimination persists.

As it is difficult to measure and control changes in bias in the field, I empirically test

---

3The model is based on the framework of dynamic contest theory (Konrad (2009))
the theoretical predictions using a laboratory experiment. Participating subjects were randomly assigned the type (favored and unfavored) and anonymously matched into pairs to compete in a two-period competition with and without the possibility of overcoming bias. Types were assigned using neutral language - A for the initially favored and B for the initially unfavored. Subjects chose their bids (as a proxy for effort choices) out of their endowment to influence their probability of winning the prize. The more they bid, the higher were their chances of winning a fixed prize. In the first period of the repeated contest, A was thrice as likely to win the prize than B for the same bid. There is no favoritism in period two if the unfavored type B wins in period one of the dynamic bias treatment while the period-two favor remains the same irrespective of the initial winner in the constant bias treatment.

Consistent with the theoretical predictions, I find that both types make higher bids when discrimination can be overcome. However, the increase is significantly higher for the favored type (4.5 standard deviation) than the unfavored type (1.5 standard deviation). Thus, the threat of losing favor creates a stronger incentive than the possibility of attaining fairness. This difference is economically large and statistically significant. The unfavored types win in 27% of the period-one competitions when bias is constant but only 22% of the period-one competitions when bias can be overcome. Moreover, the effect of dynamic bias on unfavored type’s winning probability in period two is small and insignificant: 23% in the case of constant bias as opposed to 25% in the case of dynamic bias. Thus, the unfavored type wins less often in period one and no more often in period two when bias is dynamic rather than constant.

As dynamic bias cannot mitigate discrimination, I extend my theory and the experimental analysis to examine two possible conditions under which one may expect it to cause a decline in discrimination. First, when bias can not only be overcome but reversed by their initial success Bohren et al. (2019); Fryer Jr (2007) to subsequently favor the originally unfavored type, and second when the favored type cares sufficiently lesser about the future than the unfavored type so that the threat of losing favor in future becomes weaker. Even under reversible bias, the unfavored type does not have a higher chance of winning than constant bias. I derive a relevant threshold of
relative future concern, i.e., the ratio of the favored type’s future concern to that of the unfavored type. If the relative future concern is below this threshold, the effect of surmountable bias on the favored type’s effort is less than the unfavored type, and discrimination declines. While there is global evidence of identity-based heterogeneity in concern for future welfare and patience Falk et al. (2018), the threshold requires the favored candidate to care substantially lesser about the future, which is rare to find. Thus, the dynamic nature of discriminatory bias is unlikely to mitigate discrimination by itself, and the need for policy intervention remains. The next project discusses persistent discrimination despite weak or low biases from the employer’s perspective.

Two often considered policy tools to tackle discrimination are tax rebates for the unfavored group and affirmative action. Initial tax rebates will increase the payoff from winning for the unfavored type leading to their higher incentive and effort to win initially, allowing the bias to decline. Similarly, affirmative action or quota for the unfavored group will help overcome bias, especially if it declines due to exposure to the unfavored group. Whether the decision makers’ implicit bias also increases with quotas to compensate for utility loss from a forced selection of the unfavored type remains a relevant question to determine the effectiveness of quotas. On the other hand, tax rebates are not personally costly to decision-makers such as biased employers because the tax rebate expense is borne by governments. Thus tax rebates are less likely to distort the evaluator’s choices and can correct for asymmetric incentives created by surmountable bias.

This paper advances the economic literature on discrimination from the finding that bias is dynamic to understanding the response behavior of people competing in dynamically biased settings and proposing policy solutions to make dynamic bias effective in overcoming discrimination (Coate & Loury (1993); Fryer Jr (2007); Bohren et al. (2019)). The paper offers theoretically-informed experimental evidence that historical biases would persist without policy correction, even in dynamic environments. It also provides testable theoretical predictions on the comparative effectiveness of policies. Moreover, the paper adds to our understanding of the interaction of socio-cultural institutions such as discrimination and inter-group competition. We typically consider discriminatory biases to be like slow-moving social institutions (Roland (2004)), resulting in
inequality, slow growth, and development. I show that even when bias can decline, equality in opportunities is hindered by the incentives emerging from the use of competition between people to allocate economic gains such as job positions. Further, by teasing out the effect of relative future concerns on the persistence of bias, it contributes to the growing literature that explores the relationship between culture, patience, and economic outcomes (M. K. Chen (2013); Hübner & Vannoorenberghe (2015)). Finally, by working in the framework of contests, it contributes to the widely applicable literature on dynamic theory (Konrad (2009)) and experiments (Dechenaux et al. (2015)) in contest theory.

The rest of the paper is structured as follows. Section 2 presents the model of reducible and reversible bias while allowing for heterogeneity in future concern among favored and unfavored groups and gathers the main theoretical results. Section 3 describes the experimental design to test the theoretical predictions. Section 4 presents empirical findings based on the laboratory experiment. Section 5 discusses and compares the potential policy tools to mitigate discrimination and concludes with potential extensions.

2. The Model

Consider two agents, A and B who exert efforts or spend resources to win a lottery contest (Konrad (2009); Tullock (1967)). Let A’s effort bid be denoted by $e_a$ and B’s effort bid be denoted by $e_b$. Let the designations A and B denote the combination of agents’ identity (such as gender) and ability (cost of effort). However, to focus on identity based discrimination, I assume that the agents have equal ability and the marginal cost of effort is constant and equal to 1 for both A and B. The payoff from winning the contest is given by $V$ (such as the after-tax wage from getting hired). For simplicity, let us assume that both A and B are risk neutral.

---

4While this paper is focused on the case of discrimination, the model and experimental results are more generally applicable to any economic context where competition is used as a tool to allocate economic gains between people in asymmetric positions due to any cause such as bias, ability, experience and confidence. For example, in war, a small country may be less likely to win against a big country but if it does, then resource transfer and experience will increase its likelihood to win again.

5Assuming convex cost does not change the qualitative results.

6The effect of non-neutral risk preferences is ambiguous (Sheremeta (2011)).
There is a priori bias in favor of type A such that for equal effort, A has a higher chance of winning than B. Define $f$ as the intensity of initial bias in favor of A such that $f > 1$. Please refer to the appendix for a richer micro-founded derivation of the favor $f$. Higher the $f$, higher is the initial bias in favor of A or, higher the difference between $f$ and 1, higher is the initial bias in favor of A. Thus, when A exerts effort $e_a$, her effective effort is $fe_a$. Effective effort of B is the same as her effort $e_b$. Players’ probability of winning is given by their relative effective effort, following the lottery Contest Success Function\(^7\). The following asymmetric contest success function describes the relationship of players’ efforts to their probability of winning in period 1.

$$P_a(e_a, e_b), P_b(e_a, e_b) = \begin{cases} \frac{fe_a}{fe_a + e_b}, \frac{e_b}{fe_a + e_b} & \text{if } \max\{e_a, e_b\} > 0 \\ \frac{f}{f + 1}, \frac{1}{f + 1} & \text{if } e_a = e_b = 0 \end{cases}$$

Candidates choose the effort that maximizes their expected payoff with respect to the bias they face and their ability. Thus, optimal $e_a, e_b$ maximize $U_a, U_b$ given $f$ and $V$, i.e., $e_a, e_b$ maximize:

$$U_a = \frac{fe_a}{fe_a + e_b}V - e_a$$

and

$$U_b = \frac{e_b}{fe_a + e_b}V - e_b$$

The bias in the second period depends on the outcome of the first period. The level of bias goes down or reverses if and only if the unfavored type B wins in period 1. It remains the same if the favored type A wins in period 1.

$$f_2 = \begin{cases} \alpha f_1 = \alpha f & \text{if B wins in period 1 and bias is reducible or reversible} \\ f_2 = f & \text{if A wins in period 1} \end{cases}$$

\(^7\)A Contest Success Function (CSF) provides each player’s probability of winning as a function of all players’ efforts (Skaperdas, 1996).
where $\alpha \in (0, 1)$ and represents the degree of reducibility of bias. Higher $\alpha$ means that the bias cannot be easily overcome or surmounted. Bias is reversible when $\alpha f < 1$ and reducible when $\alpha f \geq 1$. Let $\beta_a > 0$ and $\beta_b > 0$ denote A and B’s concern for future welfare. Following Falk et al. (2018)’s evidence on heterogeneous patience, I allow for $\beta_a \neq \beta_b$. Further, $\beta_i$ can be $> 1 \forall i \in a, b$ such that period 2 represents aggregated future which may be considered more important than the present.

Reducible and reversible bias create asymmetric incentives for A and B to win in period 1. When B wins, the bias falls, and B’s probability of winning in period 2 increases. But, it also means that the second period contest will be more symmetric unless the reversal causes more than initial bias in the opposite direction. The equilibrium efforts in a contest are increasing in symmetry as shown in the appendix and found previously (Konrad (2009)). Therefore, increased symmetry due to reducible bias creates a disincentive for the unfavored type to win in period 1 and indeed overcome the bias. On the other hand, when A wins, favorable bias for A as well as asymmetry in contest remain intact. The equilibrium effort in second period will be smaller for both types if A wins in period 1 and higher if B wins in period 1. Reducible bias creates a higher incentive to win in period 1 for A.

As a result, A incurs greater effort than B in period 1 and wins with greater probability than it would without reducible bias. However, if A has sufficiently lower future concern than B, then A’s additional incentive to win in period 1 almost disappears. In that case, B outbids A in the first period. Therefore, reducible bias increases A’s initial probability of winning and the probability of bias reduction is even smaller than initial bias, unless $\beta_a$ is sufficiently smaller than $\beta_b$. When bias is reversible, then the channel of asymmetry becomes less important as B’s initial win also keeps the contest asymmetric due to favorable bias towards B. Therefore, the fall in B’s initial probability of winning is less under reversible bias than reducible bias. However, in both the cases, discrimination persists unless A’s future concern is smaller enough.

**Proposition 1.** When the players are forward-looking and have positive future concern given by
\( \beta_a \) and \( \beta_b \) for A and B respectively and assumptions 1 and 2 are given, then:

1. In period 1,

   a) Equilibrium effort of the favored type A exceeds that of B, unless, A’s future concern is sufficiently lower than B’s future concern. In other words, if \( \beta_a > \gamma \beta_b \), then \( e_{1a} > e_{1b} \) and if \( \beta_a < \gamma \beta_b \), then \( e_{1a} < e_{1b} \), where \( \gamma = \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} < 1 \)

   b) Discrimination against B and B’s probability of winning \( P_{1b} \) is decreasing, U-shaped, and increasing in \( \alpha^{-1} \) if \( \beta_a > \beta_b \), \( \beta_a \in (\gamma \beta_b, \beta_b) \) and \( \beta_a < \gamma \beta_b \) respectively where \( \gamma = \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} < 1 \).

2. In period 2, A and B exert equal effort which is decreasing in \( f_2 \)

Proof of this and the following proposition is in appendix A.

Proposition 2 demonstrates another consequence of reducible and reversible bias. It is not only that the short run or initial outcomes like the realized probability of winning for the unfavored types are poor under reducible and reversible bias, but also that the unfavored type has to exert even higher effort to reach a lower probability of winning than they would under constant bias.

**Proposition 2.** When the players are forward-looking and have positive future concern given by \( \beta_a \) and \( \beta_b \) for A and B respectively then in period 1:

a. Equilibrium effort of both types is higher under reducible as well as reversible bias as compared to constant bias.

b. The increase in A’s effort is higher than the increase in B’s effort due to reducible bias unless A’s future concern is sufficiently smaller than B i.e., \( \beta_a < \gamma \beta_b \) where, \( \gamma = \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} < 1 \).

Next section presents the experimental design to test the theoretical predictions of the model.
3. Experimental Design and Procedure

The experiment is designed to study the impact of dynamic bias in a dynamic lottery contest. I also empirically test if dynamic bias is effective in improving unfavored groups’ chances of winning (i) if bias is reversible (ii) if the favored type A cares sufficiently lower about the future than the unfavored type B (i.e., $\beta_a < \gamma \beta_b$) I implement this with a between-subjects design and use the parameters as summarized in the treatment table 2 below. Each cell denotes the combination of initial favor towards A ($f_1$), period-two favor towards A if B had won in period-one ($f_2$), A’s future concern ($\beta_a$) and B’s future concern ($\beta_b$). The payoff from winning is 100 points for both types, and the period-two level of favor towards A is always equal to the period-one level of favor when A had won in period one.

Table 1: Treatment table

<table>
<thead>
<tr>
<th>$(f_1, f_2$ if B wins in period 1, $\beta_a, \beta_b$)</th>
<th>Constant Bias</th>
<th>Reducible bias</th>
<th>Reversible bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_a = \beta_b$</td>
<td>(3,3,5,5)</td>
<td>(3,1,5,5)</td>
<td>(3,1/2,5,5)</td>
</tr>
<tr>
<td>$\beta_a = 0.02\beta_b$</td>
<td>(3,3,0,5)</td>
<td>(3,1,0,5)</td>
<td></td>
</tr>
</tbody>
</table>

As $\gamma < 1$, I specify the second case with $\beta_a = \beta_b$. The choice of equal $\beta$s for case (ii) also has an intuitive appeal for contexts where we do not have a reason to believe that the favored and the unfavored types will differ in their patience or future concern. Moreover, it allows me to compare the results in this paper with the findings of other papers on the dynamics of discrimination. In the baseline treatment, bias is insurmountable or ‘Constant’ i.e., it cannot be overcome.

The chosen parameter values of V and f contextualize the data with the existing literature. The choice of $\beta$s is to ensure a clear difference in predicted efforts under different treatments. A higher than 1 value of $\beta$ is justified with the interpretation of value from period 2 as the reduced form value from all the future periods. The choice of $f_2$ is intuitively appealing and easy to apply in the lab as it means that the bias is reinforced in the second period when A wins in the first period and completely goes away when B wins in the first period.
Table 2: Theoretical Predictions of bids in period 1

<table>
<thead>
<tr>
<th>Bids - ((e_{1a}, e_{1b}))</th>
<th>Constant Bias</th>
<th>Reducible bias</th>
<th>Reversible bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_a = \beta_b)</td>
<td>18.75, 18.75</td>
<td>41.31</td>
<td>38.33</td>
</tr>
<tr>
<td>(\beta_a = 0.02\beta_b)</td>
<td>18.75, 18.75</td>
<td>24.4, 45.8</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2 describes the theoretical predictions of the bids in period 1 of the favored and the unfavored types \((e_{1a}, e_{1b})^8\). Both types are expected to bid higher under reducible bias and reversible bias as compared to constant bias. The increase in bids due to reducible bias is higher for type A than type B if relative \(\beta = 1\) and lower if relative \(\beta = 0.02\). When \(\beta_a = \beta_b\), then period 1 discrimination increases due to reducible bias and when \(\beta = 0.02\beta_b\), then period 1 discrimination reduces due to reducible bias. There are three key hypotheses regarding the players’ bids in the first period:

Hypothesis 1. When \(\beta_a = \beta_b\), each type bids higher under reducible bias than constant bias but the increase is greater for the favored type

Hypothesis 2. When \(\beta_a = \beta_b\), each type bids higher under reversible bias than constant bias but the increase is greater for the favored type

Hypothesis 3. When \(\beta_a = 0.02\beta_b\), each type bids higher under reducible bias than constant bias but the increase is greater for the unfavored type

An experimental session consisted of 4 parts followed by puzzles and demographic survey. Part 1 measured the players’ baseline behavior in contests and consisted of 5 rounds of fair or unbiased symmetric contest where each round consisted of two periods. Part 2 measured attentiveness with a simple task of setting maximum number of sliders to given numbers between 0 and 100, in a minute. Part 3 was the main treatment task and comprised of 10 rounds of the biased contest where each round consisted of two periods. Part 4 was one round of a single-period of unbiased contest for a prize worth 0 to measure subjects’ innate preference for winning regardless

---

^8The theoretical predictions and subsequent findings pertaining to period 2 are in appendix.
of the prize. The sessions ended with puzzles such as the cognitive test, Holt & Laury (2002) risk preference lottery task, Hanoi task to measure far-sightedness, and a demographic survey. All the parts are explained in detail below.

The experiment was programmed with the software oTree (D. L. Chen et al. (2016)) and executed online with undergraduate students at a public university in the US. 328 subjects participated in the study with about 65 per treatment. Each individual participated in only one session, and no subject knew anything about this project or had any experience in participating in a similar experiment. There were several quizzes added to ensure alertness and comprehension of instructions. As per the experimental design, the subjects were randomly assigned to different treatments and roles (A or B) and randomly and anonymously matched into pairs of A and B. In each round, they played two periods of a contest with a constant endowment of 100 points and winning prize of 100 points in each period. They chose how much they want to bid out of 100 to influence their probability of winning and kept the rest. To invoke favor while using neutral language, I implemented the following novel design. Subjects were informed that A’s bid is multiplied by 3 while B’s bid is multiplied by 1 to determine the number of A and B type balls put into the bid bag. One ball was drawn at random from the bid bag to determine the winner. This rule changed in period 2 of the reducible bias treatment such that A’s bid was also multiplied by 1 if and only if B had won in period 1. The rule also changed in period 2 of the reversible bias treatment such that B’s bid was multiplied by 2 and A’s bid by 1 if and only if B had won in period 1.

A and B’s future concern were invoked in the following way. In the case of equal future concern, the total payoff for each type is payoff in period 1 plus 5 times the payoff in period 2. In the case of unequal future concerns, the total payoff for type A is payoff in period 1 plus 10 percent of the payoff in period 2, while the total payoff for type B is payoff in period 1 plus 5 times the payoff in period 2. The parameters were chosen to ensure large enough difference in point prediction of bids under different treatments. Moreover, a larger than 1 future concern is relevant while thinking of period 2 as the aggregate future. Each session consists of 10 rounds of the repeated asymmetric
contest with anonymous and random re-matching with someone else of the opposite type in each round. One round was then picked at random for actual payment ($1 = 100 points). Every subject engaged in 5 rounds of repeated symmetric contest and a simple effort task to measure attentiveness before the main part of the 10 rounds of the repeated asymmetric contest. The screenshots and video instructions for a typical experimental session under the reducible bias treatment are available in the online appendix.

The decision screen provided information about the subject’s type, the value of the prize and own, and other type’s balls per point of bid. It also had sliders and corresponding input box fields for subjects to submit the prediction of their match’s bid and their bid. They had the choice of typing the bids or moving the slider (both were synchronized). Subjects could see what they keep for each bid. They also saw their probabilities of winning and losing, their own, and their match’s round payoff under each case. The probability of bias remaining the same or changing for every combination of predicted opponent’s bid and own bid was also reported. The subjects were free to move the sliders and try as many combinations as they wanted in each period. The decision screen also has a reminder box which states that the total payoff in a round for both types is \([\text{Payoff in period } 1 + (5 \times \text{Payoff in period } 2)]\) (under equal future concerns) and similarly for the case of unequal future concerns.

At the end of each round, a results screen informed the subjects whether they had won or not, their bid, their match’s bid, their prediction, their prediction payoff, bid task payoff, and the total payoff in that round. The current round number was notified at the top of each decision and results screen. At the end of each round, a history page was displayed, which informed the subjects of their bid, their match’s bid, and their payoff in round 1 and 2 of all the previous rounds. I used the random lottery payment mechanism to ensure that the subjects treat every round as a separate task, and the stakes are not distributed over rounds. Even though this relies on the assumption that subjects are expected utility maximizers, which also affects the main question of interest in this paper, Hey & Lee (2005) show that under random lottery payment mechanism subjects do answer as if they were separating the tasks over rounds. At the end of part 1, one of the last five rounds
was selected at random for payment and subjects were told their earnings from part 1.

The next part was also a bidding task with only one round of 1 period. Subjects are given an endowment of 50 points. They could choose how much to bid for a prize of 0 points. There were no types, i.e., it is a symmetric one-shot contest. They were told that their chances of winning are given by their share in the total bid by them and their match and that they will be told if they won or lost, but their bid will be forfeited irrespective of the outcome. This part was to measure subjects’ preference for winning for its own sake as they have to incur costs for no prize. Experiments in contests often find overbidding relative to the Nash predictions. Preference for winning is an important reason for overbidding. Bids in a contest for prize worth 0 as a measure of preference for winning was described in Sheremeta (2010).

This was followed by a survey that consists of the standard cognitive reflection test to measure ability, incentivized tower of Hanoi task to measure the inclination and ability to do backward induction and a demographic survey and a modified Holt & Laury (2002) lottery task to measure risk preference. Before the main treatment, I conduct 5 rounds of symmetric repeated contest for benchmark bidding behavior so that I can tease out the effect of reducible and reversible bias in asymmetric contest. A simple slider task was also added to ensure and measure alertness. Payments consist of the accumulated earnings throughout the experiment. 100 points were equivalent to 1 USD (this is informed in the welcome sheet and video). Each session lasted about 60 minutes, and the subjects’ average payment was 20 USD.

4. Experimental Results

I first compare the effect of reducible bias on the first-period bids of the favored type A and the unfavored type B. Recall that with equal bids, A is three times more likely to win the contest in period 1 in equilibrium. Contest becomes fair and symmetric in period 2 if B had won in period 1 and bias is reducible. If bias is reversible and B had won in period 1, then the contest becomes biased towards B in period 2 such that B is twice more likely to win the second period contest, with equal bids. Examining how reducible bias and reversible bias affect the bids enables me to
ascertain their effect on discrimination. Next, I present results on the effect of reducible bias when the favored type A has a sufficiently lower future concern than B.

**Equal future concerns**

In what follows, I restrict the sample to the case of equal future concern for both the favored and the unfavored type. I first compare the case of reducible bias with constant bias followed by a comparison of reversible bias with constant bias. Column (1) of Table 3 shows that when bias is constant, regressing the players’ first-period bid on type, controlling for their bids in the symmetric contest (when there was no bias), reveals no significant difference in bids based on type. Figure 1 illustrates this result by showing the average period-1 bids by type under constant bias. Column (2) of Table 3 repeats the analysis for the case when bias is reducible. It shows that when bias is reducible, the favored type bids significantly higher than the unfavored type as also illustrated in figure 1. Next I present the regression results for constant and reducible bias within the same model to ascertain the effect of reducible bias on each type. I regress the average first period bids on dummies corresponding to type (favored or unfavored), Reducible bias (1 if yes, 0 if constant),

![Figure 1](image.png)

*Figure 1.* Comparison of the favored and unfavored types’ period 1 bids under equal future concerns.
the interaction of type and reducible bias and the average bid in symmetric contest⁹ (Table 3, column 3). There is a significantly higher increase in first period bids of the favored type relative to the unfavored type due to reducible bias: the interaction effect between type and reducible bias is positive and significant. This implies that reducible bias incentivizes the favored type more than the unfavored type. Taken together, these results suggest that when bias can be overcome by success of the unfavored type, the likelihood of their first period success becomes even smaller as the likelihood of success is a function of bias as well as relative effort or bids. The threat of losing favor is stronger than the possibility of gaining fairness. This creates a discrimination trap when both types have equal patience or future concern.

Next, I present the results examining the effect of reversible bias. Recall that the initially unfavored type B gets a favorable bias of 2 in the second period if and only if they had won in the first period. There is a smaller change in asymmetry under reversible bias than reducible bias, thereby increasing the importance of the threat of losing and the opportunity to gain favorable bias by the two types. Column (2) of Table 4 shows that when bias is reversible, regressing the players’ first-period bid on type, controlling for their bids in the symmetric contest (when there was no bias), reveals a positive but not significant difference in the favored and the unfavored type’s first period bids as also illustrated in figure 1. In column 3 of table 4, I present the regression results for constant and reversible bias within the same model to get at the effect of reversible bias on each type. I regress the average first period bids on dummies corresponding to type (favored or unfavored), Reversible bias (1 if yes, 0 if constant), the interaction of type and reversible bias and the average bid in symmetric contest. There is a higher increase in first period bids of the favored type relative to the unfavored type due to reversible bias: the interaction effect between type and reducible bias is positive but not significant. While, the standard errors (clustered at the session level) are not small for the case of reversible bias, the regression coefficients suggest that reversible bias also fails to increase the initial likelihood of winning of the unfavored type B. Thus, when bias is reversible, the actual likelihood that reversal will be observed depends on the initial level of bias.

⁹I also collected data on risk preference, attentiveness, cognitive reflection test and backward induction test. These have no explanatory power in the data and the results do not change when I control for them.
Table 3: Equal future concerns - Effect of reducible bias on first period bids of the favored and unfavored types

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant Bias</th>
<th>Reducible Bias</th>
<th>Constant and Reducible Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Favored</td>
<td>1.58</td>
<td>11.65***</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>(4.95)</td>
<td>(1.92)</td>
<td>(4.9)</td>
</tr>
<tr>
<td>Reducible</td>
<td></td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(6.7)</td>
</tr>
<tr>
<td>Favored × Reducible</td>
<td></td>
<td></td>
<td>10.22*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.07)</td>
</tr>
<tr>
<td>Bid in Symmetric</td>
<td>0.57**</td>
<td>0.52***</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.10)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Prediction of other’s bid</td>
<td>0.19</td>
<td>0.38***</td>
<td>0.30***</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Intrinsic preference for winning</td>
<td>0.49***</td>
<td>0.07</td>
<td>0.30***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.21)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.01*</td>
<td>-2.45</td>
<td>-7.8**</td>
</tr>
<tr>
<td></td>
<td>(4.14)</td>
<td>(5.14)</td>
<td>(3.07)</td>
</tr>
<tr>
<td>Observations</td>
<td>80</td>
<td>70</td>
<td>150</td>
</tr>
</tbody>
</table>

Notes: (1) * p-value ≤ 0.1; ** p-value ≤ 0.05; *** p-value ≤ 0.01
(2) Standard errors from OLS regressions are clustered at session level and reported in parentheses
(3) Favored = 1 if favored type, 0 otherwise
(4) Reducible = 1 if bias is reducible, 0 if bias is constant
(5) Intrinsic preference for winning is the bid in the contest for 0 prize

only. The possibility of gaining favorable bias does not incentivize people more than the threat of losing it. Comparing this result with the finding on the effect of reducible bias suggests that when each type has the same future concern, positive discrimination from reversible bias is likely to be achieved but fairness from reducible bias is not.

Unequal future concerns - Sufficiently lower future concern of the favored type relative to the unfavored type

Now, I restrict the sample to the case of sufficiently lower future concern of the favored type relative to unfavored type (0.02) and analyze the effect of reducible bias on the first period bids
Table 4: Equal future concerns - Effect of reversible bias on first period bids of the favored and unfavored types

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant Bias (1)</th>
<th>Reversible Bias (2)</th>
<th>Constant and Reversible Bias (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favored</td>
<td>1.58 (4.21)</td>
<td>16.69* (8.27)</td>
<td>1.74 (4.84)</td>
</tr>
<tr>
<td>Reversible</td>
<td></td>
<td></td>
<td>1.51 (9.43)</td>
</tr>
<tr>
<td>Favored × Reversible</td>
<td></td>
<td></td>
<td>9.5 (10.2)</td>
</tr>
<tr>
<td>Bid in Symmetric</td>
<td>0.57*** (0.13)</td>
<td>0.20 (0.14)</td>
<td>0.43*** (0.13)</td>
</tr>
<tr>
<td>Prediction of other’s bid</td>
<td>0.19 (0.09)</td>
<td>0.18 (0.31)</td>
<td>0.39** (0.19)</td>
</tr>
<tr>
<td>Intrinsic preference for winning</td>
<td>0.49*** (0.10)</td>
<td>0.07 (0.06)</td>
<td>0.29*** (0.09)</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.01 (4.14)</td>
<td>20.3 (21.6)</td>
<td>-4.4 (8.2)</td>
</tr>
<tr>
<td>Observations</td>
<td>80</td>
<td>62</td>
<td>142</td>
</tr>
</tbody>
</table>

Notes: (1) * p-value ≤ 0.1; ** p-value ≤ 0.05; *** p-value ≤ 0.01
(2) Standard errors from OLS regressions are clustered at session level and reported in parentheses
(3) Favored = 1 if favored type, 0 otherwise
(4) Reversible = 1 if bias is reversible, 0 if bias is constant
(5) Intrinsic preference for winning is the bid in the contest for 0 prize

of the two types. Column (1) of Table 5 shows that when bias is constant, regressing the players’ first-period bid on type, controlling for their bids in the symmetric contest reveals no significant difference in the bids of the favored and the unfavored types. Figure 2 illustrates this finding by showing the average period-1 bids by type under constant bias. Column (2) of Table 5 repeats the analysis for the case when bias is reducible. It shows that when bias is reducible and the favored type has a sufficiently smaller future concern than the unfavored type, then the favored type bids significantly lower than the unfavored type as also illustrated in figure 2.

Bringing the constant and reducible bias within the same model I again regress the average
Note: Error bars represent 90% confidence intervals.

*Figure 2.* Comparison of the favored and unfavored types’ period 1 bids under unequal future concerns. The figure shows a bar chart comparing the average bids in Period 1 under constant and reducible bias. Error bars represent 90% confidence intervals.

First period bids on dummies corresponding to type (favored or unfavored), Reducible bias (1 if yes, 0 if constant), the interaction of type and reducible bias and the average bid in symmetric contest only for the sub-sample with unequal future concerns. (Table 5, column 3). There is a significantly higher increase in first period bids of the unfavored type relative to the favored type due to reducible bias: the interaction effect between type and reducible bias is negative and significant. This implies that reducible bias incentivizes the unfavored type more than the favored type in this case. This is in contrast with the findings in the case of equal future concerns. When bias can be overcome by success of the unfavored type, the likelihood of her initial success required for reduction in bias increases when the favored type does not care enough about the future. The favored type’s response to the threat of losing favor is dominated by the unfavored type’s incentive to attain fairness as she cares more about the future. Sufficiently lower patience of the favored type is one possible case in which reducible bias can indeed reduce discrimination. For example, employers may be biased against candidates with lower experience and favor the senior and more experienced candidates. But, the future concern or patience also goes down with age (and hence experience) Falk et al. (2018). Therefore, when bias can be overcome, the inexperienced candidates with higher
patience are more incentivized to attain fairness than the experienced senior candidates with lower patience. This result is consistent with small evidence of discrimination based on experience (or even evidence of age-discrimination) as it is easier to overcome.

In summary, the effect of reducible bias is significantly larger for the favored type than the unfavored type under equal future concerns which leads to an increase in observed discrimination in period 1 and a discrimination trap overall. Reversible bias leads to similar increase in initial bids of the two types with equal future concerns generating no significant effect on the unfavored type’s initial chances of winning and hence allowing for actual bias reversal when it is reversible. However, when the favored type has a sufficiently lower future concern than the unfavored type, then the unfavored type reacts more strongly to the incentive to attain fairness in the future. In this case, discrimination trap breaks and fairness can be attained over time. These results provide support for the theory that reducible bias leads to increase in observed discrimination unless the favored type has lower enough future concern than the unfavored type or if the bias reverses and not just reduces.

5. Conclusion

This paper identifies and presents experimental evidence of one reason why discrimination can persist. I use a combination of theoretical and experimental methods to examine the effect of flexibility of discriminatory bias on the economic outcomes of agents. Organizations often rely on competition between agents for resource allocation such as elections, college admission tests, job interviews, litigation, sports. Thus, I model the problem in the framework of contest theory (Tullock (1967)). In doing so, I acknowledge the evidence on heterogeneity in the patience level of different populations and carefully study the effect of agents’ relative future concerns (Falk et al. (2018)). My research includes testing the theoretical predictions of my model using laboratory experiment. The aspect of flexibility of biases and its effect on the outcomes of competitive institutions is especially relevant in current times of social media when beliefs may be reinforced or changed in novel ways (Allcott & Gentzkow (2017)).
Table 5: Unequal future concerns - Effect of reducible bias on first period bids of the favored and unfavored types

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant Bias (1)</th>
<th>Reducible Bias (2)</th>
<th>Constant and Reducible Bias (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favored</td>
<td>8.12 (6.02)</td>
<td>-13.37** (3.5)</td>
<td>7.95* (3.69)</td>
</tr>
<tr>
<td>Reducible</td>
<td></td>
<td></td>
<td>26.3*** (2.75)</td>
</tr>
<tr>
<td>Favored × Reducible</td>
<td></td>
<td></td>
<td>-20.81*** (5.07)</td>
</tr>
<tr>
<td>Bid in Symmetric</td>
<td>0.45 (0.25)</td>
<td>0.29 (0.19)</td>
<td>0.37** (0.13)</td>
</tr>
<tr>
<td>Prediction of other’s bid</td>
<td>0.28 (0.06)</td>
<td>0.31 (0.20)</td>
<td>0.29* (0.11)</td>
</tr>
<tr>
<td>Intrinsic preference for winning</td>
<td>0.24*** (0.14)</td>
<td>0.09 (0.18)</td>
<td>0.20 (0.10)</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.8 (1.8)</td>
<td>23.08 (11.8)</td>
<td>-6.1 (4.9)</td>
</tr>
<tr>
<td>Observations</td>
<td>56</td>
<td>60</td>
<td>116</td>
</tr>
</tbody>
</table>

Notes: (1) * p-value ≤ 0.1; ** p-value ≤ 0.05; *** p-value ≤ 0.01
(2) Standard errors from OLS regressions are clustered at session level and reported in parentheses
(3) Favored = 1 if favored type, 0 otherwise
(4) Reducible = 1 if bias is reducible, 0 if bias is constant
(5) Intrinsic preference for winning is the bid in the contest for 0 prize

Policy intervention is necessary even when bias can be overcome as dynamic discrimination is unlikely to improve the unfavored groups’ opportunities and chances of succeeding while competing against favored groups. I analyze two policy interventions: tax rebate for the unfavored group and affirmative action. Even if only at the initial stage, tax rebate will make the return from winning higher for the unfavored group, thereby increasing their incentive and hence the effort to win. As the chances of winning improve for the unfavored group, they will be able to overcome discrimination. As the cost of this policy is not borne by the decision-makers such as employers, it does not distort their choice of contestant selection. Affirmative action or reverse discrimination (setting quotas/reservation for unfavored groups) to correct implicit biases might also mean lower
discrimination in dynamic settings than constant. However, possibility of backlash from affirmative action further increasing the initial bias or favor towards the favored type suggests that it will be less effective.

Apart from empirically comparing the effectiveness of tax rebates and affirmative action in mitigating discrimination when bias is dynamic, there are two relevant directions for future research emerging from this paper. Firstly, it would be interesting to study the case of an inter-group contest. An alternative interpretation of future concerns is ‘concern for group welfare’ such that different cohorts compete in different periods, and players care about the welfare of their group members. The analysis will change if multiple members of each group compete in every period as intra-group dynamics would also be important. While the winner gets the prize, the utility from change in favor in the next period, based on which group’s member wins in the current period, is like a club good and would generate free-riding incentives. It would also be interesting to see how subjects behave in a repeated group contest in the laboratory. Secondly, I plan to test the model with explicit types, specifically gender and racial discrimination. The predictions of this paper are generally applicable to any context of discrimination and have been tested with minimal group identity. However, there may be some gender or race specific characteristics affecting people’s response to the threat of losing favor or the opportunity of gaining it.

References


Appendix A

Proof of propositions

Using backward induction to solve for the subgame perfect Nash equilibrium of the 2 period contest, Consider the second and last period -

If player A won in the first period, then favor f remains the same in the second period. The second period contest success function is given by:

$$p_2^A(e_{a2}, e_{b2}), p_2^F(e_{a2}, e_{b2}) = 
\begin{cases} 
\frac{f e_{a2}}{f e_{a2} + e_{b2}}, \frac{e_{b2}}{f e_{a2} + e_{b2}} & \text{if } \max\{e_{a2}, e_{b2}\} > 0 \\
\frac{f}{f + 1}, \frac{1}{f + 1} & \text{if } e_{a2} = 0 = e_{b2}
\end{cases}$$
In the Nash Equilibrium of this contest:

**Equilibrium efforts:**

\[ e_{a2} = \frac{f}{(f+1)^2} V, \quad e_{b2} = \frac{f}{(f+1)^2} V \]  \hspace{1cm} (1)

**Equilibrium probabilities of winning:**

\[ p_2^a = \frac{f}{f+1}, \quad p_2^b = \frac{1}{f+1} \]

**Equilibrium expected utilities:**

\[ U_2^a = \frac{f^2}{(f+1)^2} V, \quad U_2^b = \frac{1}{(f+1)^2} V \]  \hspace{1cm} (2)

If player B won in the first period, then favor f goes down to \( af < f \). Contest success function is thus given by:

\[ p_2^a(e_{a2}, e_{b2}), p_2^b(e_{a2}, e_{b2}) = \begin{cases} 
\frac{\alpha f e_{a2}}{\alpha f e_{a2} + e_{b2}}, \frac{e_{b2}}{\alpha f e_{a2} + e_{b2}} \quad & \text{if } \max\{e_{a2}, e_{b2}\} > 0 \\
\frac{\alpha f}{\alpha f + 1}, \frac{1}{\alpha f + 1} \quad & \text{if } e_{a2} = 0, e_{b2} = 0 
\end{cases} \]

In the Nash Equilibrium of this contest:

**Equilibrium efforts:**

\[ e_{a2} = \frac{\alpha f}{(\alpha f + 1)^2} V, \quad e_{b2} = \frac{\alpha f}{(\alpha f + 1)^2} V \]  \hspace{1cm} (3)

**Equilibrium probabilities of winning:**

\[ p_2^a = \frac{\alpha f}{\alpha f + 1}, \quad p_2^b = \frac{1}{\alpha f + 1} \]

**Equilibrium expected utilities:**

\[ U_2^a = \frac{(\alpha f)^2}{(\alpha f + 1)^2} V, \quad U_2^b = \frac{1}{(\alpha f + 1)^2} V \]  \hspace{1cm} (4)

From (1) and (3), we know that \( e_{a2} = e_{b2} \) and is increasing in \( \alpha \) (as \( \alpha < 1 \)).

This proves proposition 1.2. \( \square \)
The first period contest success function is given by:

\[ \begin{align*}
    p_1^a(e_{a1}, e_{b1}), p_1^b(e_{a1}, e_{b1}) = & \begin{cases} 
    \frac{fe_{a1}}{fe_{a1} + e_{b1}}, \frac{e_{b1}}{fe_{a1} + e_{b1}} & \text{if } \max\{e_{a1}, e_{b1}\} > 0 \\
    \frac{f}{f + 1}, \frac{1}{f + 1} & \text{if } e_{a1} = 0 = e_{b1}
    \end{cases}
\end{align*} \]

Therefore, they maximize their expected utilities given by,

\[ \begin{align*}
    U_1^a(e_{a1}, e_{b1}) = & \frac{fe_{a1}}{fe_{a1} + e_{b1}}V_a - e_{a1} \\
    U_1^b(e_{a1}, e_{b1}) = & \frac{e_{b1}}{fe_{a1} + e_{b1}}V_b - e_{b1}
\end{align*} \]

where,

\[ \begin{align*}
    V_a = & \left[ 1 + \beta_a \left( \frac{f^2}{(f + 1)^2} - \frac{(\alpha f)^2}{(\alpha f + 1)^2} \right) \right]V \\
    \text{and} \\
    V_b = & \left[ 1 + \beta_b \left( \frac{1}{(\alpha f + 1)^2} - \frac{1}{(f + 1)^2} \right) \right]V
\end{align*} \]

This generates the following equilibrium outcome of period 1,

Equilibrium efforts:

\[ e_{a1} = \frac{fV_a^2V_b}{(fV_a + V_b)^2}, e_{b1} = \frac{fV_b^2V_a}{(fV_a + V_b)^2} \]  \hspace{1cm} (5)

Equilibrium probabilities of winning:

\[ \begin{align*}
    p_1^a = & \frac{fV_a}{fV_a + V_b}, p_2^b = \frac{V_b}{fV_a + V_b} 
\end{align*} \]  \hspace{1cm} (6)

Equilibrium expected utilities:

\[ \begin{align*}
    U_1^a = & \frac{fV_a}{fV_a + V_b} \left[ V - \frac{fV_aV_b}{fV_a + V_b} \right], U_1^b = \frac{V_b}{fV_a + V_b} \left[ V - \frac{fV_aV_b}{fV_a + V_b} \right] 
\end{align*} \]  \hspace{1cm} (7)
From (5),
\[ e_{a1} > e_{b1} \iff V_a > V_b \iff \frac{\beta_a}{\beta_b} > \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} = I < 1 \]
Note that, \( V_a \) and \( V_b \) which are decreasing in \( \alpha \).

Further,
\[
\frac{\partial (V_a)}{\partial \alpha} = \frac{2f(f + 1)^2(f \alpha + 1) - f^2 \beta_a(\alpha + 2f + 1) + f^2 \beta_a(1 - \alpha)}{(f + 1)^2(f \alpha + 1)^2 + f \beta_b(1 - \alpha)(f \alpha + f + 2)}
- \frac{(-f \beta_b(f \alpha + f + 2) + 2f(f + 1)^2(f \alpha + 1) + f^2 \beta_b(1 - \alpha))(f + 1)^2(f \alpha + 1)^2 + f^2 \beta_a(1 - \alpha)(\alpha + 2f + 1))}{((f + 1)^2(f \alpha + 1)^2 + f \beta_b(1 - \alpha)(f \alpha + f + 2))^2}
\]
\[
< 0 \quad \text{if} \quad \frac{\beta_a}{\beta_b} > I
\]
\[
\text{U-shaped} \quad \text{if} \quad \frac{\beta_a}{\beta_b} \in (1, I), \ \text{change of slope at} \ \alpha = \alpha^*(f)
\]
\[
> 0 \quad \text{if} \quad \frac{\beta_a}{\beta_b} < 1
\]

where \( \alpha^*(f) \) is increasing in \( f \) and is found by setting the above equation equal to 0.

Moreover, \( p_1^a \) is increasing in \( \frac{V_a}{V_b} \) and \( p_1^b \) is decreasing in \( \frac{V_a}{V_b} \). Thus, \( p_1^a \) is increasing, U-shaped and decreasing in endogeneity of \( f \) i.e., \((\alpha^{-1})\).

Similarly, \( U_1^a \) and \( U_2^b \) are decreasing in \((\alpha^{-1})\), decreasing in own \( \beta \) and increasing in other player’s \( \beta \).

This proves Proposition 1.1. \( \square \)

Appendix B

Micro-foundations of the favor \( f \)

**Contestants** - Consider two agents, A and B who exert efforts or spend resources to:

(i) persuade an evaluator that they are better such as the job market contest for getting hired or promoted, or,

(ii) to get selected in an institutional set-up such as the candidates contesting an election in a political system.
Let A’s effort to generate persuasion resources be denoted by \( e_a \) and B’s efforts be denoted by \( e_b \). The contesting agent \( i \)’s type \( \gamma_i \) is the combination of their identity (such as gender) and ability. Ability is captured by cost functions for A and B, and given by \( c_A(.) \) and \( c_B(.) \). I assume that \( c_i'(.) > 0, c_i''(.) \leq 0, \) and \( c_i(0) = 0 \) where \( i \in A,B \). The utility from winning the contest is given by \( V \) (such as the wage from getting hired). For simplicity, let us assume that both A and B are risk neutral\(^\text{10} \).

Candidates of identity A and B compete twice in economic contexts such as the job market and political elections which they value equally at \( V \). The future concern relevant for repeated contest between A and B is denoted by \( \beta_a \) and \( \beta_b \) respectively. Note that, it could be the same individuals from identity groups A and B competing against each other repeatedly over time, or it could be different individuals from groups A and B. The interpretation of the future concern is the level of patience or own future welfare in the case when the same individuals A and B compete repeatedly. However, when different individuals from identity groups A and B compete over time for the same job, the future concerns \( \beta_a \) and \( \beta_b \) are interpreted as intra-group welfare concern (Akerlof & Kranton (2000, 2010)).

**Evaluator or Institution** - Evaluators (such as hiring managers) or institutions (such as elections with statistically large number of voters) select the winner of the contest.

**Initial Bias** - The evaluator and the institutions may be biased in favor of one candidate or the other such that the favored candidate is more likely to win when both exert equal effort (or expend equal resources).

**Evaluator** – The evaluator wants to select the better matched candidate. A candidate \( i \) is better matched if they have a higher ability net of the disutility that the evaluator faces from their selection. For example, when a manager is choosing between a male candidate and a female candidate, the manager cares about two things - ability of the candidate and the future experience of working with the hired candidate. Therefore, the manager makes the selection decision based

\(^\text{10}\)The effect of non-neutral risk preferences is ambiguous (Sheremeta (2011)).
on expected abilities of the male and the female candidate and the perceived disutility from working with a man and a woman. Evaluators in different economic contexts may weigh ability and preference differently. For example, the owner of a firm may care more about the productivity and ability of the hired candidate while the manager may care more about the perceived disutility from working with men and women while making the selection decision.

Recall that the two types of candidates are denoted as A and B. An evaluator \( j \) gets the utility 1 if the *better matched* candidate is selected and 0 otherwise i.e., \( U_j = I(\text{better candidate is chosen}) \) where \( I \) is the indicator variable. Let \( \pi \) be the evaluator’s prior belief that A is the better matched candidate.

The candidates A and B exert effort or spend resources given their ability or cost functions \( c_a \) and \( c_b \).

A utility maximizing evaluator forms a posterior belief \( \pi_1 \) of A being the better candidate and selects A with probability \( \pi_1 \) and B with probability \( 1 - \pi_1 \). This selection rule generates the following functional form of the evaluator’s posterior belief or A and B’s probability of winning when the prior belief is \( \pi \), A exerts effort \( e_a \), and B exerts effort \( e_b \).

\[
P_a = \pi_1 = \frac{\pi e_a}{\pi e_a + (1 - \pi)e_b} \quad \text{and} \quad P_b = 1 - \pi_1 = \frac{(1 - \pi)e_b}{\pi e_a + (1 - \pi)e_b}
\]

when \( e_a + e_b > 0 \)

And,

\[
P_a = \pi_1 = \pi \quad \text{and} \quad P_b = 1 - \pi_1 = (1 - \pi)
\]

when \( e_a = e_b = 0 \)

This is the standard Tullock lottery contest function Tullock (1967); Konrad (2009) following the derivation described in Skaperdas & Vaidya (2012) such that candidates exert effort to persuade the evaluator that they are the better candidate.
Institution – Institutions such as religion and political system or official language of communication and education versus mother tongue also play a role in candidate selection. Biases may be structural and beyond the control of any specific individual or set of individuals (McCrudden (1982); Moro & Norman (2004)). For example, a woman may be less likely to find paid-work in a religious context which strictly defines gender-roles and deems women as household workers. Female priests are not easily hired for conducting rituals such as wedding and funeral. Similarly, in a democratic election for political leader, popularity of the candidates is more important and may be unrelated to their ability and understanding of policy. Therefore, personality politics emerges and makes it difficult for shy individuals to thrive in a majority voting system irrespective of qualifications and economic performance. Another example is the official language which makes it difficult for non-native speakers to succeed against native speakers even if the evaluators are not biased (Lang (1986)). Similarly, due to organizational and structural differences in the education system, it is difficult for foreign educated students to get their first job as an outsider.

DEFINITION 1: Initial Bias - An evaluator or institution $j$ is biased in favor of candidate $A$ if $\pi > 0.5$ i.e., for equal persuasion effort or resources, $A$ has a higher chance of winning than $B$. In other words, $A$ is the favored candidate if $\pi > 0.5$. WLOG, let $A$ be the favored candidate. Define $f$ as the intensity of initial bias in favor of $A$ such that $f = \frac{\pi}{1 - \pi} > 1$. Higher the $f$, higher is the initial bias in favor of $A$.

Optimal Effort - Candidates choose the effort that maximizes their expected payoff with respect to the bias they face and their ability. Thus, optimal $e_a, e_b$ maximize $U_a, U_b$ given $f, c_a, c_b$ and $V$, i.e., $e_a, e_b$ maximize:

$$U_a = \frac{f e_a}{f e_a + e_b} V - c_a(e_a)$$

\[\text{Or, higher the difference between } f \text{ and } 1, \text{ higher is the initial bias in favor of } A.\]
and

\[ U_b = \frac{e_b}{f e_a + e_b} V - c_b(e_b) \]

In the case of no bias, i.e., when \( f = 1 \), then,

\[ \frac{e_a}{e_b} = \frac{\partial c_b}{\partial c_a} = \frac{MC_b(e_b)}{MC_a(e_a)} \]

where \( MC_i \) is the marginal cost of effort of candidate \( i \). This implies that under no bias, candidates’ probability of winning is determined solely by their ability i.e.,

\[(P_a|f = 1) = \frac{MC_b(e_b)}{MC_a(e_a) + MC_b(e_b)}\]

and,

\[(P_b|f = 1) = \frac{MC_a(e_a)}{MC_a(e_a) + MC_b(e_b)}\]

**Discrimination** - Discrimination is the disparate opportunities (chances of selection) of candidates based on their identity (A or B), rather than abilities. When \( f > 1 \), then A and B will have different probabilities of winning even if they have equal ability. There is no realized discrimination if the candidates are incentivized such that their relative ability exactly determines their probability of winning, i.e., I define discrimination as follows:

**DEFINITION 2** - Discrimination: Candidate \( i \) is discriminated against by an evaluator or institution \( j \) if \( i \)'s realized probability of winning smaller than their no-discrimination probability of winning i.e., \( (P_i|f = 1) \). Let

\[ D_i = (P_i|f = 1) - P_i \]

where \( i \in \{A, B\} \) denote the difference between candidate \( i \)'s economic opportunities or chances of winning without and with bias.

**Repeated Contest** The same candidates A and B may compete again such as in cases of promotion and appraisal within the firm and political candidates in the next round of elections. They may be re-evaluated the second time by the same evaluator as the first time (manager deciding who gets the highest appraisal every year), or, by a different evaluator (employers for freelancers). I also
allow for different individuals of type (or identity) A and B to compete after the first contest. For example, after a man and a woman compete for a position or a job, another set of candidates apply for the same job.

**Subsequent Bias** - Bias in the subsequent contest will be affected by the outcome and information feedback from the initial contest and the model of inference of the evaluator. The information feedback and relevant model of inference from the first contest depends whether a new evaluator or the same evaluator does the selection in the subsequent contest. When a new evaluator \( j' \) evaluates A and B in the second contest, they know the winner of the first contest. Following are some of the models of inference which an evaluator may follow that determines their type defined as \( \pi_2 \) (winner of initial contest) or correspondingly \( f_2 \) (winner of the initial contest) i.e., the favor towards A in the subsequent contest as a function of the winner of the initial contest.

1. Bayesian updating: Under the Bayesian model of inference evaluators apply Bayes’ rule to form updated beliefs based on the information about the winner of the initial contest (Becker et al. (1963)). Prior belief of evaluator \( j' \) that A is better: \( \pi > 0.5 \) is updated based on the winner of the initial contest. Higher the ability or lower the cost of effort, higher are a candidate’s chances of winning or selection. Thus, winning is a positive signal of ability and the belief \( \pi \) or probability that A is better increases when A wins while decreases when B wins: \( \pi_2 > \pi \) if A wins in initial contest and \( \pi_2 < \pi \) if B wins in the initial contest. Due to the bias in favor of A in the initial period, winning is a stronger signal of ability for B than A. Bayesian updating implies higher reduction in favor \( f \) when B wins and smaller increase in the favor when A wins. If \( j' \) internalizes that A was more likely to win due to the bias and yet won, then the bias may even reverse in favor of B when B wins. This is because the evaluator is correcting for initial bias as B must be high ability to win despite the bias (i.e., \( \pi_2 < 0.5 \) or \( f < 1 \) if B had won).

2. Conformism updating: Under the conformation model of inference, evaluators are more likely to draw inference from an information or evidence which conforms to their belief and expectation, and ignore the counter-expectation information (Nickerson (1998)). The evaluator \( j' \) believes that A is more likely to be the better candidate and enjoys a favor in the initial contest due
to which A was more likely to be the winner of the initial contest. In other words, \( j' \) expects A to have won in the initial contest. If A indeed had won, then the expectation of \( j' \) gets confirmed and hence their belief that A is the better candidate becomes even stronger. However, the increase in \( \pi \) cannot be too large as \( \pi \) is already higher than 0.5 and bounded by 1. Thus, if A had won, there is a small increase in the belief \( \pi \) of a conformist evaluator: \( \pi_2 > \pi \) if A had won. But if B had won, then \( j' \) ignores that it is an evidence which signals high ability of B and instead attributes B’s win to chance or luck (Gill & Prowse (2014); Lightbody et al. (1996)). Thus, if B had won, there is no change in the belief \( \pi \) of a conformist evaluator: \( \pi_2 = \pi \) if B had won.

2. Counter-intuitive evidence based updating: Under this model of inference, evaluators update their belief when they observe unexpected events, while their beliefs are reinforced upon observing expected events. As the evaluator expects A to win, there is no change in the belief that A is better if A had won i.e., \( \pi_2 = \pi \) and \( f_2 = f \), if A had won. But, winning of B is considered relevant evidence as it was unexpected. Thus, the favor towards A is lower or even reversed in the subsequent contest if B had won in the initial contest: \( \pi_2 < \pi \) and \( f_2 < f \) (and maybe even less than 1).

Bias may be dynamic also when the same evaluator gets the exposure of working with the unfavored type B as a result of B winning. If the evaluator has a high prior that A is the better candidate and is biased against B, then exposure to B can help mitigate this bias as found in several empirical studies on the effect of exposure to members of a disadvantaged group (Boisjoly et al. (2006); Beaman et al. (2009)). When B had won in the initial contest, then the evaluator gets an exposure to the type which is biased against and may realize a lower disutility from B’s selection than expected. Thus, favor toward A is lower in subsequent contest if B had won in initial contest.\(^\text{12}\)

No new exposure or experience is attained, however, when A had won in the initial contest and hence the level of favor remains the same.

In the case of institutional bias, the subsequent period bias is affected by whether and how

\(^\text{12}\)One might argue that the evaluator may in fact realize an even higher disutility from working with B and hence exposure may increase the favor towards A. But, that is counter-evidence as studies show increased tolerance and empathy toward each other due to increased exposure, on average.
much the contest structure gets affected by the outcome i.e., the winner of the initial contest. I posit that in an institution which is biased in favor of A, winning of B can potentially change the structure and reduce the bias. If the institution is biased but dynamic/progressive, then B’s winning can reduce the bias. For example, a society influenced by media which covers and highlights the success of a minority student, is more likely to open up opportunities for other minority students either due to increased participation (Carvalho & Pradelski (2018)) or due to accommodating rules and institutional culture created by the successful minority students. But an orthodox context such as religious institution is less likely to evolve or change the structural biases.

The subsequent or second period bias depends on how sensitive the bias is to the winner of the initial contest. The empirical literature on the dynamics of bias has found evidence of a reduction in bias after a positive signal or exposure to the unfavored type while no change in bias after the success of the favored type. Next, I examine the effect of such dynamics of bias on the behavior or effort choice of the candidates A and B and the equilibrium level of discrimination against the initially unfavored type B. In summary, consider a repeated contest between A and B such that the bias in the second period depends on the outcome of the first period. The level of bias goes down or reverses if and only if the unfavored type B wins in period 1. It remains the same if the favored type A wins in period 1.

\[
\begin{align*}
    f_2 = \begin{cases} 
    \alpha f_1 = \alpha f & \text{if B wins in period 1 and bias is reducible or reversible} \\
    f_2 = f & \text{if A wins in period 1}
    \end{cases}
\end{align*}
\]

where \(\alpha^{-1}\) represents the degree of reducibility of bias. Higher \(\alpha\) means that the bias cannot be easily overcome or surmounted. Bias is reversible when \(\alpha f < 1\) and reducible when \(\alpha f \geq 1\)

This derives the contest success function of period one and two analyzed in the paper.