

# Persistence of Discrimination Theory and Experimental Evidence

Nishtha Sharma\*

(JOB MARKET PAPER)

[Click here for the latest version](#)

## Abstract

Why does discrimination persist when there is evidence that discriminatory bias can be overcome? I answer this question with a novel application of contest theory and provide experimental evidence. I show that a favored candidate's threat of losing privilege is more salient than an unfavored candidate's opportunity of overcoming bias in a competition. My findings reveal that reducible and reversible bias cause the favored candidate to increase their effort significantly more than the unfavored candidate unless the unfavored candidate cares sufficiently more about the future. Discrimination persists even when bias can be overcome due to the favored candidate's resistance to change. (*JEL* C73, C92, D63, D74, D90, J15, J16, J71)

---

\*University of California, Irvine (email: nishths@uci.edu). Acknowledgements-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. This project was approved by the Institutional Review Board at the University of California, Irvine. I refer to the economic agents discussed in this paper using gender-neutral pronouns: they/them/their.

*“It is harder to crack prejudice than an atom.”* Albert Einstein

While economic discrimination remains a topic of research interest, popular media and survey-based scholarly writing suggest that it will be less of a problem as people are becoming more willing to change their biases<sup>1</sup>. Recent causal studies have also found that people from unfavored communities such as African Americans, women, immigrants, minorities and scheduled caste can overcome the discrimination that they face while competing for economic gains such as jobs and political positions<sup>2</sup>. They argue that evaluators such as employers and voters update their attitude and belief about the unfavored candidate who initially succeeds and gets their foot in the door. Evaluators judge the originally unfavored with lower or even reversed bias during repeated competition if they have won before. Nevertheless, identity-based discrimination persists not only in hiring but also political voting, access to loans, housing amenities, and healthcare, among other economic resources for which people compete<sup>3</sup>. The persistent evidence on discrimination is puzzling as we expect the opportunity of overcoming bias to encourage the unfavored candidate to achieve the requisite initial success<sup>4</sup>.

This paper contributes to our understanding of the dynamics of discrimination by explicitly studying the behavior of the candidates competing with changeable bias. When people compete in the presence of discrimination, the unfavored candidate is encouraged by the opportunity of overcoming bias, while such a possibility threatens the favored candidate. I theoretically analyze and compare their incentives and behavior when bias can and cannot be overcome. I use the framework of dynamic contest theory to examine repeated competition between the favored and the unfavored candidates (Konrad, 2009). I model dynamic discrimination as a two-player, two-period contest such that, all else equal, the favored candidate has a higher chance of winning the first-period contest. However, the bias declines for the second period if the unfavored candidate wins in the first period.

---

<sup>1</sup>See (Bobo et al., 2012; Tesler & Sears, 2010; Reardon & Fahle, 2017) for a discussion of people’s willingness to change their discriminatory behavior.

<sup>2</sup>There are both theoretical models and empirical evidence of bias decline and reversal upon initial success of unfavored people (Bohren et al., 2019; Mengel et al., 2019; Beaman et al., 2009; Fryer Jr, 2007; Boisjoly et al., 2006; Groot & Van Den Brink, 1996; Lewis, 1986)

<sup>3</sup>Experimental evidence for discrimination has been documented from as early as 1974 to 2021 (Christensen & Timmins, 2021; Quillian et al., 2020; Button et al., 2020; Neumark et al., 2019; Neumark, 2018; Baert, 2018; Agan & Starr, 2018; Bertrand & Duflo, 2017; Quillian et al., 2017; Bertrand & Mullainathan, 2004; Jolson, 1974).

<sup>4</sup>For example, based on their finding that accomplished female academics in STEM fields are favored over male academics Williams & Ceci (2015) state that “these results suggest it is a propitious time for women launching careers in academic science”.

My theoretical model demonstrates that when competition between candidates is used as a tool to allocate economic resources and bias can be overcome, then there is more to lose from losing for the favored candidate than there is to gain from winning for the unfavored candidate. The favored candidate's incentive to maintain the bias is larger, preventing the mitigation of discrimination.

Based on global evidence that people from different identity groups systematically differ in patience or their concern for the future (Falk et al., 2018; H. Chen et al., 2005; Schwartz, 1978), I allow for unequal levels of future concern for the favored and unfavored candidates and derive a threshold level of relative future concern. The opportunity to overcome bias will increase the unfavored candidate's initial probability of winning only if they care *sufficiently more* about the future than the favored candidate. Next, I analyze the possibility of bias reversal due to the unfavored candidate's initial success (Bohren et al., 2019; Fryer Jr, 2007). I show that even reversible bias can improve the unfavored candidate's current chances of winning only when they care sufficiently more about the future than the favored candidate, which is rare. The main theoretical prediction is that both candidates exert higher effort under reducible and reversible bias than constant bias; and the increase in effort is higher for the favored candidate than the unfavored candidate unless the favored candidate cares sufficiently less about the future. The stronger incentive to maintain than overcome or gain privilege explains the slow change in discrimination despite the growing opportunity of overcoming and even reversing the bias.

In addition to developing a novel dynamic contest game, I conducted a laboratory experiment to test the theoretical predictions empirically. I employed a between-subjects design such that each participant was equally likely to participate in any one of the following five treatments. (i) Bias cannot be overcome, and both candidates care about the future equally; (ii) Bias will be mitigated if the unfavored candidate wins initially and both candidates care about the future equally; (iii) Bias cannot be overcome, and the favored candidate cares about the future sufficiently lesser than the unfavored candidate; (iv) Bias will be mitigated if the unfavored candidate wins initially and the favored candidate cares about the future sufficiently lesser than the unfavored candidate; (v) Bias will be reversed if the unfavored candidate wins initially. Although an ideal test would randomly assign favorable bias to otherwise equivalent candidates and precisely measure and compare their behavior across the above treatments while they compete for an economic resource such as a job. However, such a test is not only infeasible in the field, but it is also unethical to subject people to discriminatory bias and manipulate their economic opportunities for a research

project. A laboratory experiment has the relative advantage of facilitating a simulation of the above treatments.

Half of the participants in every treatment were randomly assigned to be the favored candidate. Every candidate was equally likely to be the favored candidate thus maintaining a priori equity. Favored and unfavored candidates were anonymously and randomly matched into pairs to compete twice for a fixed prize in each of the two periods. Participants chose their effort bids out of a fixed endowment in both periods, the remaining endowment was converted into dollar earnings (Sheremeta, 2010). The higher the bid, the higher was their chance of winning the contest prize, but the lesser of their endowment remained for dollar conversion<sup>5</sup>. Participants' payoff was the sum of their period-one and future concern times period-two payoff. The candidates were informed that their bid choice also affected the second period's level of bias through the winner's identity (favored or unfavored) when bias could be overcome. Finally, I conducted several puzzles to measure participants' competitiveness, attention, innate taste for winning, risk preference, myopia, and cognitive abilities.

I found that both candidates choose higher effort bids when bias can be overcome. However, the increase is significantly higher for the favored type (4.5 standard deviations) than the unfavored type (1.5 standard deviations). Previous studies on the dynamics of discrimination would predict no difference in the unfavored candidate's probability of winning the period-one competition when bias can and cannot be overcome. In contrast, as my theory predicted, I found that the unfavored candidates had a significantly lower frequency of winning when bias could be overcome. The results did flip for the case when the favored candidate had sufficiently lower concern for the future. The unfavored candidate increased their bid (18 standard deviations) more than the favored candidate (8.6 standard deviations) due to sufficiently higher stakes in the future resulting in a significant increase in the unfavored candidates' winning frequency. While there is evidence of identity-based heterogeneity in future concern, the required relative future concern threshold is too large making it unlikely to mitigate discrimination despite being a plausible channel (Falk et al., 2018; H. Chen et al., 2005; Schwartz, 1978).

An effective policy tool given dynamic discrimination would either increase the returns from winning for the unfavored group (such as lower income tax for people identified as unfavored) or not rely on competition between favored and unfavored groups (reservation or

---

<sup>5</sup>Participants were told that if they are type A, then their bid choice will be multiplied by three and that many 'A' labeled balls will be put in a virtual bid bag. For the B type, participants knew that the same number of 'B' type balls would be put in the bid bag as their choice of effort bid.

quotas). However, such interventions may be infeasible or exert huge public cost and even have potential backlash effects (Belmonte & Di Lillo, 2021; Moro & Norman, 2003; Maio & Esses, 1998). Therefore, I argue and conclude that a long-run solution would be to invest in early childhood exposure to diversity and awareness programs that directly lower the initial bias unconditionally (Boisjoly et al., 2006; Pope et al., 2018).

***Related Literature*** – The economic literature on discrimination focuses on two main sources of discrimination: taste-based and belief-based, which interact with cultures and institutions. Taste-based models of discrimination consider bias to be like an exogenous and unchanging preference parameter leading to the persistence of discrimination (G. S. Becker, 1957; Cain, 1986; Roland, 2004). This paper considers the evidence on changeable bias and shows that even when bias can decline or reverse, equality in opportunities is hindered by the incentives emerging from the use of competition between people as an institutional tool to allocate economic gains such as job positions. Models of belief-based discrimination consider the possibility of bias decline or reversal when evaluators update their initially incorrect beliefs (Fryer Jr, 2007; Bohren et al., 2019). However, their work focuses on the source and mechanism of change of the bias. This paper shows that regardless of the evaluator’s model of inference<sup>6</sup>, candidates’ response to dynamic discrimination will have non-trivial consequences for the equality of outcomes.

Belief-based or statistical models of discrimination mainly argue that discrimination persists due to coordination failure among members of the unfavored group stuck in bad equilibria of low human capital investment, which self-fulfills the lower belief in their ability (Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977; S. J. Lundberg & Startz, 1983; S. Lundberg & Startz, 1998; Coate & Loury, 1993; Mailath et al., 2000). However, the human capital investment gap has closed at a fast pace without commensurate closure in the employment rate (for example, the racial gap in education in the U.S. is small and insignificant while the employment gap remains high (Neumark, 2018)). My model is a two-player contest and does not rely on intra-group coordination to explain discrimination. It reconciles these opposite trends in education and employment equality by showing that favored and unfavored individuals with equal education may exert unequal effort in competing for economic gains as a consequence of asymmetric incentives emerging from dynamic bias.

This paper advances the economic literature on discrimination from the finding that bias is dynamic to understanding the response of people competing in dynamically biased

---

<sup>6</sup>How much the bias declines or reverses upon the unfavored candidate’s success depends on whether the evaluator’s updating rule is Bayesian or conformism or counter-prediction as discussed in detail in the appendix.

settings (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. 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 contests<sup>7</sup>.

The rest of the paper is structured as follows. Section I 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 II describes the experimental design to test the theoretical predictions. Section III presents empirical findings based on the laboratory experiment. Section IV concludes with potential extensions and discusses the policy implications of the theory and empirical findings.

## I 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<sup>8</sup>. 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<sup>9</sup>

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  $f$ . The higher

---

<sup>7</sup>While 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.

<sup>8</sup>Assuming convex cost does not change the qualitative results.

<sup>9</sup>The effect of non-neutral risk preferences is ambiguous (Sheremeta, 2011).

the  $f$ , the higher is the initial bias in favor of A or, the higher the difference between  $f$  and 1, the 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<sup>10</sup>. 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.

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

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,

---

<sup>10</sup>A Contest Success Function (CSF) provides each player's probability of winning as a function of all players' efforts (Skaperdas, 1996).

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, the favorable bias for A as well as the 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 A 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.

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,

[label=)]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  $\gamma = \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} < 1$ . 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  $f_2$

The 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.

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:

□

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, 
$$\gamma = \frac{(2 + \alpha f + f)}{f + \alpha f + 2f^2} < 1.$$

The next section presents the experimental design to test the theoretical predictions of the model.

## II 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 the unfavored candidate’s 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

	$(f_1, f_2 \text{ if B wins in period 1, } \beta_a, \beta_b)$		
	<b>Constant Bias</b>	<b>Reducible bias</b>	<b>Reversible bias</b>
$\beta_a = \beta_b$	(3,3,5,5)	(3,1,5,5)	(3,1/2,5,5)
$\beta_a = 0.02\beta_b$	(3,3,0.1,5)	(3,1,0.1,5)	

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

Bids - $(e_{1a}, e_{1b})$			
	<b>Constant Bias</b>	<b>Reducible bias</b>	<b>Reversible bias</b>
$\beta_a = \beta_b$	18.75, 18.75	41,31	38,33
$\beta_a = 0.02\beta_b$	18.75, 18.75	24.4, 45.8	-

Table 2 describes the theoretical predictions of the bids in period 1 of the favored and the unfavored types  $(e_{1a}, e_{1b})$ <sup>11</sup>. 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

<sup>11</sup>The theoretical predictions and subsequent findings pertaining to period 2 are in appendix.

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 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. The subjects had the choice of typing the bids or moving the slider (both were synchronized). Subjects could see what they kept for each bid. They also saw their probabilities of winning and losing, and 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 [Payoff in period 1 + (5 \* 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 subjects as to 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.

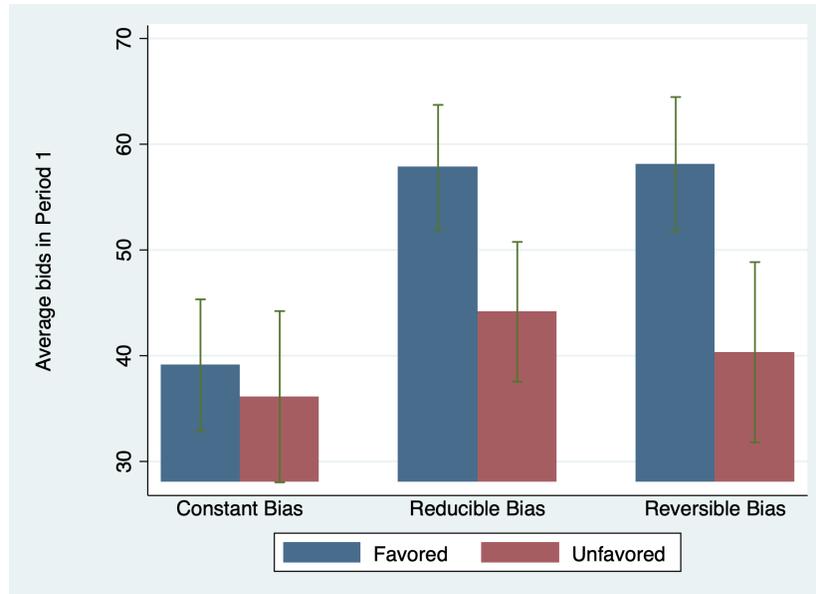
### **III Experimental Results**

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.

#### **III.a 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



Note: Error bars represent 90% confidence intervals.

Figure 1: Comparison of the favored and unfavored types' period 1 bids under equal future concerns

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), the interaction of type and reducible bias and the average bid in symmetric contest<sup>12</sup> (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.

<sup>12</sup>I 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.

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.

Table 3: Equal future concerns - Effect of reducible bias on first period bids of the favored and unfavored types

Variable	Constant Bias (1)	Reducible Bias (2)	Constant and Reducible Bias (3)
Favored	1.58 (4.95)	11.65*** (1.92)	1.6 (4.9)
Reducible			2.7 (6.7)
Favored $\times$ Reducible			10.22* (5.07)
Bid in Symmetric	0.57** (0.17)	0.52*** (0.102)	0.56*** (0.11)
Prediction of other's bid	0.19 (0.09)	0.38*** (0.07)	0.30*** (0.07)
Intrinsic preference for winning	0.49*** (0.10)	0.07 (0.21)	0.30*** (0.14)
Constant	-9.01* (4.14)	-2.45 (5.14)	-7.8** (3.07)
Observations	80	70	150

Notes: (1) \* p-value  $\leq 0.1$ ; \*\* p-value  $\leq 0.05$ ; \*\*\* p-value  $\leq 0.01$

(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

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 only. The possibility of gaining favorable bias does not incentivize people more than the threat of losing it.

Table 4: Equal future concerns - Effect of reversible bias on first period bids of the favored and unfavored types

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

Notes: (1) \* p-value  $\leq 0.1$ ; \*\* p-value  $\leq 0.05$ ; \*\*\* p-value  $\leq 0.01$

(2) Standard errors from OLS regressions are clustered at session level and reported in parentheses

(3) Favored = 1 if favored type, 0 otherwise

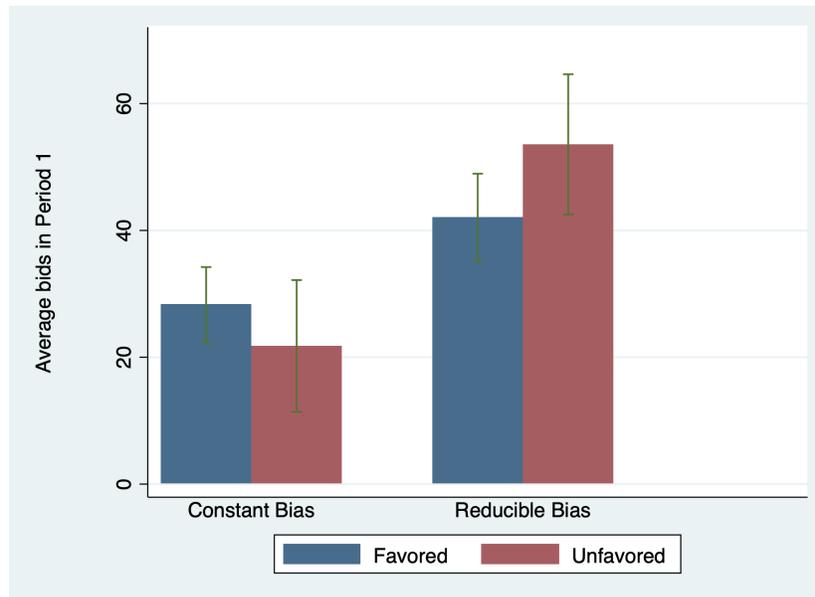
(4) Reversible = 1 if bias is reversible, 0 if bias is constant

(5) Intrinsic preference for winning is the bid in the contest for 0 prize

### III.b 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 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.



Note: Error bars represent 90% confidence intervals.

Figure 2: Comparison of the favored and unfavored types' period 1 bids under unequal future concerns

Bringing the constant and reducible bias within the same model I again regress the average 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.

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.

Table 5: Unequal future concerns - Effect of reducible bias on first period bids of the favored and unfavored types

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

Notes: (1) \* p-value  $\leq 0.1$ ; \*\* p-value  $\leq 0.05$ ; \*\*\* p-value  $\leq 0.01$

(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

## IV Conclusion

This paper identifies and presents experimental evidence for one reason why discrimination can persist. I use a combination of theoretical and experimental methods to examine the effect of changeable 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). I show that the stronger incentive to maintain than overcome or gain privilege explains the slow change in discrimination despite the growing opportunity of overcoming and even reversing discriminatory bias.

The persistence of discrimination despite the opportunity for overcoming bias warrants policy intervention. An effective policy tool given dynamic discrimination would either increase the returns from winning for the unfavored group (such as lower income tax for people identified as unfavored) or not rely on competition between favored and unfavored groups (reservation or quotas). However, such interventions may be infeasible or exert a large public cost and even have potential backlash effects (Belmonte & Di Lillo, 2021; Moro & Norman, 2003; Maio & Esses, 1998). Therefore, I conclude that a long-run solution would be to invest in early childhood exposure to diversity and awareness programs that directly target the initial bias and are found to work (Dhar et al., 2018; Pope et al., 2018; Boisjoly et al., 2006).

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 even asymmetric competition more generally. I have tested the theoretical predictions only 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

- Agan, A., & Starr, S. (2018). Ban the box, criminal records, and racial discrimination: A field experiment. *The Quarterly Journal of Economics*, *133*(1), 191–235.
- Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, *30*(2), 175–187.
- Akerlof, G. A., & Kranton, R. (2010). Identity economics. *The Economists' Voice*, *7*(2).
- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.
- Arrow, K. J. (1973). *Information and economic behavior* (Tech. Rep.). Harvard University Cambridge MA.
- Baert, S. (2018). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. *Audit studies: Behind the scenes with theory, method, and nuance*, 63–77.
- Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, *124*(4), 1497–1540.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral science*, *8*(1), 41–55.
- Becker, G. S. (1957). *The theory of discrimination*. Chicago: University of Chicago Press.
- Belmonte, A., & Di Lillo, A. (2021). Backlash against affirmative action: Evidence from the south tyrolean package. *European Economic Review*, *137*, 103802.
- Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. In *Handbook of economic field experiments* (Vol. 1, pp. 309–393). Elsevier.

- Bertrand, M., & 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.
- Bobo, L. D., Charles, C. Z., Krysan, M., & Simmons, A. D. (2012). 3. the real record on racial attitudes. In *Social trends in american life* (pp. 38–83). Princeton University Press.
- Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, *109*(10), 3395–3436.
- Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., & Eccles, J. (2006). Empathy or antipathy? the impact of diversity. *American Economic Review*, *96*(5), 1890–1905.
- Button, P., Dils, E., Harrell, B., Fumarco, L., & Schwegman, D. (2020). *Gender identity, race, and ethnicity discrimination in access to mental health care: Preliminary evidence from a multi-wave audit field experiment* (Tech. Rep.). National Bureau of Economic Research.
- Cain, G. G. (1986). The economic analysis of labor market discrimination: A survey. *Handbook of labor economics*, *1*, 693–785.
- Carvalho, J.-P., & Pradelski, B. (2018). Identity and underrepresentation. *Available at SSRN*.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.
- Chen, H., Ng, S., & Rao, A. R. (2005). Cultural differences in consumer impatience. *Journal of Marketing Research*, *42*(3), 291–301.
- Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, *103*(2), 690–731.
- Christensen, P., & Timmins, C. (2021). *The damages and distortions from discrimination in the rental housing market* (Tech. Rep.). National Bureau of Economic Research.
- Coate, S., & Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 1220–1240.

- Dechenaux, E., Kovenock, D., & Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18(4), 609–669.
- Dhar, D., Jain, T., & Jayachandran, S. (2018). *Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india* (Tech. Rep.). National Bureau of Economic Research.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645–1692.
- Fryer Jr, R. G. (2007). Belief flipping in a dynamic model of statistical discrimination. *Journal of Public Economics*, 91(5-6), 1151–1166.
- Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2), 351–376.
- Groot, W., & Van Den Brink, H. M. (1996). Glass ceilings or dead ends: Job promotion of men and women compared. *Economics Letters*, 53(2), 221–226.
- Hey, J. D., & Lee, J. (2005). Do subjects separate (or are they sophisticated)? *Experimental Economics*, 8(3), 233–265.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5), 1644–1655.
- Hübner, M., & Vannoorenberghe, G. (2015). Patience and long-run growth. *Economics Letters*, 137, 163–167.
- Jolson, M. A. (1974). Employment barriers in marketing: How do employers react to the race or religion of an applicant for a marketing position? *Journal of Marketing*, 38(2), 67–69.
- Konrad, K. A. (2009). Strategy and dynamics in contests. *OUP Catalogue*.
- Lang, K. (1986). A language theory of discrimination. *The Quarterly Journal of Economics*, 101(2), 363–382.
- Lewis, G. B. (1986). Gender and promotions: Promotion chances of white men and women in federal white-collar employment. *Journal of Human Resources*, 406–419.

- Lightbody, P., Siann, G., Stocks, R., & Walsh, D. (1996). Motivation and attribution at secondary school: The role of gender. *Educational Studies*, 22(1), 13–25.
- Lundberg, S., & Startz, R. (1998). On the persistence of racial inequality. *Journal of Labor Economics*, 16(2), 292–323.
- Lundberg, S. J., & Startz, R. (1983). Private discrimination and social intervention in competitive labor market. *The American Economic Review*, 73(3), 340–347.
- Mailath, G. J., Samuelson, L., & Shaked, A. (2000). Endogenous inequality in integrated labor markets with two-sided search. *American Economic Review*, 90(1), 46–72.
- Maio, G. R., & Esses, V. M. (1998). The social consequences of affirmative action: Deleterious effects on perceptions of groups. *Personality and Social Psychology Bulletin*, 24(1), 65–74.
- McCrudden, C. (1982). Institutional discrimination. *Oxford J. Legal Stud.*, 2, 303.
- Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535–566.
- Moro, A., & Norman, P. (2003). Affirmative action in a competitive economy. *Journal of Public Economics*, 87(3-4), 567–594.
- Moro, A., & Norman, P. (2004). A general equilibrium model of statistical discrimination. *Journal of economic theory*, 114(1), 1–30.
- Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3), 799–866.
- Neumark, D., Burn, I., & Button, P. (2019). Is it harder for older workers to find jobs? new and improved evidence from a field experiment. *Journal of Political Economy*, 127(2), 922–970.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175–220.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review*, 62(4), 659–661.

- Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, *64*(11), 4988–4995.
- Quillian, L., Lee, J. J., & Honoré, B. (2020). Racial discrimination in the us housing and mortgage lending markets: a quantitative review of trends, 1976–2016. *Race and Social Problems*, *12*(1), 13–28.
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, *114*(41), 10870–10875.
- Reardon, S., & Fahle, E. (2017). Education. state of the union: The poverty and inequality report. *Pathways Magazine*.
- Roland, G. (2004). Understanding institutional change: Fast-moving and slow-moving institutions. *Studies in comparative international development*, *38*(4), 109–131.
- Schwartz, B. (1978). Time, patience, and black people: a study of temporal access to medical care. *Sociological Focus*, *11*(1), 11–20.
- Sheremeta, R. M. (2010). Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior*, *68*(2), 731–747.
- Sheremeta, R. M. (2011). Contest design: An experimental investigation. *Economic Inquiry*, *49*(2), 573–590.
- Skaperdas, S. (1996). Contest success functions. *Economic theory*, *7*(2), 283–290.
- Skaperdas, S., & Vaidya, S. (2012). Persuasion as a contest. *Economic theory*, *51*(2), 465–486.
- Tesler, M., & Sears, D. O. (2010). *Obama's race*. University of Chicago Press.
- Tullock, G. (1967). The welfare costs of tariffs, monopolies, and theft. *Economic Inquiry*, *5*(3), 224–232.
- Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on stem tenure track. *Proceedings of the National Academy of Sciences*, *112*(17), 5360–5365.

## A 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 :

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

In the Nash Equilibrium of this contest:

Equilibrium efforts: 
$$e_{a2} = \frac{f}{(f+1)^2}V, e_{b2} = \frac{f}{(f+1)^2}V \quad (1)$$

Equilibrium probabilities of winning: 
$$p_2^a = \frac{f}{f+1}, p_2^b = \frac{1}{f+1}$$

Equilibrium expected utilities: 
$$U_2^a = \frac{f^2}{(f+1)^2}V, U_2^b = \frac{1}{(f+1)^2}V \quad (2)$$

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

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

In the Nash Equilibrium of this contest:

Equilibrium efforts: 
$$e_{a2} = \frac{\alpha f}{(\alpha f + 1)^2}V, e_{b2} = \frac{\alpha f}{(\alpha f + 1)^2}V \quad (3)$$

Equilibrium probabilities of winning: 
$$p_2^a = \frac{\alpha f}{\alpha f + 1}, p_2^b = \frac{1}{\alpha f + 1}$$

Equilibrium expected utilities: 
$$U_2^a = \frac{(\alpha f)^2}{(\alpha f + 1)^2}V, U_2^b = \frac{1}{(\alpha f + 1)^2}V \quad (4)$$

From (1) and (3), we know that  $e_{a2} = e_{b2}$  and is increasing in  $\alpha$  (as  $\alpha < 1$ ). This proves proposition 1.2.

The first period contest success function is given by :

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

Therefore, they maximize their expected utilities given by,

$$U_1^a(e_{a1}, e_{b1}) = \frac{fe_{a1}}{fe_{a1} + e_{b1}} V_a - e_{a1}$$

$$U_1^b(e_{a1}, e_{b1}) = \frac{e_{b1}}{fe_{a1} + e_{b1}} V_b - e_{b1}$$

where,

$$V_a = \left[ 1 + \beta_a \left( \frac{f^2}{(f+1)^2} - \frac{(\alpha f)^2}{(\alpha f + 1)^2} \right) \right] V \quad \text{and} \quad V_b = \left[ 1 + \beta_b \left( \frac{1}{(\alpha f + 1)^2} - \frac{1}{(f+1)^2} \right) \right] V$$

This generates the following equilibrium outcome of period 1,

$$\text{Equilibrium efforts:} \quad e_{a1} = \frac{fV_a^2V_b}{(fV_a + V_b)^2}, \quad e_{b1} = \frac{fV_b^2V_a}{(fV_a + V_b)^2} \quad (5)$$

Equilibrium probabilities

$$\text{of winning:} \quad p_1^a = \frac{fV_a}{fV_a + V_b}, \quad p_1^b = \frac{V_b}{fV_a + V_b} \quad (6)$$

Equilibrium expected

$$\text{utilities:} \quad U_1^a = \frac{fV_a}{fV_a + V_b} \left[ V - \frac{fV_aV_b}{fV_a + V_b} \right], \quad U_1^b = \frac{V_b}{fV_a + V_b} \left[ V - \frac{fV_aV_b}{fV_a + V_b} \right] \quad (7)$$

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,

$$\begin{aligned} \frac{\partial \left( \frac{V_a}{V_b} \right)}{\partial \alpha} &= \frac{2f(f+1)^2(f\alpha+1) - f^2\beta_a(\alpha+2f+1) + f^2\beta_a(1-\alpha)}{(f+1)^2(f\alpha+1)^2 + f\beta_b(1-\alpha)(f\alpha+f+2)} \\ &\quad - \frac{(-f\beta_b(f\alpha+f+2) + 2f(f+1)^2(f\alpha+1) + f^2\beta_b(1-\alpha))((f+1)^2(f\alpha+1)^2 + f^2\beta_a(1-\alpha)(\alpha+2f+1))}{((f+1)^2(f\alpha+1)^2 + f\beta_b(1-\alpha)(f\alpha+f+2))^2} \\ &< 0 \quad \text{if } \frac{\beta_a}{\beta_b} > I \\ &\text{U shaped} \quad \text{if } \frac{\beta_a}{\beta_b} \in (1, I), \text{ change of slope at } \alpha = \alpha^*(f) \\ &> 0 \quad \text{if } \frac{\beta_a}{\beta_b} < 1 \end{aligned}$$

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.

## B B. Micro-foundation 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  $c_A(\cdot)$  and  $c_B(\cdot)$ . I assume that  $c'_i(\cdot) > 0$ ,  $c''_i(\cdot) \leq 0$ , 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<sup>13</sup>.

<sup>13</sup>The effect of non-neutral risk preferences is ambiguous (Sheremeta, 2011).

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 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<sup>14</sup>.

**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:

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

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,  $\frac{e_a}{e_b} = \frac{\partial c_b / \partial c_b}{\partial c_a / \partial c_a} = \frac{MC_b(e_b)}{MC_a(e_a)}$  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

$$D_i = (P_i | f = 1) - P_i$$

---

<sup>14</sup>Or, higher the difference between  $f$  and 1, higher is the initial bias in favor of A.

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 (G. M. 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.<sup>15</sup> 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 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

---

<sup>15</sup>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.

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.

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

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 functions of period one and two, which are analyzed in the paper.