RANKING PROBLEMS IN THE PRESENCE OF IMPLICIT BIAS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science

by

Magd Bayoumi

December 2022

RANKING PROBLEMS IN THE PRESENCE OF IMPLICIT BIAS

Magd Bayoumi, M.S.

Cornell University 2022

Implicit bias is the unconscious attribution of particular qualities (or lack of) to a member from a particular social group (e.g. defined by race or gender). Studies on implicit bias have shown that these unconscious stereotypes can have adverse outcomes in various social contexts, such as job screening, teaching, or policing. This dissertation advocates for an application of fairness based re-ranking methods to improve the fairness to all items which, to some surprise, comes with little cost to or can even improve the utility.

We present our key contributions in ranking when in the presence of implicit bias. This includes the development of a theorem where we prove that under simplifying assumptions on the utilities of items, simple, well-studied, constraints can ensure that the utility does not decrease with respect to a naive ranking. Finally, we augment our theoretical results with empirical findings on real-world distributions from the IIT-JEE (2009) dataset.

BIOGRAPHICAL SKETCH

Magd was born in Kansas before growing up in Pennsylvania and, more recently North Carolina. He came to Ithaca, New York in 2017 for his undergraduate studies in Computer Science at Cornell University where his interest in Machine Learning and its connections to the world grew. After three wonderful years in Ithaca, he continued at Cornell for his M.S. in Computer Science at Cornell where he has been fortunately advised by Prof. Thorsten Joachims. During his M.S. studies, he was a co-lecturer for the undergraduate Natural Language Processing course and an advisor for multiple Master of Professional Studies (MPS) and undergraduate projects within Machine Learning. His research interest includes data-driven decision making problems, including contextual bandits and reinforcement learning as well as their impact individuals and the world. He additionally has a large interest within Natural Language Processing and Machine Learning systems. After graduating from Cornell, he will join the Integrated Analytics team at MunichRE as a Machine Learning Engineer working on ensuring the safe deployment, monitoring, development, and success of production machine learning systems.

To my family, for making me who I am,
and everyone who has helped me along the way.

ACKNOWLEDGEMENTS

First and Foremost, I am deeply indebted to my advisor, Thorsten Joachims, for his invaluable guidance, dedicated support and endless kindness. His influence on my academic life has been gigantic in terms of demonstrating me how to conduct first-class research, shaping my research tastes and offering me tremendous help in my career. His sharp insights and pioneering vision in the field have always inspired me to think about the broader impact of my research, and study the problems that are beneficial to the society and human-beings in a long term. Beyond academic support, he is very generous in sharing his life experience and advice, and always willing to help whenever needed. I could not have imagined having a better advisor for my M.S. study.

I am also indebted to Claire Cardie for her support as my academic committee member. She has been a tremendous influence demonstrating how to present and share topics of technical complexity in a digest-able manner with Natural Language Processing. Her faith in my lecturing, and weekly advice was invaluable.

Thanks should also go to my fellow machine learning colleagues: Yi Su, Ashudeep Singh, Luke Wang, and Aaron Tucker for our collaborations and weekly thought provoking group meetings.

Finally, I owe the most to my family, my parents and my sister. Thank you for raising me with all the love and kindness in the world. Thank you for letting me know how valuable I am and supporting me unconditionally. I could not have any achievement without your enduring love and support. This thesis is dedicated to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Over the past two decades, *implicit bias* [38] has become a center point of discussion on discrimination. Research within the field of implicit bias argues that unconscious attitudes towards members of differing demographic groups - defined by gender, race, ethnicity, and other characteristics - can have a significant impact on the way members of these groups are evaluated. This may affect outcomes across a wide variety of fields and societal institutions [5, 11, 51], as also highlighted by recent events in the popular press [7, 46, 75]. For example, in employment, men are perceived as more skilled and given a higher starting salary even when qualifications are the same [67], and in leadership positions, it was observed that women had to demonstrate roughly twice as much evidence of skill as men to be seen as equally skilled [45, 73]. This trend is observed in a variety of employment settings over the years with other demographic groups [13, 16, 67]. This impact in outcomes occurs across many institutions such as education [12], law [37, 64], and medicine [36].

To fight these biases, a significant effort has been placed into anti-bias training with the goal of eliminating or reducing implicit biases [6, 3, 82]. Such programs have been shown to have limited efficacy [50]. Furthermore, algorithms are increasingly taking over prediction and ranking tasks. For example, when shortlisting candidates [15, 8, 41, 74, 1] algorithms can learn from and encode existing biases present in prior hiring data against gender [27] or race [70]. This results in new algorithmic biases that disproportionately impact particular groups. Thus, it is important to explore interventions or algorithms that can mitigate these implicit biases and result in better outcomes.

As a running example, we will consider the process of recruitment or hiring. Although we use this as an example, these interventions would apply to any domain where people or items are selected or ranked such as school admissions or artist lineups for a festival. Hiring usually has multiple stages: first the applicants are ranked in order of their perceived relevances, then a short list of candidates are interviewed before finally one or more are hired [15]. The Rooney Rule is a commonly known rule to combat biases during the shortlisting phase. It was introduced by the National Football League [24] and later adopted by other industries [20, 2, 52, 60]. The main idea is to include underprivileged candidates with a higher (hidden) potential in the shortlist. This provides these candidates the opportunity to be interviewed. Whereas without this rule, these candidates may not have been selected for an interview and thus be prematurely declined (based on a potentially biased perception) without fully assessing their qualifications.

Kleinberg et al. [42] study the Rooney Rule under a theoretical model with two groups. They characterize conditions under which the Rooney rule increases the utility of the selection. However, before the shortlist is created the applicants must first be ranked. For example, LinkedIn Recruiter predicts a candidate's "likelihood of being hired" from their activity, Koru Hire analyzes a candidates (derived) personality traits to generate a "fit score", and HireVue "grades" candidates to produce an "insight score" [15]. In a ranking, as well as in [42], the candidates are sorted and the utility is defined by a weighted sum of the true utilities of the ranked candidates where the weight decreases the further down the ranking the applicant is places. This weighting in rankings is common practice and due to the fact that candidates lower in the list receive less attention compared to items placed higher in the list[40]. This in turn translates into

being less likely to be shortlisted, and contributing less to the total utility. Even a small bias can have a large impact on the individuals. If the top k candidates are close in skill-level, even with a minuscule bias the disadvantaged group will consistently rank lower and lead to extreme differences in exposure. This can be seen in Appendix Table A.3. Thus it is important to understand strategies to mitigate bias in the ranking phase and understand their effectiveness.

1.1 Main contributions

This dissertation presents a generalization of the implicit bias model in Kleinberg et al. [42] and interventions to recover utility while improving fairness. We consider a re-ranking procedure that maximizes the biased utility under a fairness constraint. We show that, under certain conditions, there is a re-ranking which will maintain the utility while increasing the fairness of the ranking from the system. This theorem provides a lower bound on the performance of a generalized version of Rooney Rule-like intervention when there are two groups (i.e., we allow stochastic rankings rather than a deterministic ranking or selection).

We then empirically show under a wide variety of conditions that the re-ranking that maximizes biased utility under a fairness constraint has an even higher utility while improving fairness, providing an advantage in both respects. We evaluate the performance of ranking under this procedure on two real-world datasets, the IIT-JEE 2009 dataset and the New York City SAT scores datasets. In both cases, we observe that this procedure improves the true utility of the ranking. While we discuss these results in the context of hiring, or

admission into some institution such a re-ranking procedure could be effective whenever the observed utilities are biased against a particular group.

1.2 Organization

The overall structure of the thesis is as follows.

Part I: Overview provides the introduction and background knowledge in this dissertation. We give brief introductions to the ranking problem formulation, fairness, and implicit biases. After this, we give a comprehensive examination of the current fair ranking (and selection) literature, which includes notions of fairness, fair ranking methods, impact of implicit biases and interventions in the presence of implicit bias.

Part II: Ranking in the presence of implicit biases consists of two parts. In chapter 3 we introduce a framework and present a theoretical analysis of the utility and fairness of a common prior fairness approach in a setting with implicit biases. In chapter 4, we move a step further and apply this technique within a wide variety of synthetic experiments. We validate these results on two real-world datasets.

CHAPTER 2

BACKGROUND ON FAIR RANKING AND SELECTION

We provide general background and literature review for fair ranking and selection in this section. Further literature related to the specific approaches is introduced in the corresponding chapters.

Designing effective ranking, recommendation or retrieval systems requires tackling similar challenges to general machine learning based classification systems - with additional challenges that stem from the fact that these systems make comparative judgements across items; a high position in the ranking is a limited resource. Information retrieval systems often have some mechanism (e.g. a machine learning model or test) to estimate the *relevance* (or *probability of relevance*) of the items [9, 44]. While user utility is generally the broader objective [55], the *Probability Ranking Principle* [57] is heavily used towards this end. This places items in descending order of their probability to be relevant to the user. For a variety of user utility metrics - such as mean average precision [68], mean reciprocal rank [69], and cumulative gain based metrics [40, 39] - this principle maximizes the expected utility to the users [39].

Furthermore, users tend to click more on higher positioned items which are more (estimated to be) relevant items (according to the Probability Ranking Principle). This *position bias* [26] means that exposure (*expected attention*) to items decreases while moving from the top rank to the bottom one; for example users may evaluate items from the top rank until they find a satisfactory one.

It is thus important to be ranked highly as a small difference in relevance estimation could result in a large difference in exposure (see Appendix Table

A.3). Depending on the ranking context, (e.g. ranking movies vs ranking job candidates) high ranking positions increase the likelihood of rewards.

2.1 Fairness in Information Retrieval

Due to the importance of rankings for providers (i.e. song artists or job candidates) and as part of the increased focus on machine learning impacts, there has been a lot of interest in fairness and equity for providers rather than just utility for the users. There are a multitude of definitions, criteria, and evaluation metrics to estimate the fairness of a system [25, 19, 31, 43, 47, 48, 78]. Given the complex environment in which retrieval / recommender systems are built, and the multitude of stakeholders involved that may have differing goals [34], and worldviews [56], there is no universal definition of fairness. However, many definitions can generally be classified into two categories: whether we want to treat similar individuals similarly (*individual fairness*) [30] or if different groups of individuals, defined by some characteristic such as demographic information, should be treated in a similar manner (*group fairness*) [148]. Dwork et al discuss [30] discuss the connection between these two ways of considering fairness. In Dwork et al [30], Pitoura et al [54], and Yang et al [77] they point out that group fairness does not guarantee individual fairness. However, Dwork et al [30] and Biega et al [14] show that under certain conditions that individual fairness may imply group fairness.

Fairness notions from classification can - to some extent - be utilized within the ranking setting. They typically only require additional consideration of the comparative nature of rankings and of how utility is modeled [18]. In com-

parison to a relevance-only ranking, adding fairness considerations often leads to a multi-objective or a constrained objective, where the usual utility objective comes with a fairness constraint or objective focused on the providers [56, 76].

One set of works [10, 22, 35, 79, 80] reasons about probability-based fairness in the top-k ranking positions putting the focus onto group fairness. These works usually provide a minimum (and for some cases a maximum) number or proportion of items/individuals from a protected group that are to be distributed evenly across the ranking. These methods do not usually allow later compensation if the fairness constraints are not met at the top-k positions.

Singh and Joachims [61] propose to view fairness from a perspective of exposure. Singh and Joachims [61] assign exposure scores (sometimes called attention) to each ranking position based on expected user click probability [23]. They argue that exposure is a limited resource on any platform (due to position bias) and advocate for a fair distribution of exposure to ensure fairness to providers. Geyik et al. [35] extend the prior work further executing a large scale A/B test on the LinkedIn network. They find that with fairness aware rankings, there is no change within their business metrics however the presented individuals are now more representative. Other works follow a similar style [14, 28, 65, 81] of assigning exposure scores to consider fairness within ranking systems. In contrast to the probability based fairness these methods have brought up group fairness [61, 49], and also definitions for individual fairness [61, 14, 17]. Additionally they do allow compensations in lower positions as well.

In Singh and Joachims [61], they consider a few different notions of fairness common within classification. We utilize the definitions of fairness from Singh

and Joachims [61] later in this work. The definition of the demographic parity fairness constraint is in Constraint 1, disparate treatment fairness constraint in Constraint 2, and disparate impact fairness constraint in Constraint 3. Our view of fairness is motivated by [61], focusing on the perspective of exposure.

Although lots of prior work [61, 28, 14, 65, 81] and our work center around ensuring fairness in a ranking at a particular point in time, there are exceptions. Biega et al [14], Suhr et al [63], and Surer et al [65] propose to ensure fairness through equity in amortized exposure i.e. over multiple instances of ranking. In our work, we focus on ensuring fairness in a particular point of time.

We utilize the following definitions from [61] for utility, fairness constraints and how we optimize fair rankings.

Definition 1 (Utility of a ranking)**.** Given an exposure function $\mathbf{v} : \mathbb{N} \rightarrow (0, 1)$, a relevance function $rel : C \rightarrow (0, 1)$, and a ranking R then the utility is defined as:

$$U(R) = \sum_{c \in C} \mathbf{v}(rank(c \mid R)) rel(c) \tag{2.1}$$

Constraint 1 (Demographic parity)**.** This constraint enforces that the average exposure of the items in both the groups is equal. We define a group $G_k = [g_1, \ldots, g_n]$ of items as well as an exposure function $\mathbf{v} : \mathbb{N} \rightarrow (0, 1)$, and a ranking R. Denoting average exposure in a group with

$$Exposure(G_k \mid R) = \frac{1}{|G_k|} \sum_{g_i \in G_k} \mathbf{v}(rank(g_i \mid R)) \tag{2.2}$$

The constraint can be expressed as:

$$Exposure(G_0 \mid R) = Exposure(G_1 \mid R) \tag{2.3}$$

Constraint 2 (Disparate treatment)**.** This constraint treats allocation of exposure as a treatment, and assigns exposure proportional to the relevance of a particular group. We define a group $G_k = [g_1, \ldots, g_n]$ of items as well as an exposure function $\mathbf{v} : \mathbb{N} \to (0, 1)$, a relevance function $rel : C \to (0, 1)$, and a ranking R. Denoting average relevance (or merit) of a group as:

$$rel(G_k) = \frac{1}{|G_k|} \sum_{g_i \in G_k} rel(g_i) \tag{2.4}$$

The constraint is expressed as:

$$\frac{Exposure(G_0 \mid R)}{rel(G_0)} = \frac{Exposure(G_1 \mid R)}{rel(G_1)} \tag{2.5}$$

Constraint 3 (Disparate impact)**.** This constraint assures that the click through rates for the groups as determined by the exposure and relevance are proportional to their average merit. To formally define this, let us first model the probability of a document getting clicked according to the following simple click model [23]:

$$\begin{aligned} P(\text{click on item i}) &= P(\text{examine item i}) \times P(\text{item i is relevant}) \\ &= Exposure(g_i \mid R) \times rel(g_i) \end{aligned} \tag{2.6}$$

We can now compute the average click through rate of items in group G_k as:

$$CTR(G_k \mid R) = \frac{1}{|G_k|} \sum_{g_i \in G_k} Exposure(g_i \mid R) \times rel(g_i) \tag{2.7}$$

The following Disparate Impact Constraint enforces that the expected click through rate of each group is proportional to its average merit:

$$\frac{CTR(G_0 \mid R)}{rel(G_0)} = \frac{CTR(G_1 \mid R)}{rel(G_1)} \tag{2.8}$$

2.2 Implicit bias

As discussed in the prior subsection 2.1, there are quite a few works designing algorithms for fair rankings. However the primary goals of these works is to design algorithms that satisfy the constraints towards satisfying some definition of fairness. A key thing to note is that the relevances of the providers are assumed to be unbiased. Geyik et al. [35] made this assumption in their LinkedIn A/B test when applying fairness to the rankings. They found that the application of fairness had a negligible impact on business metrics. However for a negligible impact in business metrics it may be the case that there may be some misestimation (either due to an implicit bias, noise, or other reason). In contrast to these works, we begin with the premise that the utilities are systematically incorrect due to implicit bias and use fair ranking algorithms to mitigate the effect of these biases when constructing a ranking.

Within the field of psychology, studying implicit bias is a common topic of study [38, 37]. Several works study the origins of implicit bias [53, 58] and its adverse impacts [45, 59, 73]. We refer the reader to a seminal work on implicit bias [38]. We will consider a model of implicit bias inspired by [72, 42]; however other models may exist and exploring other biases and how to mitigate them could lead to interesting future expansions of this work.

A commonly used mechanism for addressing implicit bias within hiring is the Rooney Rule [24] which requires (when recruiting) that one of the candidates being interviewed must come from an underrepresented group. This rule is a protocol adopted by the National Football League (NFL) in 2002 over concerns regarding low representation of African-American individuals in head coaching positions. The Rooney Rule has even become a guideline in many areas of business [20]; for example, in recent years companies including Amazon, Facebook, Microsoft, and Pinterest have adopted some version of the Rooney Rule requiring that at least one candidate interviewed must be a woman or a member of an underrepresented minority group [52]. Within the NFL, there has been an increase in minority coaches after the creation of this rule as seen in Figure 2.1. However, this is a subject of debate on whether the Rooney Rule does produce a better outcome in terms of utility. One side is that implicit biases prevent deserving candidates from being fairly considered which the Rooney Rule helps correct for. On the other side, is the concern that if a candidate short-list contains only candidates from the majority group then it may be due to them being the strongest candidates despite the underlying bias. This is especially true if there is a shortage of candidates from other groups. In this case, the Rooney Rule may lead to consideration of weaker candidates from underrepresented groups which hurts the utility and working against the elimination of unconscious biases.

To better understand the impact of the Rooney Rule from a theoretical standpoint, Kleinberg et al. [42] introduces a model for implicit bias and under this model characterize where the Rooney Rule improves the true utility of the the selection. They additionally perform a theoretical analysis to find conditions under which the Rooney Rule improves the quality of the outcome. Celis et

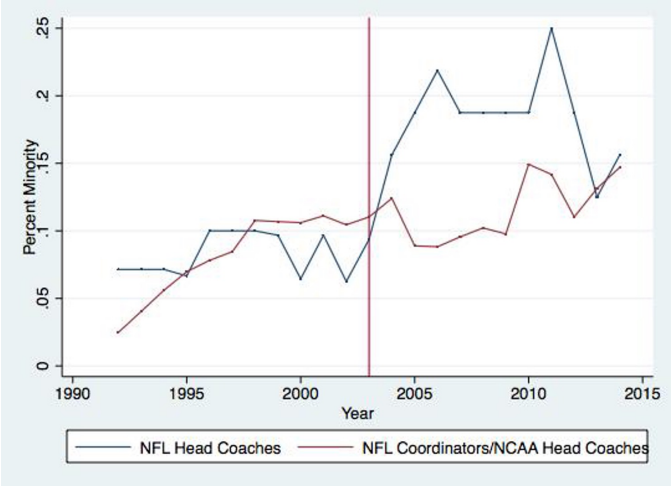

Figure 2.1: Percent minority coaches over time with the red vertical line denoting the implementation of the Rooney Rule. Source: [29]

al. [21] study the ranking problem (a generalization of selection) under implicit bias and propose simple constraints on rankings which improve the true utility of the ranking. They study a specific set of Rooney Rule constraints requiring that at least L items from the minority group be in the top k positions of a ranking. They find that they are able to increase utility in the presence of implicit bias while increasing the fairness (by inclusion of minority group members) as well. However, they consider a deterministic ranking which enforces a strict Rooney Rule. We further generalize these constraints to consider specific fairness metrics that we wish to achieve, and allow the rankings to be stochastic. This allows the ranking system design a greater degree of control in how fairness is measured (i.e., what does it means to be fair?) while maintaining a high level of utility.

Emelianov et al [32] study selection problems under a different model of bias: where the observed utility has higher than average noise for underrepresented candidates. They use a family of constraints and show that these constraints always increase the true (or latent) utility.

We consider the model of bias presented as presented in [21, 42]. Both of these works consider deterministic settings whereas we are interested in extending this to stochastic rankings which are a generalization of [21].

CHAPTER 3

THEORETICAL FINDINGS

3.1 Problem formulation

Suppose we have 2 groups (i.e. A and B or Advantaged and Biased) of candidates that are scored by some model $\mathcal{M}$. If there was no bias, then a greedy ranking of these candidates would yield the optimal utility (but not necessarily the optimal fairness). We define fairness as disparate treatment which is the exposure of group A divided by the relevance (or merit) of group A. This ratio will be the same across all groups to achieve fairness in disparate treatment

We additionally assume without loss of generality, that the Advantaged and Biased groups are sorted such that the merit of $A_1 \geq$ the merit of A_2 (i.e. $rel(A_1) \geq rel(A_2)$). This ordering is for simplicity in naming, and for illustration but not strictly required.

We assume the Advantaged and Biased relevance values come from the same arbitrary relevance distribution (e.g. Power law, uniform, linear interpolation). We additionally assume that both sets of candidate groups are non-empty, and that the exposure curve is given by an arbitrary monotonically decreasing function $\mathbf{v}(l)$ which depends only on the rank i as in the Position-Based Model [23]. Lastly, we assume there is an additive bias β which is ≥ 0 on the biased group. We define $rel_\beta(C)$ as the biased relevance of candidate C or $rel_\beta(C) = rel(C) - \beta$.

3.2 Pair-wise constraint increases fairness and maintains utility

Out first result shows that there are a simple class of constraints that can increase the fairness while maintaining the same level of utility as a naive ranking. Let $exp_G(A_i)$ denote the exposure given to advantaged group member i under the greedy ranking G and similarly for the biased group members. Let $exp_F(A_i)$ denote the exposure given to advantaged group member i under the fair ranking F and similarly for biased group members. We pair candidates of different groups together based on their rank within the group (i.e. A_1 is paired with B_1 and so on). We constrain the exposure now given to A_i and B_i must still equal the sum of their exposure in the naive ranking. Notionally, this constraint looks like:

$$exp_G(A_i) + exp_G(B_i) = exp_F(A_i) + exp_F(B_i) \tag{3.1}$$

The intuition behind this constraint comes from the relevance distribution being the same for advantaged group members and biased group members. The top advantaged member and top biased member will have the same true relevance although the biased groups observed relevances is potentially lower. Since their true relevance is the same, how we allocate exposure between the pair does not impact any utility metrics such as Discounted Cumulative Gain [40] but will impact the fairness metric of disparate treatment.

Pair-wise constraint construction

We let $k = min(|A|, |B|)$. We enforce the following equality $exp_G(A_i) + exp_G(B_i) = exp_F(A_i) + exp_F(B_i)$ for $i \leq k$. Additionally, we enforce that the exposure given

to A_i and B_i can only be derived from their original positions in ranking G. As an example, if A_1 was in position 1 and B_1 was in position 3 then A_1 and B_1 can only derive their exposure from positions 1 and positions 3 within the new fair ranking (i.e. they may not utilize any other positions). We derive a fairness constraint under the biased relevances that is later proven in Theorem 1 to not decrease utility.

We will satisfy the disparate treatment constraint which is given by:

$$\frac{\sum_{B_i \in B} exp(B_i)}{\sum_{B_i \in B} rel_\beta(B_i)} = \frac{\sum_{A_i \in A} exp(A_i)}{\sum_{A_i \in A} rel(A_i)} \tag{3.2}$$

The amount of relevance re-allocated between the Advantaged and Biased groups defined below as η is parameterized by the exposure $exp(\cdot)$ and relevance functions $rel(\cdot)$

Definition 2 (Relevance of group)**.** Given a function $rel : C \to (0, 1)$ that maps a candidate to a relevance value (or probability), the relevance of a group $B = \{B_1, \ldots, B_m\}$ is $rel(B) = \sum_{B_i \in B} rel(F_i)$. Similarly for $A = \{A_1, \ldots, A_m\}$.

Definition 3 (Exposure of group)**.** Given an exposure function $exp_R : C \to (0, 1)$ that maps a candidate to their exposure under the ranking R, the exposure of a group is $exp_R(B) = \sum_{B_i \in B} exp_R(B_i)$ and . Similarly for the group A.

Definition 4 (Total exposure)**.** Given a position-based exposure function, the total exposure is $\sum_{l=1}^{|A|+|B|} \mathbf{v}(l)$ (sum of exposure of all positions), which we write as exp_{tot}.

Definition 5 (Re-allocated exposure η)**.** η is the amount of exposure shifted from the advantaged group A to the biased group B when using the fair ranking and compared to the biased ranking. Mathematically, this is given by $exp_F(B) -$

$exp_G(B)$, which is the exposure change for the biased group between the two rankings.. It is equivalently defined as $exp_G(A) - exp_F(A)$.

We solve for our re-allocated exposure η starting with the initial set of constraints that we must satisfy with our ranking. These are the fairness constraint (disparate treatment) and the total allocated exposure cannot exceed the total available exposure. These are given below

$$\frac{exp_F(B)}{rel_\beta(B)} = \frac{exp_F(A)}{rel(A)} \tag{3.3}$$

$$exp_{tot} = exp_F(B) + exp_F(A) \tag{3.4}$$

Utilizing these constraints, we can define $exp_F(B)$ and $exp_F(A)$ in terms of only the relevance of the groups and the total exposure which allows us to numerically solve for η more easily. We first rearrange the constraint in Equation 3.3 before plugging this into Equation 3.4

$$\begin{aligned}
exp_F(B) &= \frac{exp_F(A)}{rel(A)} rel_\beta(B) \\
exp_F(B) + exp_F(A) &= exp_{tot} \\
\frac{exp_F(A)}{rel(A)} rel_\beta(B) + exp_F(A) &= exp_{tot} \\
exp_F(A)(\frac{rel_\beta(B)}{rel(A)} + 1) &= exp_{tot} \\
exp_F(A) &= \frac{exp_{tot}}{\left(\frac{rel_\beta(B)}{rel(A)} + 1\right)} \\
exp_F(A) &= \frac{exp_{tot}}{\left(\frac{rel_\beta(B) + rel(A)}{rel(A)}\right)} \\
exp_F(A) &= \frac{exp_{tot} \times rel(A)}{rel_\beta(B) + rel(A)}
\end{aligned} \tag{3.5}$$

We can do something similar to Equation 3.5 for $exp_F(B)$ and arrive at the following definition,

$$exp_F(B) = \frac{exp_{tot} \times rel(A)}{rel_\beta(B) + rel(A)} \tag{3.6}$$

The reallocation defined earlier can now be numerically solved using the definition: $exp_F(B) - exp_G(B)$. Assuming we have solved for our global re-allocation η, we define a pair specific re-allocation as a proportion of η. As a reminder, the equation for η is below:

$$\eta = exp_F(B) - exp_G(B) \tag{3.7}$$

$$= exp_G(A) - exp_F(A) \tag{3.8}$$

The proportion re-allocated per pair is defined in Equation 3.9. This equation is simply the proportion of the relevance of A_i contributes to the total relevance up to the k-th advantaged candidate. We can similarly use an equation defined by individuals within the B group however, this is unnecessary due to the equality given in Eq. 3.7.

$$\frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} \tag{3.9}$$

In general, the fair exposure for each candidate B_i or A_i (when $i \leq k$) is defined in Equation 3.10 which shifts the naive greedy exposure by a proportion (as defined in Equation 3.9) of the global re-allocation η.

$$
\begin{aligned}
exp_F(B_i) &= exp_G(B_i) + \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} \\
exp_F(A_i) &= exp_G(A_i) - \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}
\end{aligned}
\tag{3.10}
$$

For any group member whose lower than rank k within their group, their exposure will remain the same as in the greedy ranking.

The last component of this construction is assigning the allocation to each position given the exposure that each candidate has. For any group member who have a lower position than rank k within their group, their position in the ranking will remain the same as in the greedy ranking. For biased group candidates whose position is higher than rank k (i.e. they have been paired with a advantaged candidate who has a similar position within the advantaged group), suppose that the biased group candidate B_i is in position j of the naive greedy ranking and the advantaged group candidate A_i is in position l in the naive greedy ranking. The candidates B_i and A_i can only be allocated to position j and l within our construction. To allocate the appropriate amount, we solve a series of linear equations in Equation 3.11. These equations ensure that the appropriate exposure is assigned along with ensuring that the final ranking matrix is doubly stochastic.

$$\begin{bmatrix} \mathbf{v}(l) & \mathbf{v}(j) & 0 & 0 \\ 0 & 0 & \mathbf{v}(l) & \mathbf{v}(j) \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} exp_F(A_i) \\ exp_F(B_i) \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{3.11}$$

We can use this set of equations to solve for x_1, x_2, x_3, x_4. When the last 4 rows are solved, we arrive at:

$$x_1 = x_4$$

$$x_2 = 1 - x_4$$

$$x_3 = 1 - x_4$$

x_4 is free

Using this set of information, we explicitly solve for x_4 under the exposure equation(s) from the first two linear equations knowing that $0 \leq x_4 \leq 1$ to satisfy the doubly stochastic nature of the ranking matrix. We find the value of x_4 in Equation 3.12. Recall that x_4 is defined as the probability that B_j is placed in position j which means $0 \leq x_4 \leq 1$. Utilizing the boundary values that x_4 can take on (0 and 1), we can develop a condition to verify of β to ensure that the Theorem 1 holds.

$$
\begin{aligned}
x_4\mathbf{v}l + \mathbf{v}(j) - x_4\mathbf{v}(j) &= exp_F(A_i) \\
x_4 &= \frac{exp_F(B_i) - \mathbf{v}l}{\mathbf{v}(j) - \mathbf{v}l} \\
&= \frac{exp_F(A_i) - \mathbf{v}(j)}{\mathbf{v}l - \mathbf{v}(j)}
\end{aligned} \tag{3.12}
$$

A feasible solution exists whenever we know $0 \leq x_4 \leq 1$. Thus we can utilize this to solve for a condition to check with regards to β. We leave the conditions along with their derivations to the Appendix in Equations B.3 and B.5.

Theoretical analysis

We theoretically analyze the utility gap (i.e. difference in utility) between the defined pair-wise constraint construction and a naive greedy ranking.

Theorem 1. *Given a position based exposure function $\mathbf{v}k$, ranked candidates from two groups $A = [A_1, \ldots, A_n]$ and $B = [B_1, \ldots, B_m]$ whose relevance values stem from the same (deterministic) function depending only on group rank, and an additive bias $\beta \geq 0$, the pair-wise constraint will not decrease the utility. By construction, we ensure that the group fairness constraint is satisfied.*

Proof. We compute the utility for the two rankings as the exposure given to a candidate multiplied by the relevance of the candidate over all candidates. We can take the difference between the fair ranking and the greedy ranking which should be ≥ 0 to ensure the utility does not decrease. Within equation 3.13, we compute this difference

$$
\begin{aligned}
U(Fair) - U(Greedy) &= \left(\left(\sum_{B_i \in B} rel(B_i) exp_F(B_i)\right) + \left(\sum_{A_i \in A} rel(A_i) exp_F(A_i)\right)\right) \\
&- \left(\left(\sum_{B_i \in B} rel(B_i) exp_G(B_i)\right) + \left(\sum_{A_i \in A} rel(A_i) exp_G(A_i)\right)\right) \\
&= \left(\sum_{B_i \in B} rel(B_i)(exp_F(B_i) - exp_G(B_i))\right) + \left(\sum_{A_i \in A} rel(A_i)(exp_F(A_i) - exp_G(A_i))\right)
\end{aligned} \tag{3.13}
$$

From our construction, the only candidates that differ in exposure are the top k members of the A and B groups. All other candidates don't have a change in rank or exposure and hence within the respective summation the differences are 0. Thus we can further simplify equation 3.13 by removing all candidates which are not the top k candidates of their respective groups.

$$
U(Fair) - U(Greedy) = \left(\sum_{i=1}^{k} rel(B_i)(exp_F(B_i) - exp_G(B_i))\right) + \left(\sum_{i=1}^{k} rel(A_i)(exp_F(A_i) - exp_G(A_i))\right) \tag{3.14}
$$

We can further expand each of the $exp_F(B_i)$ (and $exp_F(A_i)$ respectively) from the definition before in Equation 3.10 to:

$$
\begin{aligned}
exp_F(B_i) &= exp_G(B_i) + \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} \\
exp_F(A_i) &= exp_G(A_i) - \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}
\end{aligned} \tag{3.15}
$$

Using this expansion, we can even further simplify to the difference in allocation based on our construction.

$$
\begin{aligned}
U(Fair) - U(Greedy) &= \left(\sum_{i=1}^{k} rel(B_i)(exp_G(B_i) + \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} - exp_G(B_i)) \right) \\
&+ \left(\sum_{i=1}^{k} rel(A_i)(exp_G(A_i) - \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} - exp_G(A_i)) \right) \\
&= \left(\sum_{i=1}^{k} rel(B_i)(\eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}) \right) + \left(\sum_{i=1}^{k} rel(A_i)(-\eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}) \right) \\
&= \sum_{i=1}^{k} \left((rel(B_i) - rel(A_i)) \left(\eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} \right) \right) \\
&= \sum_{i=1}^{k} \left(0 \left(\eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)} \right) \right) \\
&= 0
\end{aligned}
\tag{3.16}
$$

Since the advantaged and biased relevances are derived from a deterministic relevance function (based on group rank), in equation 3.16, we utilize the fact that the true gap in advantaged group / biased group relevance for the top k candidates is equal to 0. Thus the utility of the modified ranking is not lower than the utility of the naive greedy ranking.

To ensure that our ranking does indeed satisfy the disparate treatment constraint, we can utilize the exposure equations for our construction given in Equation 3.10 in conjunction with the disparate treatment constraint to verify that it is satisfied. The intuition comes from the construction where we explic-

itly took the difference between fairly allocated exposure and the naive greedy exposure for group A (or B) and use that difference to re-allocate our exposure.

$$
\begin{aligned}
\left| \frac{\sum_{B_i \in B} exp(B_i)}{rel_\beta(B)} - \frac{\sum_{A_i \in A} exp(A_i)}{rel(A)} \right| &= \left| \frac{\sum_{B_i \in B} exp_G(B_i) + \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}}{rel_\beta(B)} - \frac{\sum_{A_i \in A} exp_G(A_i) - \eta \cdot \frac{rel(A_i)}{\sum_{j=1}^{k} rel(A_j)}}{rel(A)} \right| \\
&= \left| \frac{exp_G(B) + \eta}{rel_\beta(B)} - \frac{exp_G(A) - \eta}{rel(A)} \right| \\
&= \left| \frac{exp_G(B) + (\frac{exp_{tot} \times rel_\beta(B)}{rel_\beta(B) + rel(A)} - exp_G(B))}{rel_\beta(B)} - \frac{exp_G(A) - (\frac{exp_{tot} \times rel(A)}{rel_\beta(B) + rel(A)} - exp_G(A))}{rel(A)} \right| \\
&= \left| \frac{\frac{exp_{tot} \times rel_\beta(B)}{rel_\beta(B) + rel(A)}}{rel_\beta(B)} - \frac{\frac{exp_{tot} \times rel(A)}{rel_\beta(B) + rel(A)}}{rel(A)} \right| \\
&= \left| \frac{exp_{tot}}{rel_\beta(B) + rel(A)} - \frac{exp_{tot}}{rel_\beta(B) + rel(A)} \right| \\
&= 0
\end{aligned}
\tag{3.17}
$$

Hence this construction does satisfy the disparate treatment constraint under the biased relevances. □

CHAPTER 4

EMPIRICAL FINDINGS

In this chapter, we empirically evaluate several key properties of our approach. We first present experiments on synthetic data which allows us to vary the properties of the candidate market to explore the robustness of the method. In addition, we also assess our method on a real-world dataset for external validity derived from scores in the IIT-JEE 2009 dataset.

4.1 Synthetic experiment setup

To examine how our method performs in comparison to baselines over a range of candidate markets with different characteristics, we create synthetic datasets as follows. The candidate market has n total participants with the biased group B making up a proportion α of those candidates. The relevance values are drawn from some function *rel* which, in the simplest case, is a deterministic function of the rank within a group such as $1/x$. Similarly, the exposure function $\mathbf{v}$ is a deterministic function of the position in the ranking following the Position-Based Model in other works [23, 61]. Lastly there is a bias β used to modify the relevance values of the Biased (or minority) group. The bias can be additive as in Chapter 3 where the modified relevance value becomes $rel_\beta(C_i) = rel(C_i) - \beta$ or it may be multiplicative as in other works [42, 21, 33] $rel_\beta(C_i) = \beta rel(C_i)$. If not mentioned otherwise, we use a multiplicative bias of $\beta = 0.8$ following [33] and their findings on a real-world SAT dataset. We additionally use the following values if not mentioned otherwise, the total participants $n = 100$, the examination function $\mathbf{v}x = 1/x$, linear relevance function

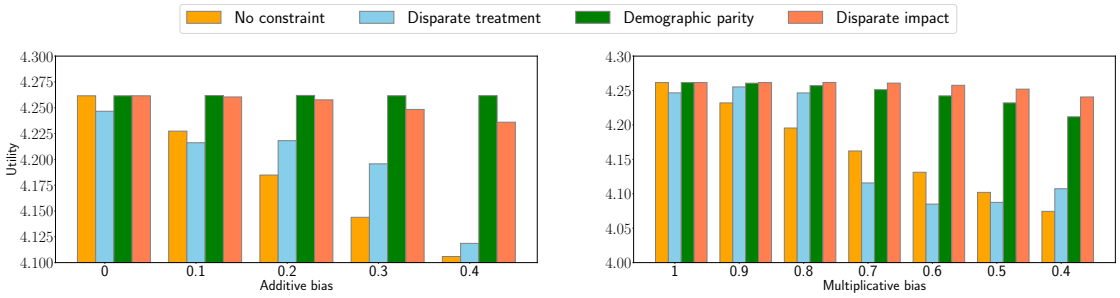

Figure 4.1: Additive and multiplicative bias value impact on interventions using synthetic dataset. The parameters are ($\beta = 0.8$, $v(x) = 1/x$, $n = 100$, $\alpha = 0.5$, $rel(x)$ is Linear from $1/n$ to 1), unless stated otherwise. The standard errors are on the order of 1e-2 and are invisible in the graph. The level of bias increases from left to right on both graphs.

of $rel(C_i) = 1 - \frac{i-1}{n}$, and a biased group proportion of $\alpha = 0.5$. We measure the quality for the various ranking policies by the utility as in other prior works [21, 33, 28, 40]. For each run, we estimate using 1000 Monte Carlo samples of the stochastic ranking policy and average over 10 runs.

We compare the naive greedy ranking against re-ranking procedures with the following fairness constraints disparate treatment, demographic parity, and disparate impact.

4.2 Synthetic experiment results

How do the methods perform for different levels of multiplicative biases?

With multiplicative bias, we refer to the level of bias within the simulation where a multiplicative bias β impacts the observed relevances as follows, $rel_\beta(C_i) = \beta rel(C_i)$. The rightmost graph in Figure 4.1 shows the utility for different levels of β. Especially for high levels of bias, we see that all the fairness

based methodologies perform better than a naive greedy ranking. Across all multiplicative biases, we see that demographic parity and disparate impact always perform similar or better. An interesting behavior to note is that disparate treatment does not consistently do better or worse than the naive ranking but varies with the bias value.

How do the methods perform for different levels of additive biases?

With additive bias, we refer to the level of bias within the simulation where a additive bias β impacts the observed relevances as follows, $rel_\beta(C_i) = rel(C_i) - \beta$. The leftmost graph in Figure 4.1 shows the utility for different levels of β. Especially for high levels of bias, we see that all the fairness based methodologies perform better than a naive greedy ranking.

How does the examination function influence the relative performance?

The examination function $\mathbf{v}(\cdot)$ models how many results people are able or willing to browse. A steep drop-off in examination probability, like $v(x) = 1/e^{x1}$, means that they are likely to only evaluate the top few results. A flat examination function, such as $v(x) = 1/log_2(x + 1)$, means that they are likely to go further down. The leftmost plot in Figure 4.2 shows the utility as we change the examination function. Across all examination functions, the fairness constraints perform better than having no constraint with little variance in performance. This suggests that similar to what was shown in Theorem 1, the exposure curve does not have a large impact on the resulting improvement.

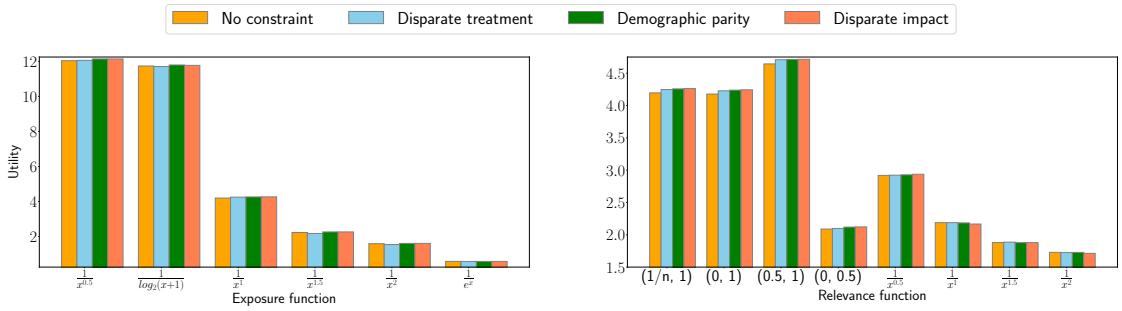

Figure 4.2: Exposure function and relevance function impact on interventions using synthetic dataset. The parameters are ($\beta = 0.8$, $v(x) = 1/x$, $n = 100$, $\alpha = 0.5$, $rel(x)$ is Linear from $1/n$ to 1), unless stated otherwise. The standard errors are on the order of 1e 2 and invisible in the graph.

How does the relevance function influence the relative performance?

The relevance function $rel(\cdot)$ models how skilled different candidates are. A steep drop-off in examination probability, like $rel(x) = 1/e^{x1}$, means that they are likely to only consider the most skilled few candidates. A flat examination function, such as $rel(x) = 1/log_2(x + 1)$, means that they are likely to consider more candidates since more are skilled. The rightmost plot in Figure 4.2 shows the utility as we change the relevance function. Across all relevance functions, the fairness constraints perform better than having no constraint with little variance in performance. This suggests that similar to what was shown in Theorem 1, the relevance curve does not have a large impact on the resulting improvement.

How does the size of the candidate pool affect the methods?

In this experiment, we vary the size of candidate pool to understand how this affects the effectiveness of the methods. Results are shown in the rightmost plot of Figure 4.3. As candidate pool size increases, all the ranking methods achieve higher utility, which is expected since there are more opportunities for relevant

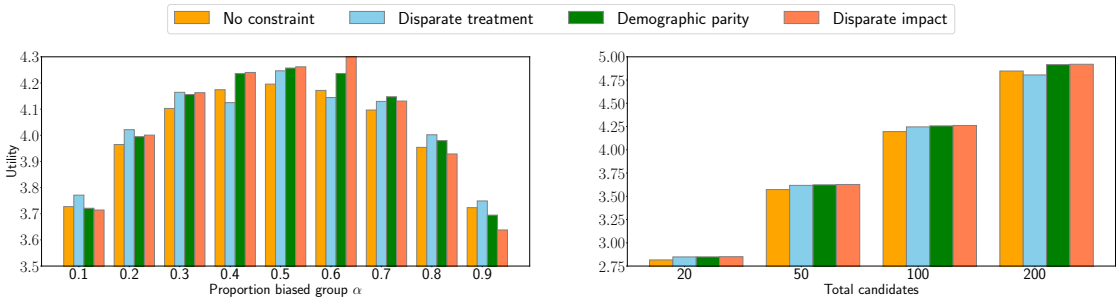

Figure 4.3: Candidate pool size and minority group proportion impact on interventions using synthetic dataset. The parameters are ($\beta = 0.8$, $v(x) = 1/x$, $n = 100$, $\alpha = 0.5$, $rel(x)$ is Linear from $1/n$ to 1), unless stated otherwise. The standard errors are on the order of 1e 2 and invisible in the graph.

candidates. More interestingly, the relative performance among the methods is largely unaffected by market size.

How does the size of the minority group affect the methods?

In this experiment, we vary the proportion of candidates that are from the minority group to to understand how this affects the effectiveness of the methods. Results are shown in the leftmost plot of Figure 4.3. At the tail ends of the plot where there is a large mismatch in group sizes, the disparate treatment intervention strongly outperforms all other methods. However, this trend is not true in the central region where groups are more similar in size with disparate impact and demographic parity being stronger performers. In some situations such as $\alpha = 0.4$, we see disparate impact actually under-performs no intervention.

How is fairness impacted across all of these experiments?

Within Appendix C, there are additional tables and figures to explore the fairness in terms of disparate impact. Across all synthetic experiments, at least

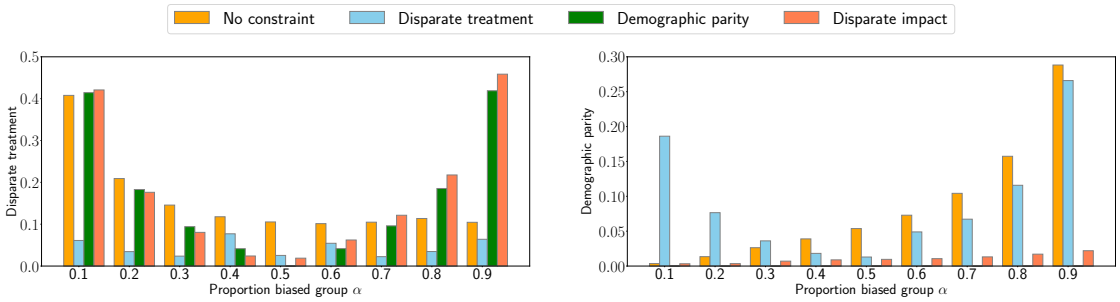

Figure 4.4: Minority group proportion impact on disparate impact and demographic parity using synthetic dataset. The parameters are ($\beta = 0.8$, $v(x) = 1/x$, $n = 100$, $\alpha = 0.5$, $rel(x)$ is Linear from $1/n$ to 1), unless stated otherwise. The standard errors are on the order of 1e 2 and invisible in the graph.

Table 4.1: Utility when Advantaged and Biased relevance distributions are different. The male or female distribution is shifted downward such that advantaged group tend to be more relevant (and vice versa). Standard error is omitted below as 2 stderr is less than $1e-2$

Ranking intervention	Higher advantaged relevance	Higher biased relevance
No constraint	4.17	4.13
Disparate treatment	4.12	4.09
Demographic parity	4.21	4.21
Disparate impact	4.20	4.21

one fairness-based method improved the fairness. If we keep $\alpha = 0.5$, then all fairness methods had increased fairness when compared to the naive greedy ranking. This suggests that more interesting behavior is seen in Table C.1 and Figure 4.4 in the regions where the two groups are of unequal size. In these regions, demographic parity and disparate impact actually increase the unfairness with respect to disparate treatment as they are over-allocating exposure relative to the relevance the group has. A smaller group will have a smaller overall relevance and hence would likely overall have a lower exposure.

How does this behavior change if advantaged and biased groups have different relevance distributions?

In this experiment, we shift either the advantaged group relevance or the biased group relevance by a constant of $\frac{1}{|A|}$ or $\frac{1}{|B|}$. This is to shift the relevance distribution to have a minimum value of 0 rather than $\frac{1}{|A|}$ or $\frac{1}{|B|}$. The utility for these two settings are noted in Table 4.1, while the unfairness is noted in Table C.6 within the Appendix. Something to note is that even with shifted distributions, the demographic parity constraint outperforms (or matches the performance of) all other methods. Further, the disparate treatment constraint worsens the utility likely due to a compounding effect of the shifted distribution and the bias factor. The upper limit of the relevance gap it can treat has been reached which raises further questions about how other differences in relevances may impact demographic parity or disparate impact. All methods do perform better in terms of fairness (disparate treatment) as seen in Table C.6.

4.3 Validation on Real World Data

We examine the effect of various constraints on naturally occurring distributions of utilities from scores in the IIT-JEE 2009 dataset. There are two disjoint groups of candidates M and F representing Male and Female candidates respectively. We first analyze the distribution of these scores D_M and D_F, and note that although similar in shape the female distribution has lower values. We assume the test is biased accounting for any systemic barriers that may exist.

If we assume that the scores of these candidates are a perfect measure of their

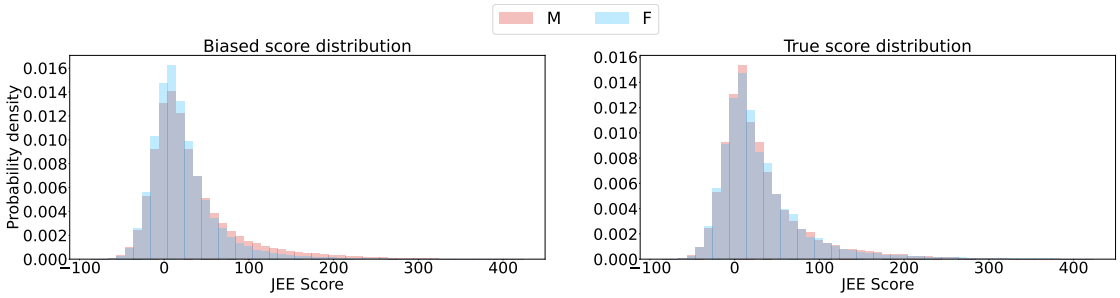

Figure 4.5: Distributions of scores in IIT-JEE 2009: Distribution of total scores of all male and all female candidates.

academic potential then any scheme to increase fairness will decrease the overall utility or potential of the candidates. However, of those with equal potential, those from underprivileged groups perform worse on standardized tests [71]. In India, fewer girls than boys attend primary school with many dropping out to work or get married [4]. This is a type of societal bias, which is different in nature than an implicit bias, but is another reason in which observed relevances can be systematically wrong against a particular group. Effectively, this means the scores are biased and the true relevance of a candidate from the female group is in fact larger than what is reflected in their score.

Dataset: JEE scores

These true and biased distribution of scores are depicted in Figure 4.5 The male and female distributions are very similar within the biased sample; However, the mean of men $\mu_M = 30.79$ (standard deviation $\sigma_M = 51.80$) is considerably higher than the mean of women $\mu_F = 21.24$ (standard deviation $\sigma_F = 39.27$).

Currently, undergraduate admissions into the IITs are decided solely on the basis of the scores attained in the Joint Entrance Exam. IIT-JEE is conducted once every year, and only students that have graduated from high school in the

previous two years are eligible. Out of the 468,280, candidates who took IIT-JEE in 2011, only 9627 candidates (2%) were admitted to an IIT. In the same year, 108, 653 women (23.2% of the total) appeared in the exam, yet only 926 were admitted into an IIT (less than 10% of the 9627 admitted) [66]. This dataset contains the scores of all the students on each section along with their gender. The candidates are scored on a scale from 35 to 160 points in all three sections, with an average total score of 28.36, a maximum score of 424 and a minimum score of 86. While the statistics of IIT JEE 2009 admissions are not available, we assume roughly the same number of students were admitted in 2009 as in 2011 and use this as our examination function. In other words our examination function is 1 for roughly the first 2% of positions and 0 everywhere else. This reduces our ranking problem into a stochastic selection problem instead. We further subsample the dataset to maintain the same gender and score distribution with a smaller candidate pool of 10,000 individuals.

Real-world experiment results

The improvement over no ranking intervention verifies that introducing various interventions can provide benefit in realistic applications even in complex conditions. Beyond the improvement in Utility, we find that the unfairness with respect to disparate treatment has also considerably decreased.

Table 4.2: Utility (± two stderr) and Disparate treatment (in expectation) on real-world dataset.

Ranking intervention	Utility	Disparate treatment (Unfairness)
No constraint	370.9 ± 0.07	0.34
Disparate treatment	371.4 ± 0.19	0.04
Demographic parity	370.4 ± 0.13	0.16
Disparate impact	370.5 ± 0.32	.30

CHAPTER 5

CONCLUSION AND FUTURE WORK

Search and Recommender systems are the arbiters of exposure in modern two-sided platforms such as LinkedIn or Etsy. Although such systems are able to generally produce relevant results, without considering the impact on both sides of the platform then it may negatively impact the users as well as the recommended entities on the platform. For the long-term well-being, ranking algorithms should be able to consider utility and fairness for both users as well as creators and producers. This is especially true in cases such as hiring in which there are implicit biases that may be amplified by such a recommender system.

This dissertation aims to provide a deep understanding of interventions that ranking systems can take to increase both the fairness for the items and utility of the user in the presence of a systematic bias such as an implicit bias. In particular, within Chapter 3, we present a sample construction to lower bound our ranking interventions. A theoretical analysis on this construction finds that we are able to re-allocate exposure between the two groups without decreasing the utility to the user. This allows systems with these interventions to attempt to reduce the unfairness within the platform while maintaining or increasing the utility to the user. We provide conditions which can be empirically checked to verify whether the bias β will allow for a possible solution.

We further validate the results in Chapter 3 with a wide array of synthetic experiments in Chapter 4 as well as validation on real-world data. We find that ranking interventions almost always improve the fairness while generally improving utility.

Some natural extensions along this line of research could be extending this work to a different set of models such as those that consider bias as differing levels of variance between groups or applying such ranking interventions in matching markets such as those in Su et al [62]. Additional extensions may consider Learning-to-Rank with Fairness Constraints with implicit biases or Fairness in Dynamic Learning-to-Rank settings with implicit biases. These pose additional challenges compared to the static re-ranking task since there may need to be additional changes to account for an additional bias that may impact utility (but may be revealed later).

APPENDIX A

EXAMPLE CONSTRUCTIONS

Here we give some toy examples relevant to our discussion in the paper. Table 1 gives an example on how position bias in ranking could further widen the already existing inequalities.

Rank	Exposure	Individual	Relevance	Group membership
1	0.5	A	0.92	Blue
2	0.25	B	0.91	
3	0.125	C	0.90	Red
4	0.0625	D	0.89	

Table A.1: Optimal greedy ranking

Group	Mean Relevance	Exposure
Blue	0.915	0.75
Red	0.895	0.1875

Table A.2: Group-level analysis

Table A.3: This is an example (inspired by Singh and Joachims [61] on how position bias can further widen inequalities. On a job platform there are four workers: A and B from the blue group while C and D are from the red group. For a certain employer, the platform wants to create a ranking of the workers. Suppose that all the workers are equally relevant to the employer. However, due to some implicit bias the relevance scores for the platform are as shown in Subtable A.1. Using the Probability Ranking Principle [57] , we can maximize utility by rankng in descending order as given in Subtable A.1. The second column of Subtable A.1 has the expected exposure of attention given for each rank. We give a group-level analysis in Subtable A.2. The mean relevance scores between the red group and blue group are not so different. On the other hand, the exposure to each group are quite different. From this example, we see how the optimal ranking can significantly widen the exposure gap even for a small gap in relevances.

APPENDIX B

BOUNDARY CONDITIONS FOR THEOREM 1

We extend the work done in 3 finding boundary conditions for x_4 from Equation 3.12. We replicate the equation below for reference. We wish to find boundary conditions for β when $x_4 = 0$ and $x_4 = 1$.

$$\begin{aligned} x_4 &= \frac{exp_F(F_i) - \mathbf{v}(l)}{\mathbf{v}(j) - \mathbf{v}(l)} \\ &= \frac{exp_F(M_i) - \mathbf{v}(j)}{\mathbf{v}(l) - \mathbf{v}(j)} \end{aligned} \tag{B.1}$$

When $x_4 = 0$,

$$\begin{aligned} x_4 &= \frac{exp_F(M_i) - \mathbf{v}(j)}{\mathbf{v}(l) - \mathbf{v}(j)} \\ 0 &= \frac{exp_F(M_i) - \mathbf{v}(j)}{\mathbf{v}(l) - \mathbf{v}(j)} \\ 0 &= exp_F(M_i) - \mathbf{v}(j) \\ exp_F(M_i) &= \mathbf{v}(j) \\ exp_G(M_i) - \eta \frac{rel(M_i)}{\sum_{j=1}^{k} rel(M_i)} &= \mathbf{v}(j) \\ \mathbf{v}(l) - \eta \frac{rel(M_i)}{\sum_{j=1}^{k} rel(M_i)} &= \mathbf{v}(j) \\ \frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) &= \eta \end{aligned}$$

We also can replace η with its definition:

$$
\begin{aligned}
\eta &= exp_F(F) - exp_G(F) \\
&= exp_G(M) - exp_F(M) \\
&= exp_G(M) - \frac{tot\_exp \cdot rel(M)}{rel_\beta(F) + rel(M)} \\
&= \frac{exp_G(M) rel_\beta(F) + exp_G(M) rel(M)}{rel_\beta(F) + rel(M)} - \frac{tot\_exp \cdot rel(M)}{rel_\beta(F) + rel(M)} \\
&= \frac{exp_G(M) rel_\beta(F) + exp_G(M) rel(M) - tot\_exp \cdot rel(M)}{rel_\beta(F) + rel(M)} \\
&= \frac{exp_G(M) rel_\beta(F) - exp_G(F) rel(M)}{rel_\beta(F) + rel(M)}
\end{aligned}
\tag{B.2}
$$

Thus,

$$
\begin{aligned}
\frac{exp_G(M) rel_\beta(F) - exp_G(F) rel(M)}{rel_\beta(F) + rel(M)} &= \frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \\
\frac{exp_G(M) rel(F) - exp_G(M)|F|\beta - exp_G(F) rel(M)}{rel(F) - |F|\beta + rel(M)} &= \frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \\
exp_G(M) rel(F) - exp_G(M)|F|\beta - exp_G(F) rel(M) &= (rel(F) - |F|\beta + rel(M) \\
&\quad \times \left(\frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \right) \\
exp_G(M) rel(F) - exp_G(F) rel(M) - (rel(F) + rel(M)) \left(\frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \right) &= |F|\beta \\
&\quad \times \left(exp_G(M) - \left(\frac{(\mathbf{v}l - \mathbf{v}(j)}{rel(M_i)} \right. \right. \\
\frac{exp_G(M) rel(F) - exp_G(F) rel(M) - (rel(F) + rel(M)) \left(\frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \right)}{|F| \left(exp_G(M) - \left(\frac{(\mathbf{v}l - \mathbf{v}(j))}{rel(M_i)} \sum_{j=1}^{k} rel(M_i) \right) \right)} &= \beta
\end{aligned}
\tag{B.3}
$$

We additionally do a similar process for when $x_4 = 1$,

$$
\begin{aligned}
x_4 &= \frac{exp_F(M_i) - \mathbf{v}(j)}{\mathbf{v}(k) - \mathbf{v}(j)} \\
1 &= \frac{exp_F(M_i) - \mathbf{v}(j)}{\mathbf{v}(k) - \mathbf{v}(j)} \\
\mathbf{v}(k) - \mathbf{v}(j) &= exp_F(M_i) - \mathbf{v}(j) \\
exp_F(M_i) &= \mathbf{v}(k) \\
exp_G(M_i) - \eta \frac{rel(M_i)}{\sum_{j=1}^{k} rel(M_i)} &= \mathbf{v}(k) \\
\mathbf{v}(k) - \eta \frac{rel(M_i)}{\sum_{j=1}^{k} rel(M_i)} &= \mathbf{v}(k) \\
\eta \frac{rel(M_i)}{\sum_{j=1}^{k} rel(M_i)} &= 0 \\
\eta &= 0
\end{aligned}
\tag{B.4}
$$

This implies that $exp_G(F) = exp_F(F)$ which would mean that the greedy solution is in fact fair. This occurs when

$$
\begin{aligned}
exp_G(M) &= \frac{tot\_exp \times rel(M)}{rel(M) + rel(F) - |F|\beta} \\
\beta &= -\frac{tot\_exp \times rel(M)}{exp_G(M)|F|} + \frac{rel(M) + rel(F)}{|F|}
\end{aligned}
\tag{B.5}
$$

Thus we can verify that β is between the values computed in Equation B.3 and B.3. This is a check after greedily ranking to verify whether their exists a feasible solution under the pair-wise construction described in Chapter 3.

APPENDIX C

EXTENDED EXPERIMENTAL RESULTS

Table C.1: Unfairness with respect to Disparate treatment for different minority group proportions.

α	No constraint	Disparate treatment	Demographic parity	Disparate impact
0.1	.41	**.06**	.41	.42
0.2	.21	**.03**	.18	.18
0.3	.15	**.02**	.09	.08
0.4	.12	.08	.04	**.02**
0.5	.11	.03	**2e-4**	.02
0.6	.10	.05	**.04**	.06
0.7	.10	**.03**	.10	.12
0.8	.11	**.03**	.19	.22
0.9	.10	**.06**	.42	.46

Table C.2: Unfairness with respect to Disparate treatment for differing additive biases.

β	No constraint	Disparate treatment	Demographic parity	Disparate impact
0	0	0	0	0
0.1	0.09	0.05	7e-4	8.8e-3
0.2	.11	0.06	6e-4	0.02
0.3	.12	.09	6e-5	0.03
0.4	.13	.13	8.2e-5	0.05

Table C.3: Unfairness with respect to Disparate treatment for differing multiplicative biases.

β	No constraint	Disparate treatment	Demographic parity	Disparate impact
1	0	0	0	0
0.9	0.08	0.01	9.8e-4	8e-3
0.8	.11	0.03	2e-4	0.02
0.7	.12	0.08	2e-5	0.03
0.6	.13	0.09	1.9e-5	0.04
0.5	.13	0.09	5.1e-5	0.06
0.4	.14	0.10	2.8e-5	0.07

Table C.4: Unfairness with respect to Disparate treatment for differing exposure functions.

v	No constraint	Disparate treatment	Demographic parity	Disparate impact
$1/\log_2(1+x)$	0.08	0.08	2e-4	0.06
$1/x^{0.5}$	0.14	0.10	1.4e-4	0.06
$1/x$	.11	0.11	0.03	0.02
$1/x^{1.5}$	.12	0.08	2e-5	0.03
$1/x^2$	.06	0.05	2e-4	7e-3
$1/e^x$	.02	0.02	0.02	0.02

Table C.5: Unfairness with respect to Disparate treatment for differing relevance functions.

rel	No constraint	Disparate treatment	Demographic parity	Disparate impact
$(1/n, 1)$	0.11	0.03	2e-4	0.02
$(0, 1)$	0.11	0.03	1.7e-5	0.02
$(0.5, 1)$	.09	0.02	1.3e-4	0.01
$(0, 0.5)$	.21	0.06	3.4e-5	0.04
$1/x^{0.5}$	0.11	0.05	2.9e-4	0.02
$1/x$	.20	0.13	7e-3	0.19
$1/x^{1.5}$	.35	0.25	2e-3	0.35
$1/x^2$	0.47	0.35	6e-4	0.46
$\mathcal{U}(0, 1)$	.10	0.07	0.01	0.03

Table C.6: Disparate impact when male and female relevance distributions are different. The male or female distribution is shifted downward such that males tend to be more relevant (and vice versa).

Ranking intervention	Higher male relevance	Higher female relevance
No constraint	0.11	0.11
Disparate treatment	0.06	0.06
Demographic parity	0.004	0.004
Disparate impact	.02	0.01

REFERENCES

[1] Ai recruiting amp; job matching platform.

[2] Fact sheet: President obama announces new commitments from investors, companies, universities, and cities to advance inclusive entrepreneurship at first-ever white house demo day.

[3] Managing bias at facebook.

[4] Why girls in india are still missing out on the education they need, Mar 2013.

[5] 2017.

[6] Race in the workplace: Mcgregor smith review, Oct 2018.

[7] Op-ed: Implicit bias puts lives in jeopardy. can mandatory training reduce the risk?, Jul 2019.

[8] Hirevue assessments and preventing algorithmic bias, Jul 2021.

[9] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[10] Abolfazl Asudeh, H. V. Jagadish, Julia Stoyanovich, and Gautam Das. Designing fair ranking schemes, 2017.

[11] Marc Bendick and Ana Nunes. Developing the research basis for controlling bias in hiring. *Journal of Social Issues*, 68:238–262, 01 2011.

[12] Linda Bergh, Eddie Denessen, Lisette Hornstra, Rinus Voeten, and Rob Holland. The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethic achievement gap. *American Educational Research Journal - AMER EDUC RES J*, 47:497–527, 05 2010.

[13] Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. Working Paper 9873, National Bureau of Economic Research, July 2003.

[14] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. Equity of attention. In *The 41st International ACM SIGIR Conference on Research &amp Development in Information Retrieval*. ACM, jun 2018.

[15] Miranda Bogen and Aaron Rieke. Help wanted: an examination of hiring algorithms, equity, and bias. 2018.

[16] Iris Bohnet, Max Bazerman, and Alexandra Geen. When performance trumps gender bias: Joint versus separate evaluation. *SSRN Electronic Journal*, 62, 03 2012.

[17] Amanda Bower, Hamid Eftekhari, Mikhail Yurochkin, and Yuekai Sun. Individually fair ranking, 2021.

[18] Pedro G. Campos, Fernando Díez, and Iván Cantador. Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1–2):67–119, feb 2014.

[19] Carlos Castillo. Fairness and transparency in ranking. *SIGIR Forum*, 52(2):64–71, jan 2019.

[20] Marilyn Cavicchia. How to fight implicit bias? with conscious thought, diversity expert tells nabe, 2015.

[21] L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi. Interventions for ranking in the presence of implicit bias, 2020.

[22] L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. Ranking with fairness constraints, 2017.

[23] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. *Click Models for Web Search*. Morgan & Claypool, 2015.

[24] Brian W Collins. Tackling unconscious bias in hiring practices: The plight of the rooney rule, Sep 2018.

[25] Sam Corbett-Davies and Sharad Goel. The measure and mismeasure of fairness: A critical review of fair machine learning, 2018.

[26] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008*

International Conference on Web Search and Data Mining, WSDM '08, page 87–94, New York, NY, USA, 2008. Association for Computing Machinery.

[27] Jeffrey Dastin. Amazon scraps secret ai recruiting tool that showed bias against women, Oct 2018.

[28] Fernando Diaz, Bhaskar Mitra, Michael D. Ekstrand, Asia J. Biega, and Ben Carterette. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM International Conference on Information amp; Knowledge Management*, CIKM '20, page 275–284, New York, NY, USA, 2020. Association for Computing Machinery.

[29] Cynthia DuBois. The impact of "soft" affirmative action policies on minority hiring in executive leadership: The case of the nfl's rooney rule. *American Law and Economics Review*, 18:ahv019, 09 2015.

[30] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Rich Zemel. Fairness through awareness, 2011.

[31] Michael D. Ekstrand, Robin Burke, and Fernando Diaz. Fairness and discrimination in retrieval and recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, page 1403–1404, New York, NY, USA, 2019. Association for Computing Machinery.

[32] Vitalii Emelianov, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau. On fair selection in the presence of implicit variance. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, page 649–675, New York, NY, USA, 2020. Association for Computing Machinery.

[33] Yuri Faenza, Swati Gupta, and Xuan Zhang. Reducing the feeder effect in public school admissions: A bias-aware analysis for targeted interventions, 2020.

[34] Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. Bridging machine learning and mechanism design towards algorithmic fairness, 2020.

[35] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. Fairness-aware ranking in search &amp recommendation systems with application to LinkedIn talent search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &amp Data Mining*. ACM, jul 2019.

[36] Alexander Green, Dana Carney, Daniel Pallin, Long Ngo, Kristal Raymond, Lisa Iezzoni, and Mahzarin Banaji. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine*, 22:1231–8, 09 2007.

[37] Anthony Greenwald and Linda Krieger. Implicit bias: Scientific foundations. *California Law Review*, 94:945, 07 2006.

[38] Anthony G Greenwald and Mahzarin R. Banaji. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102 1:4–27, 1995.

[39] Kalervo Järvelin and Jaana Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, page 41–48, New York, NY, USA, 2000. Association for Computing Machinery.

[40] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, oct 2002.

[41] Jeremy Kahn. Why hirevue will no longer assess job seekers' facial expressions, Jan 2021.

[42] Jon Kleinberg and Manish Raghavan. Selection problems in the presence of implicit bias, 2018.

[43] Julian Lamont and Christi Favor. Distributive Justice. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2017 edition, 2017.

[44] Tie-Yan Liu. Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, 3(3):225–331, mar 2009.

[45] Karen S. Lyness and Madeline E. Heilman. When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4):777–785, July 2006.

[46] Kay Manning. As starbucks gears up for training, here's why 'implicit bias' can be good, bad or very bad, Dec 2018.

[47] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning, 2019.

[48] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, 8(1):141–163, 2021.

[49] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. Controlling fairness and bias in dynamic learning-to-rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, jul 2020.

[50] Mike Noon. Pointless diversity training: Unconscious bias, new racism and agency. *Work, Employment and Society*, 32:198 – 209, 2018.

[51] Executive Office of the President, Cecilia Munoz, Megan Smith, and D.J. Patil.

[52] Christina Passariello. Tech firms borrow football play to increase hiring of women, Sep 2016.

[53] B. Keith Payne, Heidi A. Vuletich, and Jazmin L. Brown-Iannuzzi. Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences*, 116(24):11693–11698, 2019.

[54] Evaggelia Pitoura, Panayiotis Tsaparas, Giorgos Flouris, Irini Fundulaki, Panagiotis Papadakos, Serge Abiteboul, and Gerhard Weikum. On measuring bias in online information, 2017.

[55] Pearl Pu, Li Chen, and Rong Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, page 157–164, New York, NY, USA, 2011. Association for Computing Machinery.

[56] Marco Tulio Ribeiro, Anísio Mendes Lacerda, Edleno Silva de Moura, Itamar Hata, Adriano Veloso, and Nivio Ziviani. Multi-objective pareto-efficient approaches for recommend er systems. 2013.

[57] S. E. Robertson. *The Probability Ranking Principle in IR*, page 281–286. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997.

[58] Laurie Rudman. Sources of implicit attitudes. *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI*, 13:79–82, 04 2004.

[59] Melody Sadler, Joshua Correll, Bernadette Park, and Charles M Judd. The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues*, 68:286–313, 2012.

[60] Deepa Seetharaman. Facebook is testing the 'rooney rule' approach to hiring, Jun 2015.

[61] Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '18, page 2219–2228, New York, NY, USA, 2018. Association for Computing Machinery.

[62] Yi Su, Magd Bayoumi, and Thorsten Joachims. Optimizing rankings for recommendation in matching markets, 2021.

[63] Tom Sühr, Asia J. Biega, Meike Zehlike, Krishna P. Gummadi, and Abhijnan Chakraborty. Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '19, page 3082–3092, New York, NY, USA, 2019. Association for Computing Machinery.

[64] Christine R Sunstein. The law of implicit bias. *California law review.*, 94(4), 2006.

[65] Özge Sürer, Robin Burke, and Edward C. Malthouse. Multistakeholder recommendation with provider constraints. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 54–62, New York, NY, USA, 2018. Association for Computing Machinery.

[66] Tekchand, Sep 2021.

[67] Eric Luis Uhlmann and Geoffrey L. Cohen. Constructed criteria. *Psychological Science*, 16:474 – 480, 2005.

[68] Ellen M. Voorhees. Variations in relevance judgments and the measurement of retrieval effectiveness. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*,

SIGIR '98, page 315–323, New York, NY, USA, 1998. Association for Computing Machinery.

[69] Ellen M. Voorhees and Dawn M. Tice. The TREC-8 question answering track. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece, May 2000. European Language Resources Association (ELRA).

[70] Joseph Walker. Meet the new boss: Big data, Sep 2012.

[71] Gregory Walton. Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological science*, 20:1132–9, 08 2009.

[72] Christine Wenners and Agnes Wold. Nepotism and sexism in peer-review. *Nature*, 7:46–52, 05 1997.

[73] Joan C. Williams. Double jeopardy? an empirical study with implications for the debates over implicit bias and intersectionality,. 2014.

[74] Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, and Frida Polli. Building and auditing fair algorithms: A case study in candidate screening. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 666–677, New York, NY, USA, 2021. Association for Computing Machinery.

[75] Kathleen Woodhouse. Council post: Implicit bias – is it really?, Dec 2017.

[76] Lin Xiao, Zhang Min, Zhang Yongfeng, Gu Zhaoquan, Liu Yiqun, and Ma Shaoping. Fairness-aware group recommendation with pareto-efficiency. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, page 107–115, New York, NY, USA, 2017. Association for Computing Machinery.

[77] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Gerome Miklau. A nutritional label for rankings. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, may 2018.

[78] Sirui Yao and Bert Huang. Beyond parity: Fairness objectives for collaborative filtering. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus,

S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[79] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa*ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 1569–1578, New York, NY, USA, 2017. Association for Computing Machinery.

[80] Meike Zehlike, Tom Sühr, Ricardo Baeza-Yates, Francesco Bonchi, Carlos Castillo, and Sara Hajian. Fair top-¡i¿k¡/i¿ ranking with multiple protected groups. *Inf. Process. Manage.*, 59(1), jan 2022.

[81] Meike Zehlike, Tom Sühr, Carlos Castillo, and Ivan Kitanovski. FairSearch: A tool for fairness in ranked search results. In *Companion Proceedings of the Web Conference 2020*. ACM, apr 2020.

[82] Colin A. Zestcott, Irene V Blair, and Jeff Stone. Examining the presence, consequences, and reduction of implicit bias in health care: A narrative review. *Group Processes & Intergroup Relations*, 19:528 – 542, 2016.