Artificial Intelligence 302 (2022) 103609

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

On fair selection in the presence of implicit and differential variance ☆, ☆☆

Vitalii Emelianov [a,*], Nicolas Gast [a], Krishna P. Gummadi [b], Patrick Loiseau [a]

[a] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, Grenoble, France
[b] Max Planck Institute for Software Systems, Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 13 August 2020
Received in revised form 13 October 2021
Accepted 16 October 2021
Available online 20 October 2021

Keywords:
Selection problem
Fairness
Differential variance

ABSTRACT

Discrimination in selection problems such as hiring or college admission is often explained by implicit bias from the decision maker against disadvantaged demographic groups. In this paper, we consider a model where the decision maker receives a *noisy* estimate of each candidate's quality, whose variance depends on the candidate's group—we argue that such *differential variance* is a key feature of many selection problems. We analyze two notable settings: in the first, the noise variances are unknown to the decision maker who simply picks the candidates with the highest estimated quality independently of their group; in the second, the variances are known and the decision maker picks candidates having the highest expected quality given the noisy estimate. We show that both baseline decision makers yield discrimination, although in opposite directions: the first leads to underrepresentation of the low-variance group while the second leads to underrepresentation of the high-variance group. We study the effect on the selection utility of imposing a fairness mechanism that we term the γ-rule (it is an extension of the classical four-fifths rule and it also includes demographic parity). In the first setting (with unknown variances), we prove that under mild conditions, imposing the γ-rule increases the selection utility—here there is no trade-off between fairness and utility. In the second setting (with known variances), imposing the γ-rule decreases the utility but we prove a bound on the utility loss due to the fairness mechanism.

1. Introduction

Discrimination in selection and the role of implicit bias Many selection problems such as hiring or college admission are subject to discrimination [2], where the outcomes for certain individuals are negatively correlated with their membership in salient demographic groups defined by attributes like gender, race, ethnicity, sexual orientation or religion. Over the past two decades, implicit bias—that is an unconscious negative perception of the members of certain demographic groups—has been put forward as a key factor in explaining this discrimination [3]. While human decision makers are naturally susceptible to

☆ This paper is an extended and revised version of our paper "On Fair Selection in the Presence of Implicit Variance" that appeared in the proceedings of EC'20 [1].

☆☆ This paper is a participant in the 2020 ACM Conference on Economics and Computation (EC) Forward-to-Journal Program.

* Corresponding author.
E-mail addresses: vitalii.emelianov@inria.fr (V. Emelianov), nicolas.gast@inria.fr (N. Gast), gummadi@mpi-sws.org (K.P. Gummadi), patrick.loiseau@inria.fr (P. Loiseau).

https://doi.org/10.1016/j.artint.2021.103609
0004-3702/

implicit bias when assessing candidates, algorithmic decision makers are also vulnerable to implicit biases when the data used to train them or to make decisions was generated by humans.

To mitigate the effects of discrimination on candidates from underrepresented groups, various fairness mechanisms[1] are adopted in many domains, either by law or through softer guidelines. For instance, the *Rooney rule* [4] requires that, when hiring for a given position, at least one candidate from the underrepresented group be interviewed. The Rooney rule was initially introduced for hiring American football coaches, but it is increasingly being adopted by many other businesses in particular for hiring top executives [5,6]. Another widely used fairness mechanism is the so-called $^4/_5$*-rule* [7], which requires that the selection rate for the underrepresented group be at least 80% of that for the overrepresented group (otherwise one says that there is adverse impact). This rule is part of the "Uniform Guidelines On Employee Selection Procedures".[2] A stricter version of the $^4/_5$-rule is the so-called *demographic parity* constraint, which requires the selection rates for all groups to be equal. An overview of these and other fairness mechanisms can be found in [7].

Fairness mechanisms, however, have been the subject of frequent debates. On the one hand, they are believed to promote the inclusion of deserving candidates from underrepresented groups who would have otherwise been excluded in particular due to implicit bias. On the other hand, they are viewed as requiring consideration of candidates from underrepresented groups at the expense of candidates from overrepresented groups, which may potentially decrease the overall utility of the selection process, i.e., the overall quality of selected candidates.

Formal analysis of fairness mechanisms in the presence of implicit bias Perhaps surprisingly, the mathematical analysis of the effect of fairness mechanisms on utility in the context of selection problems was initiated only recently by Kleinberg and Raghavan [8] (see also an extension to ranking problems in [9]). The authors of [8] assume that each candidate i has a true latent quality W_i that comes from a group-independent distribution. They model implicit bias by assuming that the decision maker sees an estimate of the quality $\widehat{W}_i = W_i$ for candidates from the well-represented group and $\widehat{W}_i = W_i/\beta$ for candidates from the underrepresented group, where $\beta > 1$ measures the amount of implicit bias. The factor β is unknown (as it is implicit bias) and the decision maker selects candidates by ranking them according to $\widehat{W}_i$. Then Kleinberg and Raghavan [8] show that, under a well-defined condition (that roughly qualifies scenarios where the bias is large), the Rooney rule improves in expectation the utility of the selection (measured as the sum of true qualities of candidates selected for interview). This result contradicts conventional wisdom that fairness considerations in a selection process are at odds with the utility of the selection process. Rather, it formalizes the intuition that, in the presence of strong implicit bias (which makes it hard to compare candidates across groups), considering the best candidates across a diverse set of groups not only improves fairness but it also has a positive effect on utility.

The phenomenon of differential variance and its role in discrimination In this paper, we identify and analyze a fundamentally different source of discrimination in selection problems than implicit bias. Even in the absence of implicit bias in a decision maker's estimate of candidates' quality, the estimates may differ between the different groups in their *variance*—that is, the decision maker's ability to precisely estimate a candidate's quality may depend on the candidate's group. There are at least two main reasons for group-dependent variances in practice. The first arises from *candidates*: different groups of candidates may exhibit different variability when their quality is estimated through a given test. For instance, students of different genders have been observed to show different variability on certain test scores [10,11].[3] The second arises from the *decision makers*: decision makers might have different levels of experience (or different amounts of data in case of algorithmic decision making) judging candidates from different groups and consequently, their ability to precisely assess the quality of candidates belonging to different groups might be different. For instance, when hiring top executives, one may have less experience in evaluating the performance of female candidates because there have been fewer women in those positions in the past (in France for instance, there was only one woman CEO amongst the top-40 companies in 2016-2020 [12]). The quality estimate's variance might also change from one decision maker to another. For example, in college admissions, recruiters might be able to judge candidates from schools in their own country more accurately than those from international schools.

We refer to the above phenomenon as *differential variance* as the variance of the quality estimate is group-dependent. We posit that differential variance is an omnipresent and fundamental feature affecting selection problems (including in algorithmic decision making). Indeed, having different variances for the different groups is mostly inevitable and hardly fixable. In this paper, we model the differential variance phenomenon by assuming that the decision maker sees of an

[1] These mechanisms are sometimes termed "positive discrimination" (e.g., in Germany, France, China, or India) or "affirmative actions" (in the USA), often referring to their justification as corrective measures against discrimination suffered in the past by disadvantaged groups. In our work, we analyze the effect of these mechanisms in a particular setting of selection problems (with differential variance) independently of their motivation, hence we use the more neutral term "fairness mechanisms."

[2] A set of guidelines jointly adopted by the Equal Employment Opportunity Commission, the Civil Service Commission, the Department of Labor, and the Department of Justice in 1978.

[3] Note that, while this indicates that observed performance is more variable for one group than the other, it is impossible to tell whether this comes from different underlying distributions or from different measurement variances—or (more likely) from both. In fact, the general "variability hypothesis" is subject to a number of controversies. Nevertheless, this indicates potential differences between groups in the variance of the observed signals and our model can flexibly incorporate both different prior distributions and different measurement noises.

estimate of the quality of a candidate $\widehat{W}_i$ that is equal to the candidate's true latent quality W_i (possibly with an additional bias term) plus an additive noise[4] whose variance depends on the group of the candidate.

We distinguish between two notable settings. In the first setting, the noise variance is assumed to be unknown to the decision maker—we then call it *implicit variance*. In this case, a natural baseline decision maker is the *group-oblivious* algorithm[5] that simply selects the candidates with the highest estimated quality $\widehat{W}_i$, irrespective of their group. The group-oblivious selection algorithm can represent not only a decision maker unaware of the implicit variance in their estimates, but also a decision maker determined to not use group information—as it may be the case for instance in college admission based on standardized tests. In the second setting, the noise variance is known to the decision maker. In this case, a natural baseline is the *Bayesian-optimal* algorithm: this decision maker can use the group information as well as the knowledge of the distributions of latent quality and noise to select the candidates that maximize the expected quality given the noisy estimate.

As a first cornerstone, our analysis shows that, in the presence of differential variance, both the group-oblivious and the Bayesian-optimal algorithms lead to discrimination (although in opposite directions, see the overview of our results below). A natural way to address this representation inequality is to adopt fairness mechanisms proposed to address discrimination in selection such as the ones discussed above; but this poses the same question that was investigated by Kleinberg and Raghavan [8] in the case of implicit bias: *what is the effect of fairness mechanisms on the quality of a selection in the presence of differential variance?*

Our model and overview of our results To answer this question, we propose a simple model with two groups of candidates A and B: for each candidate i, the decision maker receives a noisy (and possibly biased) quality estimate $\widehat{W}_i = W_i - \beta_{G_i} + \sigma_{G_i}\varepsilon_i$, where G_i is the group to which the candidate belongs and ε_i is a standard normal random variable. The estimator has an additive bias β_{G_i} and a variance $\sigma^2_{G_i}$ that depend on the candidate's group. We assume that the true quality W_i comes from a distribution—assumed normal in our analytical results—that may be group-dependent. The decision maker then selects a fraction α (called selection budget) of the candidates.

The key feature of our model is the variance $\sigma^2_{G_i}$ that depends on the candidate's group—to model differential variance. In its general version, we also allow a bias and a latent quality distribution that depend on the candidate's group. Using this general model, we first show (Section 3.1) that both the group-oblivious and the Bayesian-optimal selection algorithms systematically lead to underrepresentation—i.e., lower selection rate—of one of the groups of candidates. Specifically, we identify a cutoff budget such that the group-oblivious selection algorithm leads to underrepresentation of the low-variance group for any budget α smaller than the cutoff (the most common case) and underrepresentation of the high-variance group for any budget α larger than the cutoff. Conversely (and for a different cutoff), the Bayesian-optimal algorithm leads to underrepresentation of the high-variance group for low budgets and of the low-variance group for high budgets. In fact, we show (Section 4.1) that this is true even in the absence of bias and with group-independent latent quality distributions—that is, if the noise variance is the only thing that depends on the candidate's group. In this particular case, the cutoff budget for both algorithms is $\alpha = 1/2$.

Then we investigate how the utility of the group-oblivious and the Bayesian-optimal baselines are affected when imposing a fairness mechanism. Specifically, we study a generalization of the 4/5-rule that we call *γ-rule*, which imposes that the selection rate for a given group is at least γ times that of the other group for some parameter $\gamma \in [0, 1]$. This includes both the 4/5-rule ($\gamma = 0.8$) and demographic parity ($\gamma = 1$) as special cases. In the general model, we identify conditions under which the γ-rule never decreases the utility of the group-oblivious algorithm (Section 3.2)—that is, there is no trade-off between fairness and selection quality for this baseline. The utility even strictly increases for γ close enough to one, including for demographic parity. Interestingly, in the special case without bias and with group-independent latent quality distributions—that is, with only implicit differential variance—, this result *always* holds for any parameters (Section 4.1). Compared to the Bayesian-optimal baseline, the γ-rule cannot increase the utility (since Bayesian-optimal is already optimal given the available information). We prove, however, a bound on the ratio of the utility of the Bayesian-optimal algorithm with and without the γ-rule imposed, which limits the decrease of utility due to imposing a fairness mechanism in this setting. Our bound is valid in the general model (Section 3.3) but takes a particularly simple form in the special case without bias and with group-independent latent quality distributions (Section 4.1).

A typical case of differential variance is when the decision maker has more uncertainty about one group, due to lack of statistical confidence (e.g., in hiring). In such a case, the high-variance group naturally corresponds to the minority group. The group-oblivious algorithm would then overrepresent the minority group (for small selection budgets), and the fairness mechanism would lead to selecting fewer of the minority group—which is counter-intuitive. We stress, however, that those are typically cases where the relevant baseline is the Bayesian-optimal algorithm, which behaves very differently. Through the Bayesian posterior quality computation, this baseline would disregard candidates for which the observed quality estimate is uninformative, that is the high-variance group. As mentioned above, we indeed find that the Bayesian-optimal algorithm underrepresents the high-variance group (i.e., the minority), and that the fairness mechanism increases the proportion of

[4] This noise may be a property of the decision maker getting a noisy perception of the candidate's quality or a property of the candidate (i.e., the variability in the candidate's performance).

[5] Throughout the paper, we use the term 'algorithm' for the selection procedure, irrespective of whether it is algorithmic decision making or not.

selected high-variance candidates—which is coherent with intuition for that case. The group-oblivious baseline is meaningful in other scenarios, typically when the decision maker is not allowed to use the group information (e.g., in college admission based on standardized tests). In such cases, the high-variance group may not be a minority group (and our model does not require that it is).

At a high-level, our results indicate that, with differential variance, the two decision makers (group-oblivious and Bayesian-optimal) lead to nearly opposite outcomes in terms of discrimination; and that the effect of imposing fairness mechanisms can be very different for both. These results imply that a policy-maker considering fairness mechanisms for a given problem should first evaluate to which decision maker the selection rule corresponds, and then choose whether or not to recommend the γ-rule based on it. Note that this should be fairly easy to distinguish between the two in practice, since one conditions on group identity while the other does not.

Organization of the paper The rest of the paper is organized as follows. We present the model in Section 2. We give all the results in the most general case in Section 3. Due to their generality, those results are sometimes complex. In Section 4, we analyze three notable cases for which the results are easier to interpret: the case without bias and with group-independent latent quality distributions (Section 4.1), the case with bias but with group-independent latent quality distributions (Section 4.2), and the case without bias but with group-dependent latent quality distributions (Section 4.3). Through numerical simulations in Section 5, we extend our analytical results, in particular to cases where the latent quality distribution does not follow a normal law. We conclude in Section 6.

Related work There is an abundant literature on fairness in machine learning, in particular on classification, that tackles the question of how to learn a classifier while enforcing some fairness notion in the outcome [13–20]. In this literature, fairness is usually seen as a constraint that reduces the classifier's accuracy and the fairness-accuracy tradeoff is analyzed. In contrast, in our work, we examine selection problems in which fairness can improve utility. Selection also differs from classification by the presence of selection budgets (i.e., maximal number of class-1 predictions), which changes the problem significantly.

The problem of selection is considered in [8] under the presence of implicit bias [3]. In their work, the authors study the Rooney rule [4] as a fairness mechanism and show that under certain conditions, it improves the quality of selection. An extension of the Rooney rule is studied under a similar model in [9], where the authors investigate the ranking problem (of which the selection problem can be seen as a special case) also in the presence of implicit bias and obtain similar results. In both papers, simple mathematical results expressing conditions under which the Rooney rule improves utility are obtained in the limit regime where the number of candidates is very large; we use the same limit regime in our work. In contrast to those papers that only consider bias, we introduce in addition the notion of differential variance to capture the difference in precision of the quality estimate for different groups. We also consider an additive bias rather than a multiplicative one as it makes more sense for normally distributed qualities. Although our model incorporates both an additive bias and differential variance (in Section 3), we purposely restrict it in Section 4 to the simplest possible form of differential variance so as to show its effect on the selection problem independently of bias. In our work, we also consider the $^4/_5$-rule [7] (or rather an extension of it that we call the γ-rule and that includes demographic parity) rather than the Rooney rule. The main difference between the two is that the $^4/_5$-rule imposes a constraint on the *fraction* of selected candidates from the underrepresented group whereas the Rooney rule or its extension in [9] imposes a constraint on the *number* of selected candidates from the underrepresented group.

Implicit bias, or simply bias (possibly from an algorithm trained on biased data) in the evaluation of candidates quality is certainly a primary factor of discrimination; but it is also one that may reasonably be fixable through the use of algorithms combined with appropriate debiasing techniques and ground truth data [21] (e.g., by learning fair representations of data [22,23]). The effects of bias can be also fixed by introducing some fairness constraints on learned prediction models. For example, in [24], the binary classification problem in the presence of label bias is studied and it is shown that adding a demographic parity constraint to an empirical risk minimization problem can lead to better generalization. Similarly, in [25], the authors study the effects of label bias on binary classification and they show that equal opportunity fairness criterion (that ensures that true positives are equal across the groups) can reduce the bias in prediction for most of the reasonable cases, as well as improve the accuracy of classification. In [26], the authors quantify a fairness-accuracy trade-off using an information theoretic approach and, in addition, they show that for the majority of traditional fairness criteria (like equal opportunity and demographic parity) there exists an ideal data distribution for which fairness and Bayesian optimality are in accordance.

The notion of differential variance first appeared (with different terminology) in the seminal work of Phelps [27] to explain racial inequality in wages. There, a Bayesian decision maker observes noisy signals of productivity of each worker. Productivities are assumed to be drawn from a common distribution while precisions of estimation differ across races. Phelps shows that a Bayesian decision maker that assigns wages equal to the expected productivity of a worker leads to inequality of wages: in the region of high values of signals the low-precision workers receive lower wages. Our model is similar that of Phelps, with additional bias and possibly group-dependent prior distributions. We also study cases where the variance is implicit—hence the decision maker cannot use Bayes' rule to estimate expected quality given noisy estimates—, and focus on utility for our main results.

This paper is an extended version of our paper "On Fair Selection in the Presence of Implicit Variance" [1]. We extend it by considering the general model with bias and group-dependent latent quality distributions, and by analyzing in parallel the two baselines of the group-oblivious and Bayesian-optimal algorithms (whereas [1] only looks at the group-oblivious baseline, that is at implicit variance). On the other hand, we do not include the results on two-stage decisions makers for conciseness. Following [1], Garg et al. [28] studied a similar model (using the term differential variance that we also adopt here). The authors propose a model of school admission with students of two groups: advantaged and disadvantaged. Each student has an intrinsic quality which is not observable to schools: only noisy signals of the quality are available. The advantaged and disadvantaged students differ in the level of precision of their signals, and can also differ in their ability to access the tests. The authors consider the case of a Bayesian school of limited capacity. They study how different policies made by the school (group-aware and group-unaware) affect the diversity level, individual fairness and overall merit of admitted students. The authors also study how dropping test scores and different abilities to access tests affects the above characteristics.

Fairness mechanisms have been a subject of a number of studies in the economic literature, in particular from empirical data. In [29], the authors study whether affirmative actions can remove stereotypes about a particular population. In [30], an empirical evaluation of the influence of affirmative actions in recruiting is performed and it is shown that it can bring quality together with equality. Our work complements those studies through a theoretical model that leads to analytical results on the effect of fairness mechanisms in the presence of differential variance.

2. Model and selection algorithms

2.1. The model of selection with differential variance

We consider the following scenario. A decision maker is given n candidates, out of which a subset of size $m = \alpha n$ is selected, $\alpha \in (0, 1)$. We assume that the set of candidates can be partitioned in two groups: group A and group B. There are n_A candidates from group A and $n_B = n - n_A$ candidates from group B. We refer to them as A-candidates and B-candidates.

Each candidate $i \in \{1, \dots, n\}$ is endowed with a true latent quality W_i. We assume that the qualities W_i are drawn *i.i.d.* from an underlying probability distribution that can be group-dependent.[6] For our analytical results, we assume that this distribution is a normal distribution of mean μ_{G_i} and variance $\eta^2_{G_i}$, where $G_i \in \{A, B\}$ is the group of candidate i.

The goal of the decision maker is to maximize the expected quality of the selected candidates: $\mathrm{E}\left[\sum_{i \in \text{selection}} W_i\right]$. When making the selection decision, the decision maker has access to a (possibly biased) noisy estimator of the true quality. We denote the estimator of the quality of candidate i by $\widehat{W}_i$. We assume that the bias and the variance of the estimator may depend on the group: for a candidate i that belongs to group $G_i \in \{A, B\}$, its estimated quality is

$$\widehat{W}_i = \begin{cases} W_i - \beta_A + \sigma_A \cdot \varepsilon_i & \text{if } i \text{ is an } A\text{-candidate,} \\ W_i - \beta_B + \sigma_B \cdot \varepsilon_i & \text{if } i \text{ is a } B\text{-candidate,} \end{cases} \tag{1}$$

where ε_i is a centered random variable from $\mathcal{N}(0, 1)$—the standard normal distribution, of mean 0 and variance 1. The variables ε_i are assumed independent and identically distributed. Note that we model the bias as an additive parameter in contrast to the multiplicative parameter in [8]. This is more suitable for our model of qualities as normally distributed random variables, which can be negative (a multiplicative bias on a negative quality would turn into a positive effect, which is not meaningful).

We denote by $\hat{\sigma}^2_{G_i} = \sigma^2_{G_i} + \eta^2_{G_i}$ the variance of the estimate $\widehat{W}_i$. Without loss of generality, we assume that the estimates' variance is larger for A-candidates than for B-candidates, that is $\hat{\sigma}^2_A > \hat{\sigma}^2_B$. We note that none of our results require that A is also the minority group, i.e., that $n_A < n_B$. It is possible to think of scenarios where the minority group has lower variance in cases where the difference in variances arises from the candidates. In the example of students' tests scores (see Section 1), for instance, one could potentially observe that males have greater variability in topics in which they are in majority. If the difference in variances arises from the decision maker and has a statistical nature, the minority group (for past selections) will have higher variance due to less data points to build the estimator.

Throughout the paper, we refer to this difference in variance as *differential variance* because we assume that the variance of the estimators differs across groups. Fig. 1 illustrates the resulting distribution of quality estimates for groups A and B for different distributions of the true latent quality (by abuse of notation, we denote by $\widehat{W}_A$ a variable that has the same distribution as $W_i + \sigma_A \varepsilon_i$ and similarly for B).

2.2. Selection algorithms

Candidates are selected in a one-stage process: for each candidate i, the decision maker observes the quality estimate $\widehat{W}_i$ as well as its group $G_i \in \{A, B\}$. The decision maker then selects m candidates out of those n. The goal of the decision maker

[6] We present here the model in its most general form. We will analyze special cases, in particular when the quality distribution is group-independent, in Section 4.

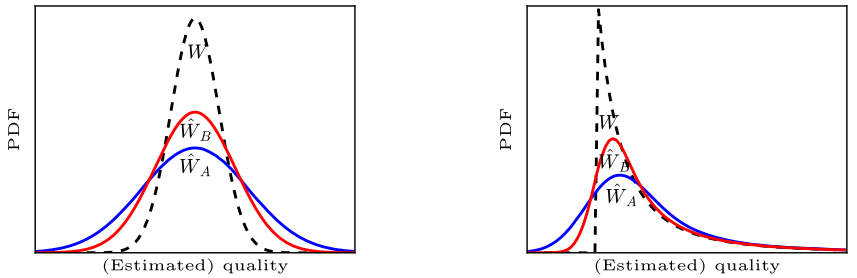


Fig. 1. Probability density function of the true latent quality W and the estimated quality $\widehat{W}$. To the purpose of illustration, the underlying distribution is assumed group-independent and the estimation is unbiased.

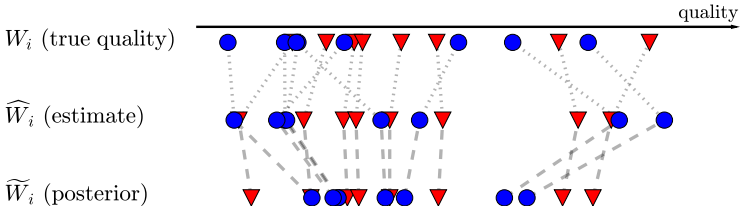


Fig. 2. Illustration of the baseline selection algorithms. Here, there are $n_{blue} = 8$ blue and $n_{red} = 8$ red candidates, and the decision maker wants to select $m = 2$ candidates. The quality is group-independent and there is no bias. The estimator variance is three times higher for the blue candidates. Here, the group-oblivious algorithm would select the 2 blue candidates. Yet, because blue candidates have higher variance, the Bayesian-optimal algorithm would select 2 red. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

is to maximize the expected quality of the selected m candidates. In this paper, we distinguish and study the following two baseline selection algorithms. Each baseline is a natural selection algorithm in a situation when the decision maker knows the parameters of the model (μ_{G_i}, $\eta^2_{G_i}$, β_{G_i} and $\sigma^2_{G_i}$) or not.

Group-oblivious algorithm One of the most natural selection rules is to sort the candidates according to $\widehat{W}_i$ irrespective of their group and to keep the best m. We call this the *group-oblivious* selection algorithm. Typical examples of the group-oblivious algorithm could be admission processes in colleges where the selection is performed with respect to standardized test results (no group information is taken into account), or selection processes where the decision maker does not know the model's parameters, and in particular where it does not know the variance of the estimator (hence the name *implicit variance* in that case). This selection algorithm might be also seen as a fair treatment because the selection does not use the group label. Yet, because of the differential variance or bias, this might lead to discrimination. We will discuss that in Theorem 1.

Bayesian-optimal algorithm When the variance of the noise is known, an alternative selection algorithm is what we call the *Bayesian-optimal* algorithm. This algorithm knows all the parameters of the problem (the quality distribution, the variances of noise σ^2_G, and the biases β_G) and chooses the candidates with the largest expected quality given the estimate $\widehat{W}_i$. Since $(W_i, \widehat{W}_i)$ is a bivariate normal random vector, then using the property of conditional expectation for normal random vectors, the expected quality of candidate given its quality estimate can be expressed as:

$$\widetilde{W}_i = \mathrm{E}(W_i|\widehat{W}_i) = \frac{\eta^2_{G_i}}{\sigma^2_{G_i} + \eta^2_{G_i}}(\widehat{W}_i + \beta_{G_i}) + \left(1 - \frac{\eta^2_{G_i}}{\sigma^2_{G_i} + \eta^2_{G_i}}\right)\mu_{G_i}. \tag{2}$$

Note that $\widetilde{W}_i$ converges to $\widehat{W}_i + \beta_{G_i}$ as $\sigma^2_{G_i}$ tends to 0 (i.e., there is no noise) and it converges to μ_{G_i} as $\sigma^2_{G_i}$ tends to ∞. Intuitively, all candidates appear similar to the decision maker as the precision of estimation degrades. We denote by $\tilde{\sigma}^2_{G_i} = \eta^4_{G_i}/\left(\eta^2_{G_i} + \sigma^2_{G_i}\right)$ the variance of the expected quality $\widetilde{W}_i$.

Perhaps more surprisingly, the Bayesian-optimal algorithm also leads to discrimination (although in the opposite way as for the group-oblivious algorithm) as we show in Theorem 2. We illustrate how the two decision making algorithms work with an example depicted in Fig. 2. In this example, the blue candidates have higher variance $\sigma_{blue} = 3\sigma_{red}$. This implies that the posteriors $\widetilde{W}_i$ are more shrank towards the mean for blue candidates than for red candidates: as a result, the Bayesian-optimal tends to select fewer blue candidates compared to red candidates. Note that the Bayesian-optimal is only optimal *in expectation* given the information available; it needs not be optimal for a given realization (on Fig. 2 the optimal selection ex-post would be one red and one blue).

2.3. The γ-rule fairness mechanism

For a given algorithm alg $\in \{\text{obl}, \text{opt}\}$, we denote by x_A^{alg} (and x_B^{alg}) the proportion of the A-candidates (and B-candidates) that are selected, where obl stands for group-oblivious and opt for Bayesian-optimal. A selection algorithm might favor one group or the other, that is $x_A^{\text{alg}} \gg x_B^{\text{alg}}$ or $x_B^{\text{alg}} \gg x_A^{\text{alg}}$. To mitigate the inequality, the decision maker can introduce selection quotas. One example is the 4/5-rule [7] that imposes that $x_A \geq \frac{4}{5} x_B$ and $x_B \geq \frac{4}{5} x_A$.

In this paper, we consider a generalization of the 4/5-rule that is parameterized by $\gamma \in [0, 1]$. We say that a selection satisfies the *γ-rule* if

$$x_A \geq \gamma x_B \qquad \text{and} \qquad x_B \geq \gamma x_A. \tag{3}$$

A selection algorithm satisfies this constraint if and only if it picks at least $m\gamma n_A/(n_B + \gamma n_A)$ A-candidates and at least $m\gamma n_B/(n_A + \gamma n_B)$ B-candidates. Indeed, the total number of selected candidates is $m = x_A n_A + x_B n_B$ which means that $x_B = (m - x_A n_A)/n_B$. The constraint $x_A \geq \gamma x_B$ is therefore true if $x_A \geq \gamma x_B = \gamma(m - x_A n_A)/n_B$ which is true if and only if $x_A \geq \gamma m/(n_B + \gamma n_A)$. Similarly, the constraint $x_B \geq \gamma x_A$ is true if and only if $x_B \geq \gamma(m - x_B n_B)/n_A$.

This means that one can easily transform a baseline into a γ-fair algorithm by first selecting at least $m\gamma n_A/(n_B + \gamma n_A)$ A-candidates and at least $m\gamma n_B/(n_A + \gamma n_B)$ B-candidates and then filling the remaining positions according the best estimated candidates (candidates with largest $\widehat{W}_i$ if the baseline algorithm is group oblivious and with largest $\widetilde{W}_i$ if the baseline algorithm is Bayesian-optimal), irrespective of their group. This is what defines the *γ-fair group-oblivious* and *γ-fair Bayesian-optimal* algorithms.

When $\gamma = 0$, the γ-fair version of a baseline algorithm reduces to the original unconstrained algorithm (the algorithm that does not to take into account fairness). When $\gamma = 1$, the γ-rule mechanism corresponds to the classical notion of *demographic parity* [16] that mandates that the selection rates be equal across different groups. We highlight the demographic parity mechanism as a special and important case of the γ-rule. Note that because n_A, n_B and m are integer variables, it might be impossible to satisfy the constraints in (3) when γ is too close to 1. In such a case, we say that an algorithm is γ-fair if the constraint (3) is satisfied up to one candidate.

2.4. Simplification of the selection problem for large n and m

In the remainder of the paper, we study the selection problem when the number of candidates is large. That is, we assume that there exist fixed fractions $p_A, \alpha \in (0, 1)$ such that

$$n_A = \lfloor p_A n \rfloor \qquad n_B = \lceil (1 - p_A) n \rceil \qquad m = \lfloor \alpha n \rfloor,$$

and let n grow. Our theoretical results are obtained in the limit where n goes to infinity (similarly to [8,9]). In Section 5.3 we show numerically that our results for $n = \infty$ continue to hold for finite selection sizes. Note that p_A represents the fraction of A-candidates in the population while α represents the global selection ratio (or budget).

For a finite n, the selection algorithms presented in Sections 2.2-2.3 are hard to analyze because the probability for a candidate to be selected depends on all other candidates. As we prove below, characterizing the performance of a selection problem is simpler when the number of candidates n is infinite because there is an equivalence between the algorithms presented in the previous sections and threshold-based algorithm. A threshold-based algorithm uses two thresholds $\hat{\theta}_A$ and $\hat{\theta}_B$ and selects all G_i-candidates, such that $\widehat{W}_i \geq \hat{\theta}_{G_i}$.[7] For given thresholds $\hat{\theta}_A$ and $\hat{\theta}_B$, we denote the expected utility of the corresponding selection by $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B)$:

$$\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B) = \mathrm{E}\left[W_i \mid \widehat{W}_i \geq \hat{\theta}_{G_i} \right].$$

Hence, the selection of a candidate does not depend on the qualities of the other individuals. Also, as we show in the next theorem, the fraction of A-candidates that are selected becomes deterministic as n goes to infinity.

Lemma 1. *For any of the selection algorithms presented in Sections 2.2-2.3,*

1. *there exists a deterministic fraction $x_A \in [0, 1]$ such that the fraction of A-candidates that are selected by the algorithm converges (in probability) to x_A as n grows;*
2. *there exist deterministic thresholds $\hat{\theta}_A, \hat{\theta}_B$ such that the expected utility of this algorithm converges to $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B)$.*

Proof Sketch. The above result is essentially a direct consequence of the law of large numbers. By the Glivenko-Cantelli theorem, the empirical distribution of the estimated qualities of the G-candidates converges to the distribution of $\widehat{W}_G$ as

[7] Note that the Bayesian-optimal algorithm can also be written that way, with appropriate thresholds, because *within a given group* the expected qualities $\widetilde{W}_i$ are in the same order as the signals $\widehat{W}_i$.

Table 1
Summary of notation.

W_i	latent quality of candidate i
$\widehat{W}_i$	estimated quality of candidate i
$\widetilde{W}_i$	expected value of latent quality of candidate i given the estimate $\widehat{W}_i$
μ_G	expected value of latent quality W_G
η_G^2	variance of latent quality W_G
σ_G^2	variance of additive noise
$\hat{\sigma}_G^2$	variance of estimated quality $\widehat{W}_G$. It equals $\sigma_G^2 + \eta_G^2$
$\tilde{\sigma}_G^2$	variance of expected quality $\widetilde{W}_G$. It equals $\eta_G^4/(\eta_G^2 + \sigma_G^2)$
x_G^{alg}	fraction of G-candidates that are selected by a given algorithm "alg"
$\hat{\theta}_G^{\text{alg}}$	threshold above which G-candidates are selected by the algorithm "alg"
ϕ, Φ, Φ^{-1}	PDF, CDF and quantile of the standard normal distribution $\mathcal{N}(0,1)$

$n \to \infty$. This shows that taking the best $\lfloor np_A x_A \rfloor$ A-candidates or taking all A-candidates above the x_A-quantile of the distribution $\widehat{W}_A$ is asymptotically equivalent as $n \to \infty$. □

For these given thresholds $\hat{\theta}_A, \hat{\theta}_B$, the fractions of selected candidates are $\mathrm{P}(\widehat{W}_i \geq \hat{\theta}_{G_i})$. Using the above definition, we denote by $\mathcal{U}(x_A)$ the expected utility of a threshold-type selection algorithm that selects A-candidates with probability x_A and that satisfies the selection size constraints in expectation:

$$\mathcal{U}(x_A) = \mathcal{V}(\hat{\theta}_A, \hat{\theta}_B), \text{ where } \hat{\theta}_A, \hat{\theta}_B \text{ are such that } \begin{cases} \mathrm{P}(\widehat{W}_i \geq \hat{\theta}_A \mid G_i = A) = x_A, \\ \mathrm{P}(\widehat{W}_i \geq \hat{\theta}_{G_i}) = \alpha. \end{cases} \tag{4}$$

Note that combining the constraints in (4) immediately gives that such an algorithm selects B-candidates with probability $x_B = (\alpha - x_A p_A)/(1 - p_A)$. Hence, it is sufficient to describe the algorithm with x_A.

The above definition of expected quality is not directly applicable to the selection algorithms presented in Section 2.2 because those algorithms are defined neither in terms of fraction of selected candidates nor in terms of thresholds. In fact, for a given selection algorithm, the fractions of selected A- and B-candidates depend on the realizations of the random variables representing the quality (W_i) and the estimated quality ($\widehat{W}_i$). As a result, these fractions (x_A and x_B) are random variables. For instance, if because of randomness the A-candidates are evaluated much worse than the B-candidates, then x_A will be 0 for the group-oblivious algorithm. Lemma 1 shows that when the population is large, these random fluctuations disappear. It shows that, when n is large, the performance of the various algorithms is simply characterized by x_A.

For a finite n, characterizing precisely the utility of an algorithm like group-oblivious is computationally difficult due to the correlations between the selection of the different agents. Lemma 1 allows us to greatly simplify the study of the performance of the various algorithms because the function $\mathcal{U}$, defined in (4), depends only on one parameter x_A, and is simpler to characterize than the expectation over a finite number of candidates n.

2.5. Summary of main notation

We denote respectively by x_A^{obl}, $x_A^{\gamma\text{-obl}}$, x_A^{opt} and $x_A^{\gamma\text{-opt}}$ the asymptotic fraction of A-candidates that are selected for the group-oblivious, the γ-fair group-oblivious, the Bayesian-optimal and the γ-fair Bayesian-optimal algorithms. We also identify an important subcase of the γ-rule for $\gamma = 1$. In this case both the γ-fair group-oblivious algorithm and the γ-fair Bayesian-optimal algorithm select A-candidates at rate $x_A = \alpha$, so there is no difference between them. We name the corresponding selection algorithm as *demographic parity algorithm*.

We denote the expected performance of the introduced algorithms by

$$\mathcal{U}^{\text{obl}} = \mathcal{U}(x_A^{\text{obl}}); \qquad \mathcal{U}^{\gamma\text{-obl}} = \mathcal{U}(x_A^{\gamma\text{-obl}}); \qquad \mathcal{U}^{\text{opt}} = \mathcal{U}(x_A^{\text{opt}}); \qquad \mathcal{U}^{\gamma\text{-opt}} = \mathcal{U}(x_A^{\gamma\text{-opt}}); \qquad \mathcal{U}^{\text{dp}} = \mathcal{U}(x_A^{\text{dp}}).$$

We summarize the other notation in Table 1.

3. Analysis of the general model

In this section, we present the main technical results of the paper in the most general model. The results that we prove in this section are quite abstract; to make things more concrete and provide more intuitive results, we will instantiate this general model in important sub-cases in Section 4.

We start by showing why the two baseline algorithms lead to discrimination, in Theorem 1 for the group-oblivious and in Theorem 2 for the Bayesian-optimal algorithm. Then, we specify in Theorem 3 conditions under which the γ-rule fairness mechanism increases the utility of selection compared to the unconstrained group-oblivious algorithm. Although it is clear that the γ-rule mechanism cannot increase the utility of the Bayesian-optimal algorithm (since it is an expected utility and

the Bayesian-optimal algorithm maximizes it by definition), we prove in Theorem 4 that the ratio of the utilities of the unconstrained Bayesian-optimal and the γ-fair Bayesian-optimal algorithms is bounded.

3.1. Discrimination of baseline selection algorithms

Recall that we assume (without loss of generality) that group A is the high-variance group, that is $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$. Then, the distribution of $\widehat{W}_A$ has longer tails compared to the distribution of $\widehat{W}_B$. Thus, if the selection size is small, A-candidates will be selected by the group-oblivious algorithm at higher rate compared to B-candidates because the probability to estimate an A-candidate as "outstanding" is higher than for B-candidates. In contrast, if the selection size is large, the chance of estimating an A-candidate as poor is larger than for B-candidates, in which case the group-oblivious algorithm selects a lower fraction of A-candidates. This can be formally stated as follows.

Theorem 1. *Assume without loss of generality that $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$. When using the group-oblivious selection algorithm, the selection rates for A- and B-candidates, x_A^{obl} and x_B^{obl}, satisfy:*

$$x_A^{\text{obl}} > x_B^{\text{obl}} \text{ if and only if } \alpha < \Phi\left(\frac{\Delta\mu - \Delta\beta}{\Delta\hat{\sigma}}\right),$$

where $\Delta\mu = \mu_A - \mu_B$, $\Delta\beta = \beta_A - \beta_B$ and $\Delta\hat{\sigma} = \hat{\sigma}_A - \hat{\sigma}_B$.

Proof Sketch. The group-oblivious algorithm sorts candidates by their estimated qualities $\widehat{W}_i$ and takes the best αn by applying a group-independent threshold $\hat{\theta}$. The expression for the selection rates $x_G^{\text{obl}} = 1 - \Phi\left(\frac{\hat{\theta} - \mu_G + \beta_G}{\hat{\sigma}_G}\right)$ and a simple rearrangement allows us to find such sizes of budget α for which selection rates for both groups become equal $x_A^{\text{obl}} = x_B^{\text{obl}}$. The result then follows from the corresponding properties of normal CDF and our assumption that $\hat{\sigma}_A > \hat{\sigma}_B$. A detailed proof is given in Appendix A.1. □

The above result implies that, for a small selection budget, the group-oblivious algorithm will select high-variance candidates at a higher rate. Note that this result does not assume that this higher variance comes from the variance of the true quality (η_A^2 and η_B^2) or from the variance of the estimates (σ_A^2 and σ_B^2). It is only assumed that the variance of $\widehat{W}_A$, equal to $\hat{\sigma}_A^2 = \sigma_A^2 + \eta_A^2$, is larger than the one of $\widehat{W}_B$.

As we show below, nearly the opposite is true for the Bayesian-optimal algorithm: for a small selection budget, in the case of group-independent variance of the latent quality ($\eta_A^2 = \eta_B^2$), a Bayesian-optimal algorithm will select fewer candidates from the high-variance group. In the case where $\eta_A^2 \neq \eta_B^2$, though, which group is underrepresented will be determined by the variances $\tilde{\sigma}$ and not $\hat{\sigma}$, see our discussion below the theorem. Note also that the specific budget threshold at which the transition happens is not the same as for the group-oblivious algorithm.

Theorem 2. *Assume that $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$. When using the Bayesian-optimal selection algorithm, the selection rates for A- and B-candidates, x_A^{opt} and x_B^{opt}, satisfy:*

$$x_A^{\text{opt}} < x_B^{\text{opt}} \text{ if and only if } \alpha < \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right),$$

where $\Delta\mu = \mu_A - \mu_B$, $\Delta\tilde{\sigma} = \tilde{\sigma}_A - \tilde{\sigma}_B$.

Proof Sketch. In the Bayesian-optimal algorithm, the candidates are sorted by their expected qualities $\widetilde{W}$ and a group-independent threshold is applied to select the best αn candidates. The expression (2) for the expected quality $\widetilde{W}_G$ allows us to compare the selection rates x_A^{opt} and x_B^{opt} for different groups A or B and to find such value of budget α for which $x_A^{\text{opt}} = x_B^{\text{opt}}$. Then using the fact that $\widetilde{W}_G$ follows normal law and the relation between $\tilde{\sigma}_A$ and $\tilde{\sigma}_B$, we obtain our result. A complete proof can be found in Appendix A.2. □

The result of Theorem 2 is consistent with the observation from Phelps [27] in a simpler setting (without bias and with group-independent distribution of the latent quality W): in the presence of differential variance, the candidates from the high-variance group will appear more similar to each other to the decision maker, hence the distribution of computed expected quality will have a longer tail for the low-variance group. As a consequence, for small enough selection budgets, candidates from the high-variance group will be selected at a lower rate.

Note that the result in Theorem 2 imposes a condition on the order between $\tilde{\sigma}_A^2 = \eta_A^4 / \left(\eta_A^2 + \sigma_A^2\right)$, the variance of $\widetilde{W}_A$, and $\tilde{\sigma}_B^2$; but it is not conditional on the relation between $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$, i.e., both $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$ and $\hat{\sigma}_A^2 \leq \hat{\sigma}_B^2$ are allowed. Hence the condition $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$ comes without loss of generality for this result. Note also that in the case where the variances of the true quality are the same for both groups ($\eta_A = \eta_B$), the two conditions from Theorems 1 and 2 are equivalent, that is $\tilde{\sigma}_A < \tilde{\sigma}_B$ if and only if $\hat{\sigma}_A > \hat{\sigma}_B$ (since it holds if and only if $\sigma_A > \sigma_B$). The main special cases that we consider in Section 4 (specifically those of Sections 4.1 and 4.2) are in this case (i.e., satisfy $\eta_A = \eta_B$).

3.2. The γ-rule mechanism can increase the utility of the group-oblivious algorithm

As we show in Theorem 1, the group-oblivious algorithm leads to overrepresentation of the high-variance group A, if the budget α is small. To mitigate this effect, the decision maker can use the γ-rule fairness mechanism introduced in Section 2.3.

In the next theorem, we provide a condition on budgets α for which using the γ-fair group-oblivious algorithm attains larger quality of selection compared to the unconstrained group-oblivious algorithm. The main message of this theorem is that if A-candidates have larger variability of their estimate compared to B-candidates (i.e., $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$) and the variance of expected quality for A-candidates is smaller than for B-candidates (i.e., $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$), then the γ-fair group-oblivious algorithm leads to larger quality of selection compared to the group-oblivious algorithm for both small and large budgets α. As said earlier, when $\eta_A = \eta_B$, these conditions are always satisfied, up to switching the groups A and B. At the same time, there may exist a region of budgets α such that the γ-rule fairness mechanism harms the quality of selection compared to the group-oblivious algorithm.

Theorem 3. *Without loss of generality, assume that the estimates of quality for A-candidates have larger variance than for B-candidates $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$. Assume also that the variance of the expected quality is smaller for A-candidates than for B-candidates ($\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$), and let us define*

$$\alpha^{\min} = \min\left\{\Phi\left(\frac{\Delta\mu - \Delta\beta}{\Delta\hat{\sigma}}\right), \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)\right\},\ \alpha^{\max} = \max\left\{\Phi\left(\frac{\Delta\mu - \Delta\beta}{\Delta\hat{\sigma}}\right), \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)\right\},$$

where $\Delta\mu = \mu_A - \mu_B$, $\Delta\beta = \beta_A - \beta_B$, $\Delta\hat{\sigma} = \hat{\sigma}_A - \hat{\sigma}_B$ and $\Delta\tilde{\sigma} = \tilde{\sigma}_A - \tilde{\sigma}_B$. We have:

(i) For any $\alpha \in (0, \alpha^{\min}) \cup (\alpha^{\max}, 1)$, the demographic parity algorithm strictly improves the selection quality compare to the group-oblivious algorithm and the γ-fair group-oblivious algorithm for $\gamma < 1$ weakly improves it:

$$\mathcal{U}^{\mathrm{dp}} > \mathcal{U}^{\gamma\text{-obl}} \geq \mathcal{U}^{\mathrm{obl}}.$$

(ii) If $\alpha^{\min} = \alpha^{\max}$, then for $\alpha = \alpha^{\min} = \alpha^{\max}$ one has $\mathcal{U}^{\mathrm{dp}} = \mathcal{U}^{\gamma\text{-obl}} = \mathcal{U}^{\mathrm{obl}}$.

(iii) Assume that $\alpha^{\min} \neq \alpha^{\max}$, then there exists $[\tilde{\alpha}^{\min}, \tilde{\alpha}^{\max}]$, where $\tilde{\alpha}^{\min} > \alpha^{\min}$ and $\tilde{\alpha}^{\max} < \alpha^{\max}$, such that for any $\alpha \in [\tilde{\alpha}^{\min}, \tilde{\alpha}^{\max}]$, the demographic parity algorithm strictly harms the selection quality compared to the group-oblivious algorithm and the γ-fair group-oblivious algorithm for $\gamma < 1$ weakly harms it:

$$\mathcal{U}^{\mathrm{dp}} < \mathcal{U}^{\gamma\text{-obl}} \leq \mathcal{U}^{\mathrm{obl}}.$$

Proof. We prove in Theorem 1 that if $\alpha < \Phi\left((\Delta\mu - \Delta\beta)/\Delta\hat{\sigma}\right)$, then the group-oblivious algorithm leads to overrepresentation of the high-variance group A. At the same time, the group-oblivious algorithm leads to underrepresentation of the group A if $\alpha > \Phi\left((\Delta\mu - \Delta\beta)/\Delta\hat{\sigma}\right)$. Similarly, we prove in Theorem 2 that if $\tilde{\sigma}_A < \tilde{\sigma}_B$, then for $\alpha < \Phi\left(\Delta\mu/\Delta\tilde{\sigma}\right)$, the Bayesian-optimal algorithm underrepresents the group A and for $\alpha > \Phi\left(\Delta\mu/\Delta\tilde{\sigma}\right)$ it overrepresents the group A.

Recall that for any value of α, the demographic parity algorithm requires that candidates from both groups, A and B, must be selected at equal rates, i.e., $x_A^{\mathrm{dp}} = x_B^{\mathrm{dp}} = \alpha$. It means that if $\alpha \in (0, \alpha^{\min}) \cup (\alpha^{\max}, 1)$, then the demographic parity algorithm will perform a selection such that either $x_A^{\mathrm{obl}} < x_A^{\mathrm{dp}} < x_A^{\mathrm{opt}}$ or $x_A^{\mathrm{opt}} < x_A^{\mathrm{dp}} < x_A^{\mathrm{obl}}$ (also $x_A^{\mathrm{obl}} \leq x_A^{\gamma\text{-obl}} < x_A^{\mathrm{opt}}$ or $x_A^{\mathrm{opt}} < x_A^{\gamma\text{-obl}} \leq x_A^{\mathrm{obl}}$ for $\gamma < 1$). In Appendix A.3, we prove that the selection quality $\mathcal{U}$ is a concave function of x_A with a single maximum at $x_A = x_A^{\mathrm{opt}}$. Hence, from this property we conclude that $\mathcal{U}^{\mathrm{dp}} > \mathcal{U}^{\mathrm{obl}}$ and $\mathcal{U}^{\gamma\text{-obl}} \geq \mathcal{U}^{\mathrm{obl}}$ for $\gamma < 1$. Finally, (iii) is due to the fact that the utility $\mathcal{U}$ is a continuous and smooth function of x_A as we prove in Appendix A.3. □

While the statement of Theorem 3 is somewhat complex due to its generality, in special cases (e.g., that of Section 4.1) we have $\alpha^{\min} = \alpha^{\max}$. This means that in the special case of Section 4.1, we are always in case (i) of Theorem 3: the γ-fair group-oblivious algorithm attains a larger utility than the corresponding baseline (or at worst an equal utility).

Note that the statement of Theorem 3 is under the assumption that $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$ *and* $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$. As discussed earlier, this assumption may not be without loss of generality if $\eta_A \neq \eta_B$. If it does not hold, then using the demographic parity algorithm could lead to a worse utility than the group-oblivious algorithm. Even in this case, however, the ratio $\mathcal{U}^{\mathrm{obl}}/\mathcal{U}^{\mathrm{dp}}$ remains bounded. Indeed, we can write $\mathcal{U}^{\mathrm{obl}}/\mathcal{U}^{\mathrm{dp}} \leq \mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$ and the ratio $\mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$ itself is upper-bounded as we show in the next section (Theorem 4).

3.3. Bounds on the decrease of utility due to imposing γ-rule on the Bayesian-optimal algorithm

By definition, the Bayesian-optimal algorithm maximizes the utility of the selection which means that imposing a γ-rule cannot increase the expected utility of the selection—in most cases it decreases it. In this section, however, we obtain a bound on the ratio of utilities for Bayesian-optimal and γ-fair Bayesian-optimal algorithms. This is stated in the following theorem:

Theorem 4. *Assume that $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$ and that $\mu_A, \mu_B \geq 0$, then for any budget α the ratio $\mathcal{U}^{\text{opt}} / \mathcal{U}^{\gamma\text{-opt}}$ satisfies the following bound:*

$$1 \leq \frac{\mathcal{U}^{\text{opt}}}{\mathcal{U}^{\gamma\text{-opt}}} \leq 1 + \begin{cases} -\frac{\alpha}{p_A + p_B/\gamma} \cdot g(\mu_G, \tilde{\sigma}_G, p_G, \alpha), & \text{if } \alpha \leq \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right) \\ \left(1 - \frac{\alpha}{p_A + p_B\gamma}\right) \cdot g(\mu_G, \tilde{\sigma}_G, p_G, \alpha), & \text{if } \alpha > \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right) \end{cases}$$

where $\Delta\mu = \mu_A - \mu_B$, $\Delta\tilde{\sigma} = \tilde{\sigma}_A - \tilde{\sigma}_B$ and $g(\mu_G, \tilde{\sigma}_G, p_G, \alpha) = \frac{p_A}{\alpha} \frac{\Delta\mu + \Phi^{-1}(1-\alpha)\Delta\tilde{\sigma}}{\sum_G p_G \mu_G + \frac{\phi\left(\Phi^{-1}(1-\alpha)\right)}{\alpha} \sum_G p_G \tilde{\sigma}_G}$.

Proof Sketch. The first inequality is due to the fact that the utility function $\mathcal{U}(x_A)$ is strictly concave and that it attains its maximum at $x_A = x_A^{\text{opt}}$ as we show in Appendix A.3.

To prove the second inequality, we need a few preparatory steps. First, using the result of Theorem 2, we obtain that for the budgets $\alpha < \Phi\left(\Delta\mu/\Delta\tilde{\sigma}\right)$, we have $x_A^{\text{opt}} \leq x_A^{\gamma\text{-opt}} < x_A^{\text{dp}}$. Using the concavity of $\mathcal{U}$ and the mean value theorem from real analysis, we obtain: $\frac{\mathcal{U}(x_A^{\text{opt}}) - \mathcal{U}(x_A^{\text{dp}})}{x_A^{\text{opt}} - x_A^{\text{dp}}} \geq \mathcal{U}'(x_A = x_A^{\text{dp}}) \implies \mathcal{U}(x_A^{\text{opt}}) - \mathcal{U}(x_A^{\text{dp}}) \leq -\alpha \cdot \mathcal{U}'(x_A^{\text{dp}})$. After, we divide both parts by $\mathcal{U}(x_A = x_A^{\text{dp}})$. The expressions for $\mathcal{U}'(x_A = x_A^{\text{dp}})$ and $\mathcal{U}(x_A = x_A^{\text{dp}})$ can be written explicitly using the equation derived in Appendix A.3. A complete proof is given in Appendix A.4. □

The expression in Theorem 4 is general but complex due to the large number of model parameters. It can be simplified as we tighten up some of the assumptions (see Section 4). There are also interesting behaviors to observe for some values of the parameters. First, if the size of group A becomes small (i.e., $p_A \to 0$), we observe that the function g converges to 0, hence the upper bound converges to 1. This is expected, since the introduction of the γ-rule mechanism will affect the selection in a tiny amount due to a small number of A-candidates. Second, as the selection budget decreases (i.e., $\alpha \to 0$), we can show using L'Hôpital's rule that the upper bound in this limit converges to $1 - \frac{p_A \Delta\tilde{\sigma}}{\sum p_G \tilde{\sigma}_G}$ for $\gamma = 1$. In other words, the difference in the expected values of qualities $\Delta\mu$ does not play any role. This is also quite natural, since for tiny selection budgets α, the competition is among the candidates with very large values of quality which is due to the variance of the distribution of latent quality but not their mean values.

4. Notable special cases of the general model

The results in Section 3 might be difficult to interpret without considering some specific cases. In this section, we decompose the effects of different factors by tightening up some of the assumptions of our model while keeping the others in place. We consider the following important special cases: In Section 4.1 we assume that there is no bias in the estimation of quality and that the quality distribution is group-independent. This is the model studied in [1], where the only quantity that depends on the candidate's group is the noise variance (to isolate the differential variance effect). In Section 4.2, we assume that the quality distribution is group-independent but the estimates are biased. In Section 4.3 we assume unbiased estimates but let the quality be group-dependent. All these subcases allow us to greatly simplify the results of Theorem 3 and Theorem 4.

4.1. Group-independent latent quality and unbiased estimates

In this section, we assume that the underlying quality distribution is group-independent (this is the classical assumption in the literature, see for instance [8,9]) and follows a normal law with mean μ and variance η^2. To isolate the effect of the variance, we also assume that quality estimates $\widehat{W}_i$ are unbiased, i.e., $\beta_{G_i} = 0$. The main result of this section is that imposing a fairness constraint in this context *cannot* decrease the utility compared to using the unconstrained group-oblivious baseline. We also simplify the bound of Theorem 4 on the decrease of the utility of the Bayesian-optimal algorithm due to the γ-rule fairness mechanism.

First, the following corollary relates selection ratios for two baseline algorithms. It can be obtained directly from Theorems 1 and 2. (Recall that in this special case of group-independent quality distribution, we have $\sigma_A^2 > \sigma_B^2$ if and only if $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$, which is also if and only if $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$.)

Corollary 1 *(Corollary of Theorems 1 and 2). Assume that the quality distribution is group-independent $W_i \sim \mathcal{N}(\mu, \eta^2)$ and that the quality estimates $\widehat{W}_i$ are unbiased $\beta_G = 0$, $\forall G \in \{A, B\}$. Assume without loss of generality that $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$. When using the group-oblivious selection algorithm and the Bayesian-optimal selection algorithm, the fractions x_G^{obl} and x_G^{opt} of selected candidates from each group satisfy:*

(i) $x_A^{\text{obl}} > x_B^{\text{obl}}$ *if and only if* $\alpha < 1/2$;
(ii) $x_A^{\text{opt}} < x_B^{\text{opt}}$ *if and only if* $\alpha < 1/2$.

Corollary 1 formalizes in simple terms, for the case of group-independent latent quality distributions and unbiased estimators, the discrimination that results from the two baseline algorithms. Notably, (i) states that for selection budgets below 1/2, the group-oblivious algorithm overrepresents the high-variance group. If the high-variance group is a minority, this is counter-intuitive. As noted in the introduction, however, these typically correspond to cases where the Bayesian-optimal baseline is more meaningful. Then, (ii) states that for small budgets, the Bayesian-optimal algorithm indeed underrepresents the high-variance group.

In Section 3 we specify a condition under which the γ-rule fairness mechanism is beneficial to the utility of the group-oblivious algorithm. In the special case of group-independent prior and unbiased estimator, the thresholds $\alpha^{\min}$ and $\alpha^{\max}$ defined in Theorem 3 coincide and are equal to 1/2. This implies the next theorem, which shows that for this case, the γ-rule fairness mechanism cannot decrease the average quality of a selection compared to the group-oblivious algorithm (without any condition on α).

Corollary 2 *(Corollary of Theorem 3). Assume that the quality distribution is group-independent $W_i \sim \mathcal{N}(\mu, \eta^2)$ and that the quality estimates $\widehat{W}_i$ are unbiased $\beta_G = 0, \forall G \in \{A, B\}$. Let, without loss generality, $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$, then for any budget $\alpha \neq 1/2$, the demographic parity selection algorithm provides a larger utility than the γ-fair group-oblivious selection algorithm with $\gamma < 1$, which in turn provides a larger utility than the group-oblivious selection algorithm:*

$$\mathcal{U}^{\text{dp}} > \mathcal{U}^{\gamma\text{-obl}} \geq \mathcal{U}^{\text{obl}}.$$

The above inequality is an equality when $\alpha = 1/2$.

Proof. This result is a special case of Theorem 3. Since the distribution of quality is group-independent and there is no implicit bias, then the condition in Theorem 3 holds, and $\alpha^{\min}$ and $\alpha^{\max}$ coincide and become equal to 1/2. □

As for the general case, the γ-rule fairness mechanism cannot increase the selection quality of the Bayesian-optimal baseline. In Theorem 4, we obtained a bound on the decrease of utility. In the next result, we show how this result simplifies in the modeling assumptions of the current subsection. We provide the bound for $\alpha \leq 1/2$ as it is the most interesting setting. The one for $\alpha > 1/2$ can also be easily deduced.

Corollary 3 *(Corollary of Theorem 4). Assume that quality distribution is group-independent $W_i \sim \mathcal{N}(\mu, \eta^2)$ and quality estimates $\widehat{W}_i$ are unbiased $\beta_G = 0, \forall G \in \{A, B\}$. Let, without loss of generality, $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$ and $\mu \geq 0$. Then for all $\alpha \neq 1/2$, the demographic parity selection algorithm provides a smaller utility than the γ-fair Bayesian-optimal selection algorithm with $\gamma < 1$, which in turns provides a smaller utility than the Bayesian-optimal selection algorithm. The utility ratio $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\gamma\text{-opt}}$ for any budget $\alpha \leq 1/2$ has the following bound:*

$$1 \leq \frac{\mathcal{U}^{\text{opt}}}{\mathcal{U}^{\gamma\text{-opt}}} \leq 1 + g(\alpha) \cdot \frac{1}{p_A + (1-p_A)/\gamma} \cdot \frac{p_A(\nu - 1)}{p_A + (1-p_A)\nu},$$

where $g(\alpha) = \frac{\alpha \Phi^{-1}(1-\alpha)}{\phi(\Phi^{-1}(1-\alpha))}$ and $\nu = \hat{\sigma}_A/\hat{\sigma}_B > 1$.

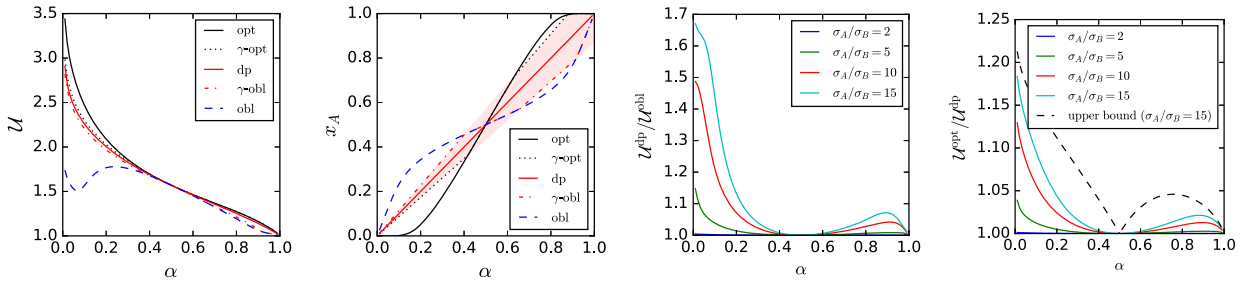
Proof. Direct from Theorem 4 when we set $\mu = \mu_A = \mu_B$ and $\eta^2 = \eta_A^2 = \eta_B^2$. □

For $\gamma = 1$, which is the case of demographic parity, we can further simplify the expression in Corollary 3. By using the fact that $g(\alpha)$ is decreasing with α and that $\lim_{\alpha \to 0} g(\alpha) = 1$, we can write

$$1 \leq \mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}} \leq 1 + \frac{p_A(\nu - 1)}{p_A + (1-p_A)\nu}.$$

Note that as ν tends to 1, meaning that there is no difference in variances between A and B group, the upper bound also tends to 1 and matches the lower bound. Most interestingly, we observe that the larger the difference in variances ν, the larger the upper bound. As ν tends to infinity, the upper bound tends to $1/(1-p_A)$. Hence, if, for instance, the high-variance group is the minority ($p_A < 1/2$), then the gap cannot be larger than 2.

In Fig. 3, we show the obtained utilities $\mathcal{U}$, the selection fractions x_A and the gap values $\mathcal{U}^{\text{dp}}/\mathcal{U}^{\text{obl}}$ and $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$ for different budgets α from 0.01 to 0.99. Fig. 3a illustrates the utilities corresponding to different selection algorithms. We observe that the utilities of the Bayesian-optimal and demographic parity selections decrease when α increases. This is expected because this graph represents the average quality of a selected candidate: the average quality decreases when the number of selected candidates increases. What is more surprising is that the behavior of the group-oblivious selection algorithm is not monotonous: the expected utility $\mathcal{U}$ increases when α goes from 0.1 to 0.3. In fact, when $\alpha < 0.1$, very few B-candidates are selected by the group-oblivious algorithm. When $\alpha \approx 0.1$–0.2, this algorithm selects a few good B-candidates which leads to an increased average performance.



(a) Utility (b) Selection fraction (c) Ratio $\mathcal{U}^{\mathrm{dp}}/\mathcal{U}^{\mathrm{obl}}$ (d) Ratio $\mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$

Fig. 3. Utility $\mathcal{U}$, selection fraction x_A and performance gaps for different budgets α. The parameters are $\mu = 1$, $\eta = 1$, $\sigma_B = 0.2$, and $p_A = 0.4$; $\sigma_A = 3$ for panels (a,b).

In Fig. 3c we show the performance gap between group-oblivious and demographic parity selection algorithms for different values of σ_A and fixed $\sigma_B = 0.2$, $\eta = 1$. The values of σ_A are such that $\sigma_A/\sigma_B = k$, $k = 2, 5, 10, 15$. We see that the gap is in general larger when the selection size α is small. This is due to the fact that as the selection size increases, the selections by the group-oblivious and demographic parity algorithms become close. The performance gap is zero when $\alpha = 0.5$ because the selections are exactly the same (due to the symmetry of the underlying quality distribution), but it becomes positive again for larger values of α. In addition, the larger the differential variance ratio σ_A^2/σ_B^2, the larger the gain that demographic parity brings.

Finally, in Fig. 3d we illustrate the performance gap between the Bayesian-optimal and the demographic parity selection algorithms for different values of σ_A and fixed $\sigma_B = 0.2$, $\eta = 1$. As in Fig. 3c, the values of σ_A are such that $\sigma_A/\sigma_B = k$, $k = 2, 5, 10, 15$. In addition, we also show the bound on the ratio $\mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$ for different values of α and for fixed $k = 15$. We see that the upper bound developed in Theorem 4 is relatively tight for small values of α, but is quite loose when $\alpha \approx 0.5$.

4.2. Group-independent latent quality distribution and biased estimates

In this section, we again assume that the underlying quality distribution is group-independent but we now assume that the estimates are both biased and with differential variance. Since the true quality distribution is group-independent, then $\mu = \mu_A = \mu_B$ and $\eta^2 = \eta_A^2 = \eta_B^2$. Recall that in this case, the conditions in Theorem 3 hold since under the assumption of $\hat{\sigma}_A^2 > \hat{\sigma}_B^2$ which is w.l.o.g., the requirement $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$ is also satisfied. The expressions for the budgets $\alpha^{\min}$ and $\alpha^{\max}$ specified in Theorem 3 can also be simplified:

$$\alpha^{\min} = \min\left\{\Phi\left(\frac{-\Delta\beta}{\Delta\hat{\sigma}}\right), \frac{1}{2}\right\}, \alpha^{\max} = \max\left\{\Phi\left(\frac{-\Delta\beta}{\Delta\hat{\sigma}}\right), \frac{1}{2}\right\}.$$

We can get several insights from this simplification. First, if both groups are subject to the same amount of bias, $\beta_A = \beta_B$, then both $\alpha^{\min}$ and $\alpha^{\max}$ coincide, $\alpha^{\min} = \alpha^{\max} = 1/2$. Hence, according to Theorem 3, the γ-rule fairness mechanism in this case is beneficial to the utility of the group-oblivious algorithm for all budgets $\alpha \neq 1/2$. For $\alpha = 1/2$, both the γ-fair group-oblivious algorithm and the group-oblivious algorithm will perform the same $\mathcal{U}^{\gamma\text{-obl}} = \mathcal{U}^{\mathrm{obl}}$ for all $\gamma > 0$. Hence, if the amount of bias is the same, the result is not different from the one when there is no bias at all (see Section 4.1), which is natural and expected. We illustrate this result in Fig. 4a which is the same as Fig. 3a.

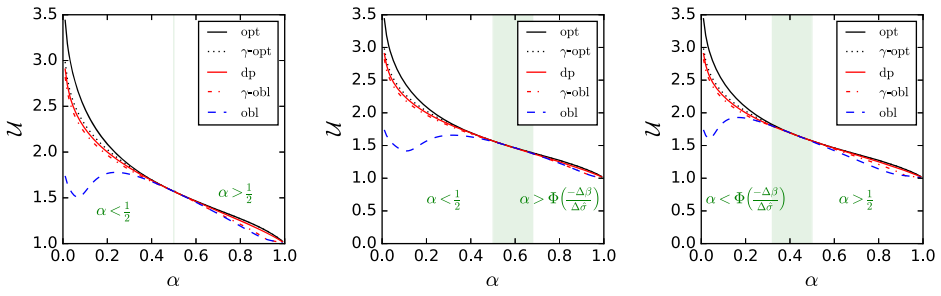
Second, if the estimate for the high-variance group A has smaller bias than for the low-variance group B, i.e., $\Delta\beta = \beta_A - \beta_B < 0$, then the γ-fair mechanism will improve the utility of the group-oblivious algorithm for all $\alpha < 1/2$. It can be seen from the fact that in this case $\Phi\left(\frac{-\Delta\beta}{\Delta\hat{\sigma}}\right) \geq 1/2$ which means that $\alpha^{\min} = 1/2$. This case is illustrated in Fig. 4b.

Perhaps counterintuitively, when implicit bias and implicit variance both affect the estimation, for some values of $\alpha \in (\alpha^{\min}, \alpha^{\max})$ specified in Theorem 3, the γ-fair group-oblivious algorithm will always perform worse than the group-oblivious algorithm. We observe the corresponding phenomenon on both Fig. 4b and Fig. 4c around the values of budgets $\alpha = 0.6$ and $\alpha = 0.4$, respectively.

Finally, note that the Bayesian-optimal algorithm and the demographic parity (implicitly) remove the biases, hence, the results and discussion from Corollary 3 can also be applied in this section.

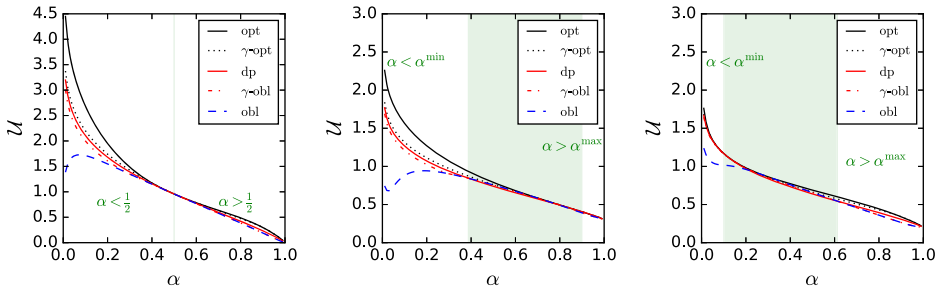
4.3. Group-dependent latent quality distribution and unbiased estimates

We now assume that there is *no bias* but that the underlying *quality distribution is group-dependent*. We can also distinguish different cases, when we isolate the effect of group-dependency of the distribution of quality by removing the implicit bias from our consideration. In this case, $\Delta\beta = 0$ and the budgets specified in Theorem 3 can be reformulated as follows:



(a) $\beta_A = \beta_B$ (b) $\beta_A < \beta_B$ (c) $\beta_A > \beta_B$

Fig. 4. The quality of selection in the presence of bias and differential variance for different budgets α. We assume that the quality distribution is group-independent, but A-candidates have larger variability of estimation compare to the B-candidates, i.e. $\sigma_A^2 > \sigma_B^2$. The quality distribution follows $\mathcal{N}(\mu = 1, \eta^2 = 1)$, the differential variance parameters are equal to $\sigma_A = 3$ and $\sigma_B = 0.2$. The bias parameters are equal to $\beta_A = 1, \beta_B = 1$ for 4a, $\beta_A = 0, \beta_B = 1$ for 4b and $\beta_A = 1, \beta_B = 0$ for 4c. The shaded green region indicates the case $\alpha \in [\alpha^{\min}, \alpha^{\max}]$, *i.e.*, when no increase of performance is guaranteed by Theorem 3.



(a) $\mu_A = \mu_B$ (b) $\mu_A < \mu_B$ (c) $\mu_A > \mu_B$

Fig. 5. The quality of selection in the presence of differential variance for different budgets α. We assume that A-candidates have larger variability of their estimates compare to B-candidates $\hat{\sigma}_A > \hat{\sigma}_B$ as well as the relative amount of noise is larger for A-candidates than for B-candidates $\tilde{\sigma}_A < \tilde{\sigma}_B$. The implicit variance parameters are equal to $\sigma_A = 3$ and $\sigma_B = 1$. The distribution of quality is $\mathcal{N}(\mu_A = 0, \eta_A = 1)$ and $\mathcal{N}(\mu_B = 0, \eta_B = 2)$ for 5a, $\mathcal{N}(\mu_A = 0, \eta_A = 1)$ and $\mathcal{N}(\mu_B = 0.5, \eta_B = 1)$ for 5b and $\mathcal{N}(\mu_A = 0.5, \eta_A = 1)$ and $\mathcal{N}(\mu_A = 0, \eta_B = 1)$ for 5c. The shaded green region indicates the case $\alpha \in [\alpha^{\min}, \alpha^{\max}]$, *i.e.*, when no increase of performance is guaranteed by Theorem 3.

$$\alpha^{\min} = \min\left\{\Phi\left(\frac{\Delta\mu}{\Delta\hat{\sigma}}\right), \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)\right\},\ \alpha^{\max} = \max\left\{\Phi\left(\frac{\Delta\mu}{\Delta\hat{\sigma}}\right), \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)\right\}.$$

We can draw several conclusions from this simplification. First, if both groups have equal means $\mu_A = \mu_B$ and if $\hat{\sigma}_A > \hat{\sigma}_B$, $\tilde{\sigma}_A < \tilde{\sigma}_B$, then the condition in Theorem 3 simplifies to $\alpha^{\min} = \alpha^{\max} = 1/2$, which is equivalent to the result in Corollary 2. Thus, in this case, the γ-rule mechanism improves the quality of group-oblivious selection for all budgets $\alpha \neq 1/2$. (If $\alpha = 1/2$, then $\mathcal{U}^{\text{obl}} = \mathcal{U}^{\gamma\text{-obl}}$ for all γ.) We illustrate this case in Fig. 5a and it is the same result as in Section 4.1. Second, if both groups have equal variances of quality $\eta_A^2 = \eta_B^2 = \eta^2$, then the condition $\tilde{\sigma}_A < \tilde{\sigma}_B$ from Theorem 3 holds automatically. We illustrate different cases of relations between μ_A and μ_B in Fig. 5b and Fig. 5c. Unfortunately, the bound on $\mathcal{U}^{\text{opt}} / \mathcal{U}^{\gamma\text{-opt}}$ in Theorem 4 cannot be further simplified for the case of group-dependent quality distribution.

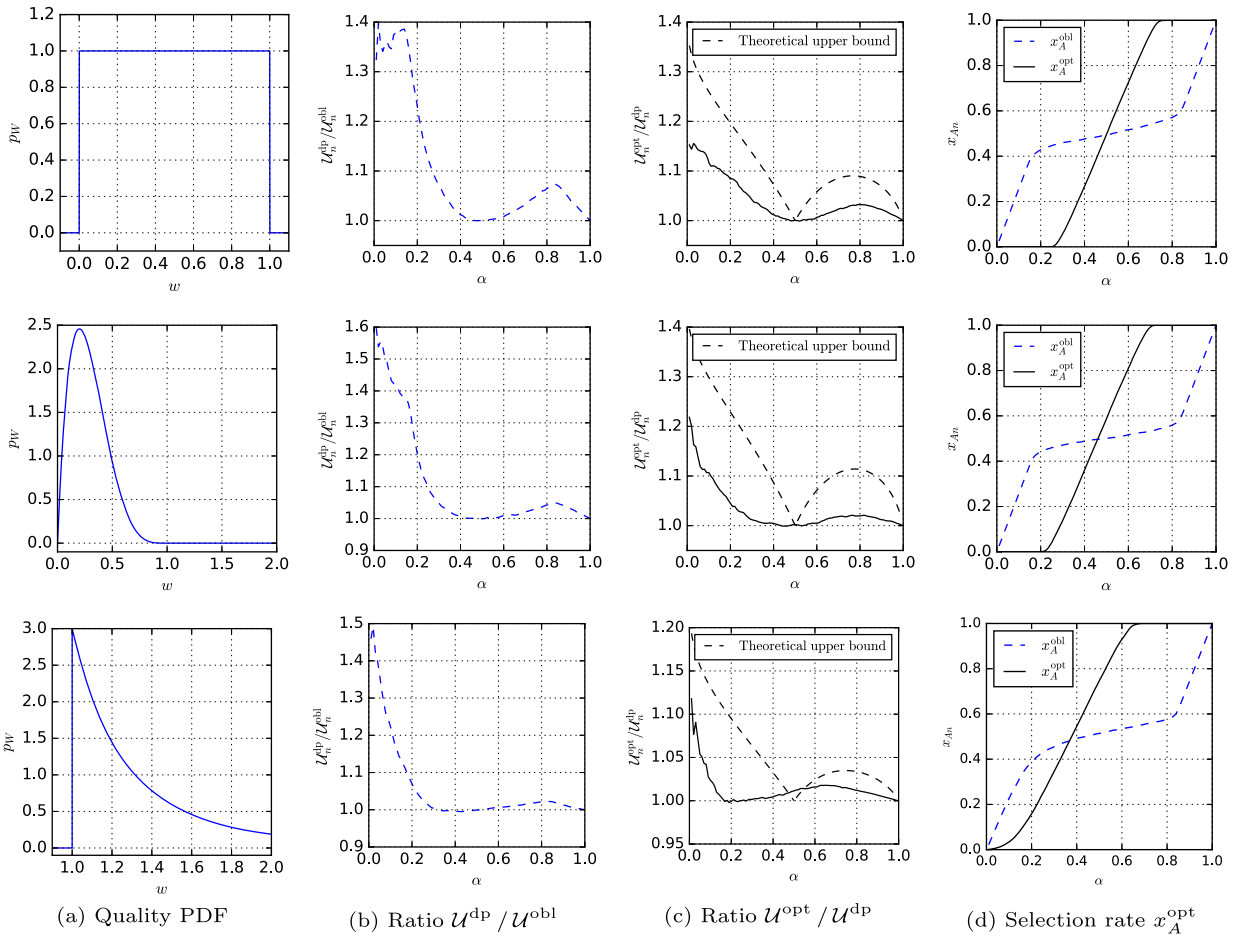
5. Experiments

In this section,[8] we challenge our theoretical results by using sets of data that do not satisfy our assumptions. We show in Section 5.1 that the results are qualitatively similar when the candidates' true quality comes from a non-normal distribution. We also observe a similar behavior when considering in Section 5.2 an artificial scenario that we construct using a real dataset coming from the national Indian exam data. We conclude in Section 5.3 with experiments that show that a case with $n = 50$ candidates behaves similarly as with $n = \infty$.

5.1. Synthetic data with non-normal quality

Our assumption in the theoretical evaluation of Sections 3-4 was that qualities W follow a normal distribution. In some cases, however, the quality distribution is quite different from normal and can be better modeled by a power law [8], this

[8] All codes are available at: https://gitlab.inria.fr/vemelian/differential-variance-code.



(a) Quality PDF (b) Ratio $\mathcal{U}^{\mathrm{dp}}/\mathcal{U}^{\mathrm{obl}}$ (c) Ratio $\mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$ (d) Selection rate x_A^{opt}

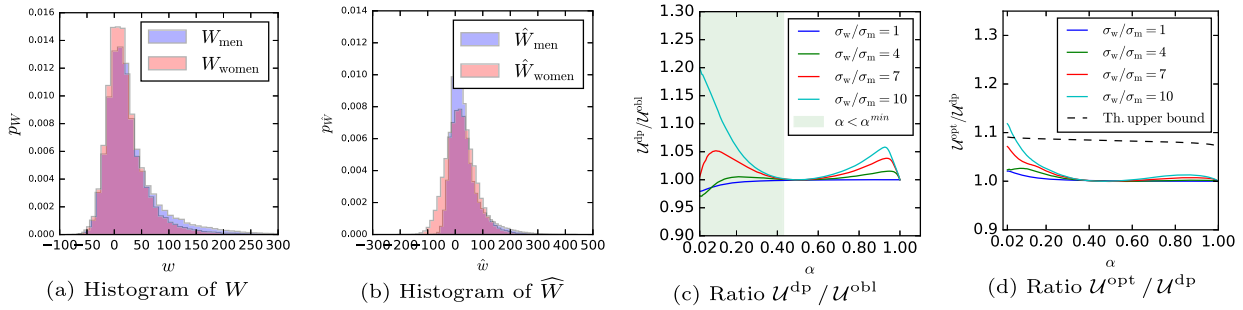
Fig. 6. Synthetic data for different prior distributions of quality W: Uniform on [0, 1], Beta(2,5), and Pareto(1,3): Effects of fairness on utility $\mathcal{U}$ and selection rate x_A. The parameters are $p_A = 0.4$, $\sigma_A = 3$ and $\sigma_B = 0.2$. The number of candidates is fixed to $n = 10,000$.

for example the case for wealth, income or number of citations [31], meaning that a minority possesses a large fraction of the aggregate quality. In this experiment, we vary the quality distribution and consider other distributions of quality W: a Uniform distribution on [0, 1], a Beta distribution with the shape parameter equal to 2 and the scale parameter equal to 5 or a Pareto distribution with a scale 1 and shape 3 (whose PDF $p_W(w) = \frac{3}{w^4}$). We generate a single dataset of size $n = 10,000$. For this dataset we perform a group-oblivious, a demographic parity and a Bayesian-optimal selection. In Fig. 6, we report the sample utilities $\mathcal{U}_n$ and sample selection rates x_{An}. Note that in this section we consider no bias and group-independent quality distribution. Each line corresponds to a different prior quality.

In Fig. 6b, we show the performance gap between the group-oblivious and the demographic parity algorithms. We see that the demographic parity improves the utility of the group-oblivious algorithm in most of the cases and that the largest gap corresponds to the smallest budget α. Note that contrary to Corollary 2, the demographic parity does not always improve the utility of the group-oblivious algorithm. Yet, the loss due to the demographic parity is never larger than 0.1% while the gain can be up to 60%.

In Fig. 6c, the performance ratio for the Bayesian-optimal algorithm and the demographic parity algorithm is shown. As expected, the demographic parity harms the utility of the Bayesian-optimal algorithm for both small and large values of budget α. As the budget α increases, the performance gap decreases. To estimate the ratio between the Bayesian-optimal algorithm and the demographic parity, we plot also the value of upper bound from Theorem 3 that is calculated under an assumption that the quality distribution is normal. We observe that the bound is not tight, however, still dominates the values of $\mathcal{U}^{\mathrm{opt}}/\mathcal{U}^{\mathrm{dp}}$ for most values of budget α.

Finally, in Fig. 6d, we show how the selection fractions x_A^{obl} and x_A^{opt} depend on α. We see that for small budgets α, the group-oblivious algorithm tends to select more from group A, while for large budgets, the situation is opposite. In contrast, the Bayesian-optimal algorithm always selects A-candidates at lower rate if the selection budget α is small.



(a) Histogram of W (b) Histogram of $\widehat{W}$ (c) Ratio $\mathcal{U}^{\text{dp}}/\mathcal{U}^{\text{obl}}$ (d) Ratio $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$

Fig. 7. Distribution of W and $\widehat{W}$ given gender, and selection ratios for **IIT-JEE dataset** [32]. Mean values and standards deviations of W for two groups are: $\mu_{\text{m}} = 30.8$, $\eta_{\text{m}} = 51.8$, $\mu_{\text{w}} = 21.2$, $\eta_{\text{w}} = 39.3$. Added noise has standard deviation $\sigma_{\text{m}} = 10$ and $\sigma_{\text{w}} = k \cdot \sigma_{\text{m}}$; $k = 4$ in plot (b).

5.2. IIT-JEE scores dataset

In this section, we aim to consider a scenario in which the underlying quality distributions are non-normal and non-symmetric, and are group-dependent. To easily construct such a case, we create an artificial scenario by using a real dataset, the IIT-JEE dataset [32], with joint entrance exam results in India in 2009. These scores are used as an admission criterion to enter the high-rated universities. The dataset consists of $n = 384,977$ records. Every record has information about one student: its name, gender, grade for Mathematics, Physics, Chemistry and total grade. In the dataset, there are 98,028 women and 286,942 men. This dataset is the same as the one considered in [9].

In order to construct a model of differential variance, we consider an artificial scenario where the field "grade" is the true latent quality W of the candidates. The mean values and standard deviations of W for the two groups are: $\mu_{\text{men}} = 30.8$, $\eta_{\text{men}} = 51.8$, $\mu_{\text{women}} = 21.2$, $\eta_{\text{women}} = 39.3$. We then suppose that an unbiased estimator $\widehat{W}$ of the grade is observed. The standard deviation of estimation for male candidates is set to $\sigma_{\text{men}} = 10$. For the women group, which is the minority group, we consider different cases: $\sigma_{\text{women}} = k \cdot \sigma_{\text{men}}$, for $k = 1, 4, 7, 10$. The distribution of grades W and observed values $\widehat{W}$ for $k = 4$ are shown in Fig. 7a and 7b.

For the dataset we perform a group-oblivious (select best m), a demographic parity selection (select best m, but maintain the demographic parity condition $x_A = x_B$ up to one candidate) and a Bayesian-optimal selection. The selection size varies from 2% to 100% of total number of candidates, i.e., out of 384,977 students the decision maker selects 7,700 students or more. A selection rate of 2% was set by IIT in 2009 [9].

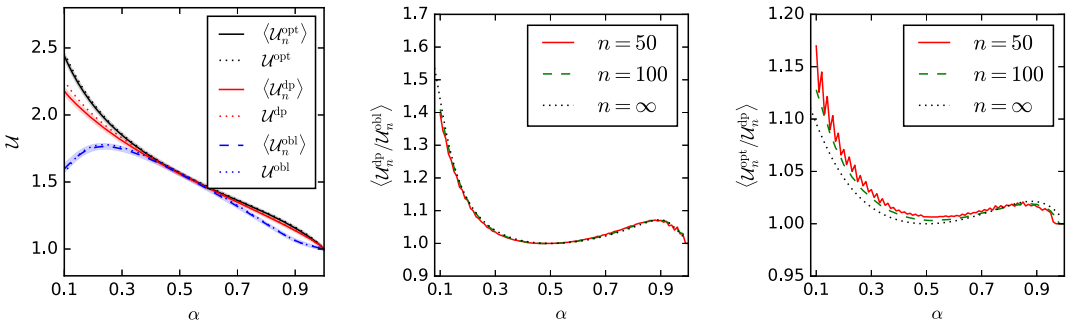
The results for the ratio of $\mathcal{U}^{\text{dp}}/\mathcal{U}^{\text{obl}}$ are given in Fig. 7c. We observe that for both small and large values of α, the demographic parity helps the utility of the group-oblivious algorithm, if the noise values of women evaluation σ_{women} are large, which agrees with the results from Theorem 3. We see that the gain can be up to around 20% if the selection size is small and up to 5% if the selection size is large. For the case where σ_{women} and σ_{men} are close, we observe no gain if the selection is large and we observe a minor loss in utility (around 2%) if the selection is small. This is due to the fact that in the dataset, there are more men with a high true latent quality W, as seen in Fig. 7a. We also plot the region (for $k = 10$) from Theorem 3 in which the utility of the demographic parity algorithm should dominate the utility of the group-oblivious algorithm if the distribution of quality is a group-dependent normal.

Finally, on Fig. 7d we show the ratio $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$ for different values of $k = 1, 4, 7, 10$. In addition to these ratios, we also plot the bound from Theorem 4 for $k = 10$. We see that the bound is quite close to the actual value of $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$ for small α.

5.3. Accuracy of the approximation for small n

As discussed in Section 2, we cannot solve the problem with finite selection sizes exactly. Instead, throughout the paper, we use an approximation that is exact as number of candidates n tends to infinity (Theorem 1). In this section, we question the accuracy of this approximation when the number n of candidates is relatively small. For our experiment, we generate datasets of different sizes $n = 50, 100$. For every size parameter n, we generate 10,000 different datasets. For a population size n, we denote by $\langle \mathcal{U}_n \rangle$ the average quality of the selected candidates over our 10,000 experiments. In each case, the true latent qualities W are generated from a normal distribution $\mathcal{N}(1, 1)$.

In Fig. 8a we plot the average utilities $\langle \mathcal{U}_n \rangle$ for a population of $n = 100$, where we select m individuals and where we vary m from 10 to 100. The shaded region corresponds to a confidence interval. We consider three selection algorithms (demographic parity, group-oblivious and Bayesian-optimal) and compare the performance for $n = 100$ with the limiting quantities $\mathcal{U}^{\text{dp}}$, $\mathcal{U}^{\text{obl}}$ and $\mathcal{U}^{\text{opt}}$. We observe that, even for $n = 100$, the average values of utility are close to the approximation. In Fig. 8b we compare the average ratio of performances $\langle \mathcal{U}_n^{\text{dp}}/\mathcal{U}_n^{\text{obl}} \rangle$ for different n. We observe that the approximation for $n = 50$ is a good prediction of the average gain provided by the use of demographic parity. Similarly, in Fig. 8c, we compare the average ratio of performances $\langle \mathcal{U}_n^{\text{opt}}/\mathcal{U}_n^{\text{dp}} \rangle$ for different n. Again, the curves for finite n are still quite close to the case where $n \to \infty$.



(a) Average utility $\langle \mathcal{U}_n \rangle$, $n = 100$ (b) Ratio $\langle \mathcal{U}_n^{\mathrm{dp}} / \mathcal{U}_n^{\mathrm{obl}} \rangle$ (c) Ratio $\langle \mathcal{U}_n^{\mathrm{opt}} / \mathcal{U}_n^{\mathrm{dp}} \rangle$

Fig. 8. Finite population size: quality of selection and expected gain of the demographic parity over the group-oblivious algorithm. The quality distribution W is $\mathcal{N}(\mu = 1, \eta^2 = 1)$ and the noise parameters are $\sigma_A = 3$, $\sigma_B = 0.2$. The number of experiments per set of parameters is $K = 10,000$. The shaded areas are the confidence intervals (corresponding to one standard deviation on the estimation of the empirical mean).

6. Conclusion

In this work, we study a simple model of the selection problem that captures the phenomenon of differential variance, that is, the decision maker has estimates of the candidates' quality with different variances for different demographic groups. We distinguish two notable cases. In the first case, the decision maker does not have information about the estimate properties (variances and biases); as a result they use a group-oblivious algorithm. In the second case, every information about the distribution of quality is known, and the decision maker is Bayesian-optimal.

First, we show that both baseline algorithms (without any fairness constraint) lead to discrimination. Then we identify conditions under which, in the first case, the γ-rule fairness mechanism (a generalization of the 4/5 rule) leads to a higher selection utility compared to using the group-oblivious baseline. In the second setting, the γ-rule mechanism is harmful to the utility of Bayesian-optimal baseline but we prove that the utility decrease is bounded. Overall, our results contribute to a recent thread of works identifying cases in which, contrary to conventional wisdom, imposing fairness mechanisms does not come at the cost of utility (or even if it does, that the cost is bounded). Beyond fitting a particular application in detail, our results are useful in thinking about the impact of possible policies. For instance, they can help evaluate the effect of imposing a given fairness mechanism, or deciding whether or not to allow access to group information in a particular application.

Our theoretical results are obtained under the assumption that the true latent quality W follows a normal law (to allow for analytical derivations). This assumption can be relaxed: we can plug into the model any distribution of latent quality (e.g., Pareto, uniform, etc.). We show numerically in Section 5 that it does not change the flavor of the main results. Extending these results theoretically is, however, challenging as in our proofs we operate with the expression for the conditional expectation of true latent quality given the noisy estimate. In a non-normal case, this conditional expectation cannot, in general, be expressed in closed form, which complicates the analysis.

Our modeling assumptions imply that a candidate's quality does not depend on the selection strategy used. If attaining a certain level of quality comes at a cost, then the interaction between decision makers and candidates may be seen as a game. It would be interesting to see how differential variance affects the incentives of candidates and how the γ-rule changes them in this game. We leave this game-theoretic formulation of the selection problem as a future direction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003), by the French National Research Agency (ANR) through grant ANR-20-CE23-0007, and by a European Research Council (ERC) Advanced Grant for the project "Foundations for Fair Social Computing" funded under the European Union's Horizon 2020 Framework Programme (grant agreement no. 789373). We thank the editor and the reviewers for their thoughtful comments.

Appendix A. Omitted proofs

In this section we provide detailed proofs of the statements given before. Namely, these are proofs of Theorem 1, Theorem 2, Theorem 3 and Theorem 4.

A.1. Proof of Theorem 1

By our assumptions, the estimates of qualities for G-candidates follow a normal law with the mean $\mu_G - \beta_G$ and the variance $\hat{\sigma}_G^2 = \eta_G^2 + \sigma_G^2$. Recall that the selection rate x_G^{obl} for the group-oblivious algorithm is a probability for the G-candidate to have an estimated quality larger than a predefined group-independent threshold: $x_G^{\text{obl}} = \mathrm{P}(\widehat{W} \geq \hat{\theta}^{\text{obl}} | G)$. Taking all that into account, the selection rate for G-candidates can be expressed as:

$$x_G^{\text{obl}} = 1 - \Phi\left(\frac{\hat{\theta}^{\text{obl}} - \mu_G + \beta_G}{\hat{\sigma}_G}\right).$$

This shows that the condition $x_A^{\text{obl}} > x_B^{\text{obl}}$ is equivalent to $\frac{\hat{\theta}^{\text{obl}} - \mu_A + \beta_A}{\hat{\sigma}_A} < \frac{\hat{\theta}^{\text{obl}} - \mu_B + \beta_B}{\hat{\sigma}_B}$, since Φ is an increasing function of its argument. Hence, by rearranging the terms we conclude that $\hat{\theta}^{\text{obl}} > \frac{\mu_A \hat{\sigma}_B - \mu_B \hat{\sigma}_A}{\hat{\sigma}_B - \hat{\sigma}_A} + \frac{\beta_B \hat{\sigma}_A - \beta_A \hat{\sigma}_B}{\hat{\sigma}_B - \hat{\sigma}_A}$. By substituting the corresponding threshold to the expression for the selection rate x_G^{obl}, we end up with the expression for the values of budgets α for which $x_A^{\text{obl}} > x_B^{\text{obl}}$. The calculations show that for the budgets $\alpha < 1 - \Phi\left(\frac{(\mu_A - \mu_B) - (\beta_A - \beta_B)}{\hat{\sigma}_B - \hat{\sigma}_A}\right) = \Phi\left(\frac{\Delta\mu - \Delta\beta}{\Delta\hat{\sigma}}\right)$ using the group-oblivious algorithm leads to overrepresentation of a high-variance group A, where we use the notation $\Delta\mu = \mu_A - \mu_B$, $\Delta\beta = \beta_A - \beta_B$ and $\Delta\hat{\sigma} = \hat{\sigma}_A - \hat{\sigma}_B$.

A.2. Proof of Theorem 2

The Bayesian-optimal algorithm selects candidates for which the expected quality $\widetilde{W}$ is larger than some group-independent but budget-dependent threshold $\tilde{\theta}$. Since $\widetilde{W}_G$ follows a normal law with the mean μ_G and the variance $\tilde{\sigma}_G^2$, we can write that $x_G^{\text{opt}} = 1 - \Phi\left(\frac{\tilde{\theta} - \mu_G}{\tilde{\sigma}_G}\right)$. In the rest of the proof, without loss of generality we assume that $\tilde{\sigma}_A^2 < \tilde{\sigma}_B^2$, hence, we can calculate that

$$x_A^{\text{opt}} < x_B^{\text{opt}} \iff \frac{\tilde{\theta} - \mu_A}{\tilde{\sigma}_A} > \frac{\tilde{\theta} - \mu_B}{\tilde{\sigma}_B} \iff \tilde{\theta} > \frac{\mu_A \tilde{\sigma}_B - \mu_B \tilde{\sigma}_A}{\tilde{\sigma}_B - \tilde{\sigma}_A}.$$

By substituting the corresponding threshold to the expression for the selection rate x_G^{opt}, we end up with the expression for the values of budgets α for which $x_A^{\text{opt}} < x_B^{\text{opt}}$. The calculations show that this is for all budgets $\alpha < \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)$, where we use the notation $\Delta\mu = \mu_A - \mu_B$ and $\Delta\tilde{\sigma} = \tilde{\sigma}_A - \tilde{\sigma}_B$.

A.3. Properties of the utility $\mathcal{U}$ (Proof of Theorem 3)

In this section, we study the properties of the utility function $\mathcal{U}(x_A)$ independently from the selection algorithm used. We give the expression for the derivative of $\mathcal{U}$ as a function of x_A. This expression allows us to prove that the utility function $\mathcal{U}$ is strictly concave. This implies that as we have $x_A^{\text{obl}} \leq x_A^{\gamma\text{-obl}} \leq x_A^{\gamma\text{-opt}} \leq x_A^{\text{opt}}$ or $x_A^{\text{obl}} \leq x_A^{\gamma\text{-obl}} \leq x_A^{\gamma\text{-opt}} \leq x_A^{\text{opt}}$, one always has $\mathcal{U}(x_A^{\text{obl}}) \leq \mathcal{U}(x_A^{\gamma\text{-obl}}) \leq \mathcal{U}(x_A^{\gamma\text{-opt}}) \leq \mathcal{U}(x_A^{\text{opt}})$, with strict inequalities whenever the above inequalities are strict.

Lemma 2. *Assume that the budget α is fixed.*

1. *The first derivative of the utility $\mathcal{U}(x_A)$ can be expressed as follows:*

$$\mathcal{U}'(x_A) = \frac{p_A}{\alpha}\left[\frac{(\hat{\theta}_A + \beta_A)\eta_A^2 + \mu_A\sigma_A^2}{\eta_A^2 + \sigma_A^2} - \frac{(\hat{\theta}_B + \beta_B)\eta_B^2 + \mu_B\sigma_B^2}{\eta_B^2 + \sigma_B^2}\right] \tag{A.1}$$

   where $\hat{\theta}_A, \hat{\theta}_B$ are such that $\mathrm{P}(\widehat{W} \geq \hat{\theta}_A \mid G = A) = x_A$ and $\sum_{G \in \{A,B\}} \mathrm{P}(\widehat{W} \geq \hat{\theta}_G \mid G) \cdot p_G = \alpha$.
2. *The utility $\mathcal{U}(x_A)$ is strictly concave.*

Proof. By definition of $\mathcal{U}$ in (4), the utility $\mathcal{U}$ equals $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B)$ where $\hat{\theta}_A, \hat{\theta}_B$ are the unique thresholds such that $\mathrm{P}(\widehat{W} \geq \hat{\theta}_A \mid G = A) = x_A$ and $\sum_{G \in \{A,B\}} \mathrm{P}(\widehat{W} \geq \hat{\theta}_G \mid G) \cdot p_G = \alpha$. Using that $\widetilde{W}_G$ and $\widehat{W}_G$ are normally distributed these quantities can be expressed as:

$$\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B) = \frac{1}{\alpha}\sum_G p_G \int_{\hat{\theta}_G}^{\infty} d\hat{w} \int_{-\infty}^{\infty} dw \left[w \cdot \frac{1}{\eta_G}\phi\left(\frac{w - \mu_G}{\eta_G}\right) \cdot \frac{1}{\sigma_G}\phi\left(\frac{\hat{w} - w + \beta_G}{\sigma_G}\right)\right],$$

$$x_G(\hat{\theta}_G) = \int_{\hat{\theta}_G}^{\infty} d\hat{w} \int_{-\infty}^{\infty} dw \left[\frac{1}{\eta_G} \phi\left(\frac{w-\mu_G}{\eta_G}\right) \cdot \frac{1}{\sigma_G} \phi\left(\frac{\hat{w}-w+\beta_G}{\sigma_G}\right) \right].$$

Using the chain rule, we can write the first derivative of selection utility:

$$\frac{d\mathcal{U}}{dx_A} = \sum_G \frac{\partial \mathcal{V}}{\partial \hat{\theta}_G} \frac{d\hat{\theta}_G}{dx_A}. \tag{A.2}$$

From the budget constraint $p_A x_A + p_B x_B = \alpha$, by differentiating both parts by x_A, we obtain that $p_A \frac{dx_A}{dx_A} + p_B \frac{\partial x_B}{\partial \hat{\theta}_B} \frac{d\hat{\theta}_B}{dx_A} = 0$ which implies that $\frac{d\hat{\theta}_B}{dx_A} = -\frac{p_A}{p_B} \frac{\partial \hat{\theta}_B}{\partial x_B}$. Then, by substituting the obtained expression for $\frac{d\hat{\theta}_B}{dx_A}$ into (A.2), we obtain that $\frac{d\mathcal{U}}{dx_A} = p_A \left(\frac{\partial \mathcal{V}}{\partial \hat{\theta}_A} \frac{\partial \hat{\theta}_A}{\partial x_A} - \frac{\partial \mathcal{V}}{\partial \hat{\theta}_B} \frac{\partial \hat{\theta}_B}{\partial x_B} \right)$. From this, the expression (A.1) follows directly.

We observe that the first derivative is linear in the selection thresholds $\hat{\theta}_A$ and $\hat{\theta}_B$. Thus, as the selection rate x_A increases, the derivative $\mathcal{U}'_{x_A}$ decreases which means that the function $\mathcal{U}(x_A)$ is strictly concave. □

A.4. Proof of Theorem 4

Assume that $\alpha < \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)$. By Theorem 2, we have $x_A^{\text{opt}} < x_A^{\text{dp}}$. As we prove in Lemma 2, the utility function $\mathcal{U}$ is concave function of x_A. Using the concavity of $\mathcal{U}$ we have $\frac{\mathcal{U}(x_A^{\text{opt}}) - \mathcal{U}(x_A^{\text{dp}})}{x_A^{\text{opt}} - x_A^{\text{dp}}} \geq \mathcal{U}'(x_A = x_A^{\text{dp}})$ which implies that $\mathcal{U}(x_A^{\text{opt}}) - \mathcal{U}(x_A^{\text{dp}}) \leq (0-\alpha) \cdot \mathcal{U}'(x_A^{\text{dp}})$, where we use the fact that $x_A^{\text{dp}} - x_A^{\text{opt}} \leq \alpha$ for all budgets $\alpha < \Phi\left(\frac{\Delta\mu}{\Delta\tilde{\sigma}}\right)$.

By dividing both sides by $\mathcal{U}(x_A^{\text{dp}})$, from the above inequality we obtain the following upper bound:

$$\frac{\mathcal{U}(x_A^{\text{opt}})}{\mathcal{U}(x_A^{\text{dp}})} \leq 1 - \alpha \cdot \frac{\mathcal{U}'(x_A^{\text{dp}})}{\mathcal{U}(x_A^{\text{dp}})}.$$

The expression for $\mathcal{U}'(x_A^{\text{dp}})$ can be written explicitly by using (A.1) and the fact that group-dependent thresholds for the demographic parity algorithm can be calculated as $\hat{\theta}_G^{\text{dp}} = \mu_G - \beta_G + \hat{\sigma}_G \Phi^{-1}(1-\alpha)$:

$$\mathcal{U}'(x_A^{\text{dp}}) = \frac{p_A}{\alpha} \left(\mu_A - \mu_B + \Phi^{-1}(1-\alpha) \left[\frac{\eta_A^2}{\sqrt{\sigma_A^2 + \eta_A^2}} - \frac{\eta_B^2}{\sqrt{\sigma_B^2 + \eta_B^2}} \right] \right) = \frac{p_A}{\alpha} \left(\Delta\mu + \Phi^{-1}(1-\alpha)\Delta\tilde{\sigma} \right).$$

The utility by the demographic parity algorithm can be calculated using the law of total expectation and the expected value of truncated normal distribution as follows:

$$\mathcal{U}(x_A^{\text{dp}}) = \sum_G p_G \mu_G + \frac{\phi\left(\Phi^{-1}(1-\alpha)\right)}{\alpha} \sum_G p_G \frac{\eta_G^2}{\sqrt{\sigma_G^2 + \eta_G^2}} = \sum_G p_G \mu_G + \frac{\phi\left(\Phi^{-1}(1-\alpha)\right)}{\alpha} \sum_G p_G \tilde{\sigma}_G.$$

Hence, from the above inequality and the expressions for $\mathcal{U}(x_A^{\text{dp}})$ and $\mathcal{U}'(x_A^{\text{dp}})$, we can obtain the following upper bound on the ratio $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$ for $\alpha < \Phi(\Delta\mu/\Delta\tilde{\sigma})$[9]:

$$\frac{\mathcal{U}(x_A^{\text{opt}})}{\mathcal{U}(x_A^{\text{dp}})} \leq 1 - \alpha \cdot \frac{p_A}{\alpha} \frac{\Delta\mu + \Phi^{-1}(1-\alpha)\Delta\tilde{\sigma}}{\sum_G p_G \mu_G + \frac{\phi(\Phi^{-1}(1-\alpha))}{\alpha} \sum_G p_G \tilde{\sigma}_G}.$$

For the γ-fair Bayesian-optimal algorithm, the upper bound on $\mathcal{U}^{\text{opt}}/\mathcal{U}^{\text{dp}}$ can be calculated in a similar manner. The values of the selection rate difference for $\alpha < \Phi(\Delta\mu/\Delta\tilde{\sigma})$ can be upper bounded as $x_A^{\gamma\text{-opt}} - x_A^{\text{opt}} \leq \frac{\alpha}{p_A + p_B/\gamma}$, since the selection by the Bayesian-optimal algorithm lies either inside the γ-region $x_A \in \left[\frac{\alpha}{p_A + p_B/\gamma}, \frac{\alpha}{p_A + \gamma p_B} \right]$ or on its boundary. For $\alpha > \Phi(\Delta\mu/\Delta\tilde{\sigma})$, the difference can be upper bounded as $x_A^{\text{opt}} - x_A^{\gamma\text{-opt}} \leq 1 - \frac{\alpha}{p_A + \gamma p_B}$.

[9] Note that the case $\alpha > \Phi(\Delta\mu/\Delta\tilde{\sigma})$ is proven similarly, except that we use $x_A^{\text{opt}} - x_A^{\text{dp}} \leq 1 - \alpha$.

References

[1] V. Emelianov, N. Gast, K.P. Gummadi, P. Loiseau, On fair selection in the presence of implicit variance, in: Proceedings of the 21st ACM Conference on Economics and Computation, Association for Computing Machinery, 2020, pp. 649–675.
[2] M. Bertrand, S. Mullainathan, Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, Am. Econ. Rev. 94 (4) (2004) 991–1013.
[3] A. Greenwald, L. Krieger, Implicit bias: scientific foundations, Calif. Law Rev. 94 (2006) 945.
[4] B. Collins, Tackling unconscious bias in hiring practices: the plight of the Rooney rule, N.Y. Univ. Law Rev. 82 (2007).
[5] M. Cavicchia, How to fight implicit bias? With conscious thought, diversity expert tells nabe, Am. Bar Assoc., Bar Lead. 40 (1) (2015), https://www.americanbar.org/groups/bar_services/publications/bar_leader/2015-16/september-october/how-fight-implicit-bias-conscious-thought-diversity-expert-tells-nabe/.
[6] C. Passariello, Tech firms borrow football play to increase hiring of women, Wall St. J. (2016), https://www.wsj.com/articles/tech-firms-borrow-football-play-to-increase-hiring-of-women-1474963562.
[7] H. Holzer, D. Neumark, Assessing affirmative action, J. Econ. Lit. 38 (3) (2000) 483–568.
[8] J.M. Kleinberg, M. Raghavan, Selection problems in the presence of implicit bias, in: Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS), 2018, 33.
[9] L.E. Celis, A. Mehrotra, N.K. Vishnoi, Interventions for ranking in the presence of implicit bias, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*), 2020, pp. 369–380.
[10] A. Baye, C. Monseur, Gender differences in variability and extreme scores in an international context, Large Scale Assess. Educ. 4 (2016).
[11] R.E. O'Dea, M. Lagisz, M.D. Jennions, S. Nakagawa, Gender differences in individual variation in academic grades fail to fit expected patterns for stem, Nat. Commun. 9 (1) (2018) 3777.
[12] Isabelle Kocher, seule femme dirigeante du CAC 40, Feb. 5, https://lentreprise.lexpress.fr/actualites/1/actualites/isabelle-kocher-seule-femme-dirigeante-du-cac-40_2117393.html, 2020.
[13] D. Pedreshi, S. Ruggieri, F. Turini, Discrimination-aware data mining, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2008, pp. 560–568.
[14] M. Hardt, E. Price, N. Srebro, Equality of opportunity in supervised learning, in: Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS), 2016, pp. 3323–3331.
[15] M.B. Zafar, I. Valera, M. Gomez Rogriguez, K.P. Gummadi, Fairness constraints: mechanisms for fair classification, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 962–970.
[16] M.B. Zafar, I. Valera, M. Gomez Rodriguez, K.P. Gummadi, Fairness beyond disparate treatment & disparate impact: learning classification without disparate mistreatment, in: Proceedings of the 26th International Conference on World Wide Web (WWW), 2017, pp. 1171–1180.
[17] A. Chouldechova, Fair prediction with disparate impact: a study of bias in recidivism prediction instruments, Big Data 5 (2) (2017) 153–163.
[18] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, A. Huq, Algorithmic decision making and the cost of fairness, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017, pp. 797–806.
[19] Z. Lipton, J. McAuley, A. Chouldechova, Does mitigating ml's impact disparity require treatment disparity?, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), 2018, pp. 8125–8135.
[20] M. Mathioudakis, C. Castillo, G. Barnabo, S. Celis, Affirmative action policies for top-k candidates selection, with an application to the design of policies for university admissions, in: Proceedings of the ACM Symposium on Applied Computing (SAC), 2020, pp. 440–449.
[21] M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, Mitigating bias in algorithmic hiring: evaluating claims and practices, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*), 2020, pp. 469–481.
[22] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork, Learning fair representations, in: Proceedings of the 30th International Conference on Machine Learning (ICML), 2013, pp. 325–333.
[23] F. Locatello, G. Abbati, T. Rainforth, S. Bauer, B. Schölkopf, O. Bachem, On the fairness of disentangled representations, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS), 2019, pp. 14584–14597.
[24] M. Wick, S. Panda, J.-B. Tristan, Unlocking fairness: a trade-off revisited, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS), 2019, pp. 8783–8792.
[25] A. Blum, K. Stangl, Recovering from biased data: can fairness constraints improve accuracy?, in: 1st Symposium on Foundations of Responsible Computing (FORC), 2019.
[26] S. Dutta, D. Wei, H. Yueksel, P.-Y. Chen, S. Liu, K. Varshney, Is there a trade-off between fairness and accuracy? A perspective using mismatched hypothesis testing, in: Proceedings of the 37th International Conference on Machine Learning and Systems (PMLR), 2020, pp. 5067–5077.
[27] E. Phelps, The statistical theory of racism and sexism, Am. Econ. Rev. 62 (4) (1972) 659–661.
[28] N. Garg, H. Li, F. Monachou, Standardized tests and affirmative action: the role of bias and variance, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT'21, Association for Computing Machinery, 2021, p. 261.
[29] S. Coate, G. Loury, Will affirmative-action policies eliminate negative stereotypes?, Am. Econ. Rev. 83 (1993) 1220–1240.
[30] L. Balafoutas, M. Sutter, Affirmative action policies promote women and do not harm efficiency in the laboratory, Science 335 (2012) 579–582.
[31] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[32] IIT-JEE dataset, https://github.com/AnayMehrotra/Ranking-with-Implicit-Bias. (Accessed 29 January 2020), 2019, Online.