On Fair Selection in the Presence of Implicit Variance

VITALII EMELIANOV, Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, France

NICOLAS GAST, Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, France

KRISHNA P. GUMMADI, Max Planck Insitute for Software Systems, Germany

PATRICK LOISEAU, Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, France

Quota-based fairness mechanisms like the so-called Rooney rule or four-fifths rule are used in selection problems such as hiring or college admission to reduce inequalities based on sensitive demographic attributes (gender, ethnicity, etc.). These mechanisms are often viewed as introducing a trade-off between selection fairness and utility (i.e., the overall quality of the selected candidates). In recent work, however, Kleinberg and Raghavan [*Proc. of ITCS '18*] showed that, in the presence of implicit bias in estimating candidates' quality, the Rooney rule can in fact increase the utility of the selection process (beyond improving its fairness).

We argue that even in the absence of implicit bias, the estimates of candidates' quality from different groups may differ in another fundamental way, namely, in their variance. We term this phenomenon *implicit variance* and we ask: can fairness mechanisms be beneficial to the utility of a selection process in the presence of implicit variance (even in the absence of implicit bias)? To answer this question, we propose a simple model in which candidates have a true latent quality that is drawn from a group-independent normal distribution. To make the selection, a decision maker receives an unbiased estimate of the quality of each candidate, with normal noise, but whose variance depends on the candidate's group. We then compare the utility obtained by imposing a fairness mechanism that we term γ-rule, which includes demographic parity ($\gamma = 1$) and the four-fifths rule ($\gamma = 0.8$) as special cases, to that of a group-oblivious baseline selection algorithm that simply picks the candidates with the highest estimated quality independently of their group. Our main result shows that the demographic parity mechanism always strictly increases the selection utility, while any other γ-rule also always increases it weakly. We extend our model to a two-stage selection process where the true quality is observed at the second stage and analyze how our results are changed in that case. We finally discuss multiple extensions of our results, in particular to different distributions of the true latent quality.

CCS Concepts: • **Theory of computation** → *Machine learning theory*; *Models of learning*.

Additional Key Words and Phrases: selection problem, fairness, implicit bias, implicit variance

1 INTRODUCTION

Discrimination in selection and the role of implicit bias. Many selection problems such as hiring or college admission are subject to discrimination [5], 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 [15]. While human decision makers are naturally susceptible to 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* [11] 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 [6, 24]. Another widely used fairness mechanism is the so-called $^4/_5$-*rule* [17], that 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 [17].

Fairness mechanisms, however, have been the subject of frequent debates. On 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 [19] (see also an extension to ranking problems in [7]). The authors of [19] 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 $\hat{W}_i = W_i$ for candidates from the well-represented group and $\hat{W}_i = W_i/\beta$ 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 $\hat{W}_i$. Then Kleinberg and Raghavan [19] 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 implicit 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'

---

[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 implicit 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.

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 [4, 23]. The second arises from the *decision makers*: decision makers might have different levels of experience[3] 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). 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 term the above issue with quality estimates '***implicit variance***' as decision makers are often unaware of their group-dependent variances. We posit that implicit 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, while estimating these variances can be a difficult task (and variance is indeed ignored in many algorithms). In this paper, we model the implicit variance phenomenon by assuming that the decision maker sees of an estimate of the quality of a candidate $\hat{W}_i$ that is equal to the candidate's true latent quality W_i plus an additive noise whose variance depends on the group of the candidate.[4] In this situation, a natural baseline decision maker to consider is the *group oblivious* selection algorithm that simply selects the candidates with the highest estimated quality, irrespective of their group, to maximize the selection utility. The group oblivious selection algorithm represents not only a decision maker unaware of the implicit variance in their estimates, but also a decision maker determined to *not use* group information.[5]

Unfortunately, our analysis shows that in the presence of implicit variance, group oblivious selection can lead to underrepresentation of groups with lower-variance quality estimates compared to groups with higher-variance quality estimates. One natural way to address this representation inequality would be 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 [19] in the case of implicit bias: *what is the effect of fairness mechanisms on the quality of a selection in the presence of implicit variance?*

Our model and overview of our results. To answer this question, we propose a simple model of implicit variance with two groups A and B: for each candidate i, the decision maker gets a quality estimate $\hat{W}_i = W_i + \sigma_{G_i} \varepsilon_i$, where G_i is the group to which the candidate belongs and ε_i is a standard normal random variable. The estimator is unbiased but has a variance $\sigma_{G_i}^2$ that depends on the candidate's group. We assume that the true quality comes from a group-independent distribution— assumed normal in our analytical results. In the one-stage selection problem, the decision maker then selects a fraction α_1 (called selection budget) of the candidates.

---

[3]Or different amounts of data in case of algorithmic decision making.

[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]If a decision maker knows the group-dependent variances, then they could use the variances together with group information of the candidates to optimize utility, see the Bayesian-optimal algorithm below.

Using this model, we first observe that, for any selection budget $\alpha_1 \neq 1/2$, the group oblivious selection algorithm (our baseline without fairness mechanism) leads to a smaller selection rate—i.e., to underrepresentation—for one of the two groups: the low-variance group if $\alpha_1 < 1/2$ (the most common case) and the high-variance group if $\alpha_1 > 1/2$. Then, we investigate how the utility of the group oblivious baseline is 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.

Our main result shows that for the one-stage selection problem with any selection budget, beyond giving a more fair representation, demographic parity strictly improves the selection utility (measured by the expected quality of a selected candidate). Moreover, the γ-rule with any $\gamma < 1$ yields a utility strictly lower than demographic parity but still weakly higher than the group-oblivious algorithm. We then consider a two-stage selection process. There, a pre-selection is first made based on the quality estimates. Then the true quality is observed for each pre-selected candidate and the selection is refined to meet a lower second-stage budget α_2. This can model hiring decisions where one first makes a short list based on CV and then refines the selection after interview; or grant selection processes that often happen in two stages with only an extended abstract at first stage and then a full proposal at the second stage. In this two-stage selection process, we show that demographic parity strictly improves the selection quality if the first-stage budget is close enough to the second-stage budget (i.e., one does not interview too many more candidates than slots available) or if the first-stage budget is large; but that it may hurt in between—although much less than the gain in the other regimes. As above, the same holds (but weakly) for the γ-rule with other values of γ. Finally, through numerical simulations, we show that our analytical results can be extended in particular to cases where the latent quality distribution is not normal.

Overall, our results show that fairness mechanisms can increase utility in selection problems with implicit variance and freed of any bias. In practical scenarios, one may (at least currently) encounter both implicit bias and implicit variance in decision making, and we do not claim that our main takeaway generalizes to those situations. Similarly, if the baseline decision maker is not group oblivious, fairness mechanisms may not increase utility. Finally we remark that in selection problems with implicit variance, the key characteristic of a group is high- *vs* low-variance, not minority *vs* majority. In such problems, the group oblivious algorithm often overrepresents the high-variance group. If the high-variance group corresponds to a minority of candidates, this may seem counter-intuitive. We stress however that (*i*) this corresponds to a group oblivious baseline without bias (which might not match certain practical scenarios); and (*ii*) our model *does not require* that the higher-variance group corresponds to a minority group. We argue that the opposite case is equally interesting in practice (in particular when the implicit variance arises from the candidates).

Related works. There is an abundant literature on fairness in machine learning, in particular on (one-stage) classification, that tackles the question of how to learn a classifier while enforcing some fairness notion in the outcome [8, 12, 16, 20, 22, 25, 28, 29]. 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 [19] under the presence of implicit bias [15]. In their work, the authors study the Rooney rule [11] 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 [7], 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 but, in contrast, we do not consider that there is implicit bias and introduce instead the notion of implicit variance to capture the difference in precision of the quality estimate for different groups. Although our model can easily be extended to incorporate implicit bias as well, we purposely restrict it to the simplest possible form of implicit variance so as to show its effect on the selection problem independently of bias. 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 [27] (e.g., by learning fair representations of data [21, 30]).

In our work, we also consider the $^4/_5$-rule [17] (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 or its extension in [7] imposes a constraint on the *number* of selected candidates from the underrepresented group.

The two aforementioned papers [7, 19] essentially analyze one-stage selection, whereby the utility is the sum of the utilities of selected candidates. In practice, many selection problems are done in two stages where the first-stage selection is refined in a second stage with access to finer information. A few recent papers specifically analyze the two-stage setting. In [18], a two-stage college admission and hiring procedure is considered. The authors study if certain fair policies—irrelevance of group membership and equal opportunity—can be satisfied. They show that it is possible to satisfy both if the college grades are not reported to the employer, but there are settings where these fairness conditions cannot be satisfied even in isolation. In [13], the authors study an optimal multistage selection. They propose two fairness notions for the multistage setting: local (per stage) and global (final stage) fairness and study their trade-off, the price of local fairness. They show that this price is bounded and is dependent on the timing of the sensitive feature revelation (e.g., gender): the later the sensitive feature is given, the lower the price is. The model that the authors consider assumes an optimal selection procedure that requires knowledge of the data generation procedure. In contrast, in our work, we study selection procedures that do not require knowledge of the data distribution and analyze when imposing a fairness mechanism at the first stage leads to an improvement of the utility of the selection after the second stage.

Fairness mechanisms has been a subject of a number of studies in the economic literature, in particular from empirical data. In [10], the authors study whether affirmative actions can remove stereotypes about a particular population. In [3], 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 implicit variance. The model of observed quality in this paper is similar in spirit to the model of statistical discrimination in [2, 26], but in contrast to those works we assume that the decision maker does not know the distribution parameters and the discrimination happens for a different reason.

2 THE MODEL

We consider the following scenario. A decision maker is given n candidates, out of which a subset is selected. Each candidate $i \in \{1, \cdots, n\}$ is endowed with a true latent quality W_i. The qualities W_i are drawn *i.i.d.* from an underlying probability distribution that is group-independent but unknown to the decision maker. For our analytical results and unless otherwise explicitly specified, we assume
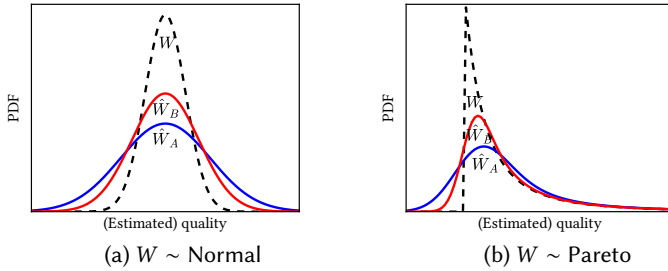
Fig. 1. Probability density function of the true latent quality W and the estimated quality $\hat{W}$.

that this distribution is a normal distribution of mean μ_W and variance $\sigma_W^2 > 0$. The goal of the decision maker is to maximize the expected quality of the selected candidates: $E\left[\sum_{i \in \text{selection}} W_i\right]$.

2.1 The implicit variance model

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. When making the selection decision, the decision maker has access to an unbiased estimator of the true quality. We denote the estimator of the quality of candidate i by $\hat{W}_i$. We assume that the variance of the estimator depends on the group: for a candidate i that belongs to group $G_i \in \{A, B\}$, its estimated quality is

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

where ε_i is a centered random variable from Normal$(0, 1)$—the standard normal distribution, of mean 0 and variance 1. The variables ε_i are assumed *i.i.d.*.

Without loss of generality, in the rest of the paper we assume that $\sigma_A^2 > \sigma_B^2$, that is that the quality estimate has higher variance for group A. 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, however, 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 *implicit variance* because we assume that the decision maker does not know the variance of the estimators: it is an unconscious phenomenon. Also, note that a different estimator, having access to different data, will have different implicit variances for the two 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 $\hat{W}_A$ a variable that has the same distribution as $W_i + \sigma_A \varepsilon_i$ and similarly for B).

2.2 One-stage and two-stage selection problems

Candidates are selected in a one-stage or two-stage process. In the first stage, for each of the candidate i, the decision maker observes the quality estimate $\hat{W}_i$ as well as its group $G_i \in \{A, B\}$. They then select m_1 candidates out of those n. The selection process stops here if the selection has only one stage. If the process has two stages, the decision maker observes the true quality W_i of the m_1 candidates that were selected at first stage. They select the m_2 candidates having the largest W_i among those m_1 candidates. We remark that when $m_1 = m_2$, the two-stage selection process reduces to the one-stage selection process.

2.3 Selection algorithms

The goal of the decision maker is to maximize the expected quality of the selected candidates. Since the true quality W_i is observed at second stage, the natural selection at second stage is to sort the candidates according to their quality W_i and to select the best m_2. For the first stage, the situation is different because the decision maker only has access to an estimator of the quality $\hat{W}_i$ whose variance depends on the candidate's group. In this paper, we compare the following first-stage selection algorithms to show that fairness mechanisms can improve the selection quality.

Group Oblivious. As $\hat{W}_i$ is an unbiased estimator of W_i whose variance is unknown to the decision maker, the most natural first-stage selection rule is to sort the candidates according to $\hat{W}_i$ irrespective of their group and to keep the best m_1. We call this the *group oblivious* selection. This selection algorithm might be seen as a fair treatment because the selection does not use the group label. Yet, because of our implicit variance model, this might lead to discriminate against high-variance or low-variance groups. We will discuss that in Lemma 3.1.

$^4/_5$-*rule and γ-rule.* Let us denote by x_A (and x_B) the fraction of the A-candidates (and B-candidates) that are selected at first stage. The group oblivious selection might favor one group or the other, that is $x_A \gg x_B$ or $x_B \gg x_A$. To mitigate the inequality, the decision maker can introduce selection quotas. One example is the well studied Rooney rule [19] that states that at least one candidate from each group should be chosen. Another example is the four-fifths rule [17] 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 this rule that is parametrized by $\gamma \in [0, 1]$. We say that a selection is γ-fair if

$$x_A \geq \gamma x_B \qquad \text{and} \qquad x_B \geq \gamma x_A. \tag{2}$$

In practice, this is easily done by sorting A- and B-candidates separately. Indeed, by using that $x_A \geq \gamma x_B$ and $x_B \geq \gamma x_A$, the total number of candidates selected at first stage satisfies

$$m_1 = x_A n_A + x_B n_B \leq x_A n_A + \frac{x_A}{\gamma} n_B = x_A n_A \frac{n_A \gamma + n_B}{n_A \gamma}, \quad \text{and} \quad m_1 \leq x_B n_B \frac{n_B \gamma + n_A}{n_B \gamma}.$$

This means that to satisfy (2), the selection algorithm picks the best estimated $\lceil m_1 \gamma n_A / (n_B + \gamma n_A) \rceil$ A-candidates and the best estimated $\lceil m_1 \gamma n_B / (n_A + \gamma n_B) \rceil$ B-candidates. Then, the remaining positions are filled with the best estimated among the remaining candidates, irrespective of their group. Note that when $\gamma = 0$, the γ-fair algorithm reduces to the group oblivious algorithm.

Demographic parity. When $\gamma = 1$, the γ-fair algorithm corresponds to the classical notion of *demographic parity* [28] that mandates that the selection rates be equal across different groups. Note that because n_A, n_B and m_1 are integer variables, it might be impossible to satisfy the constraints in (2) when γ is too close to 1. In such a case, we say that an algorithm is γ-fair if the constraint (2) is satisfied up to one candidate.

Bayesian optimal. We will compare the performance of the above selection algorithms with the performance of what we call a *Bayesian-optimal* selection algorithm. This algorithm is an idealized selection algorithm that knows all the parameters of the problem (the quality distribution and the variances σ_A^2 and σ_B^2) and chooses the candidates in order to maximize the expected quality at final stage. Recall that our model assumes that the decision maker does not know the quality distribution W nor the variances of the estimator σ_A^2 and σ_B^2. Hence, this Bayesian-optimal algorithm is not implementable in practice; we will use it as an upper bound of what could be achieved.

Note that while the previous algorithms are oblivious to what happens in the second stage, the Bayesian-optimal selection does depend on the fraction of candidates selected at second stage, α_2.

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 $0 \leq p_A \leq 1$ and $0 < \alpha_2 \leq \alpha_1 \leq 1$ such that

$$n_A = \lfloor p_A n \rfloor \qquad m_1 = \lfloor \alpha_1 n \rfloor \qquad m_2 = \lfloor \alpha_2 n \rfloor,$$

and let n grow. Our theoretical results are obtained in the limit where n goes to infinity (similarly to [7, 19]). In Section 5.3 we will 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_1 and α_2 represent the global selection ratios (or budgets) at first and second stage respectively.

As we prove below, characterizing the performance of a selection problem is simpler when the number of candidates n is infinite because a selection algorithm is characterized by three selection thresholds $\hat{\theta}_A$, $\hat{\theta}_B$ and θ as follows. The thresholds $\hat{\theta}_G$'s correspond to the first stage selection and the threshold θ corresponds to the second stage selection: if i is a G_i-candidate, they will be selected at first stage if $\hat{W}_i \geq \hat{\theta}_{G_i}$. They will pass both stages if $\hat{W}_i \geq \hat{\theta}_{G_i}$ and $W_i \geq \theta$. For given thresholds $\hat{\theta}_A$, $\hat{\theta}_B$ and θ, we denote the expected utility of the corresponding selection by $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, \theta)$:

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

For these given thresholds $\hat{\theta}_A, \hat{\theta}_B, \theta$, the fractions of selected candidates are $\mathrm{P}(\hat{W}_i \geq \hat{\theta}_{G_i})$ after the first stage, and $\mathrm{P}(W_i \geq \theta, \hat{W}_i \geq \hat{\theta}_{G_i})$ after the second stage. 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 at the first stage and that satisfies the selection size constraints in expectation:

$$\mathcal{U}(x_A) = \mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, \theta), \text{ where } \hat{\theta}_A, \hat{\theta}_B, \theta \text{ are such that} \begin{cases} \mathrm{P}(\hat{W}_i \geq \hat{\theta}_A \mid G_i = A) = x_A, \\ \mathrm{P}(\hat{W}_i \geq \hat{\theta}_{G_i}) = \alpha_1, \\ \mathrm{P}(\hat{W}_i \geq \hat{\theta}_{G_i}, W_i \geq \theta) = \alpha_2. \end{cases} \tag{3}$$

Note that combining the first two constraints in (3) immediately gives that such an algorithm selects B-candidates with probability $x_B = (\alpha_1 - x_A p_a)/(1 - p_A)$. 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 2.3 because those algorithms are defined neither in terms of fraction of selected candidate 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 ($\hat{W}_i$). 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. The following proposition shows that when the population is large, these random fluctuations disappear. It shows that, when n is large, the performance of the various algorithms are simply characterized by x_A.

PROPOSITION 2.1. *For any problem parameters and any of the first stage selection algorithms presented in Section 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 $\hat{\theta}_A, \hat{\theta}_B, \theta$ such that the expected utility of this algorithm converges to $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, \theta)$.*

SKETCH OF PROOF. 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 $\hat{W}_G$ as $n \to \infty$. 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 $\hat{W}_A$ is asymptotically equivalent as $n \to \infty$. The same argument can be used to show that second stage selection is asymptotically equivalent to selecting all candidates above a given threshold. □

In what follows, we will study directly the model when the population n is large. We denote by respectively x_A^{obl}, $x_A^{\gamma\text{-rule}}$, x_A^{dp}, and x_A^{opt} the asymptotic fraction of A-candidates that are selected at first stage for the group oblivious, the γ-fair, the demographic parity and the Bayesian-optimal algorithms. Moreover, we will denote the expected performance of the various algorithms by

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

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. Proposition 2.1 allows us to greatly simplify the study of the performance of the various heuristics because the function $\mathcal{U}$, defined in Equation (3), depends only on one parameter x_A, and is simpler to characterize than the expectation over a finite number of candidates n.

Summary of the notation. To simplify the exposition, and since candidates are interchangeable, in the remainder of the paper, we will omit the subscript i and write directly W and $\hat{W}$ for the quality and the estimated quality of a given candidate. Our notation is summarized in Table 1.

Table 1. Summary of notation.

p_G	fraction of G-candidates, n_G/n, for $G \in \{A, B\}$
σ_W^2	variance of latent quality W
σ_G^2	implicit variance of estimated quality $\hat{W}$ given group $G \in \{A, B\}$
x_G	fraction of G-candidates that are selected at first stage
$\hat{\theta}_G$	threshold for G-candidates at first stage
θ	threshold at second stage
α_1, α_2	fractions of candidates selected at stages 1 and 2 (or budgets)
$\mathcal{U}$	expected selection quality
$\phi, \Phi, \Phi^c, \Phi^{-1}$	pdf, cdf, complementary cdf and quantile of Normal$(0, 1)$

3 ONE-STAGE SELECTION

We start our study with the simplest case: one-stage selection. In this setting $\alpha_2 = \alpha_1$, which means that the decision maker observes the values of $\hat{W}$ and makes a selection that is final, i.e., no further subselection is performed after observing the exact values of W. This type of selection is the most commonly studied in the related work, see for instance [7, 19].

In this section, we compare the first-stage selection algorithms that we introduced in Section 2.3. The main result of this section is that using any γ-fair algorithm increases the performance compared to using the group-oblivious algorithm. To show this, we start by describing key properties of the one-stage algorithms in Section 3.1. The main result is then stated and proven in Section 3.2.

3.1 Behavior of the different first-stage selection algorithms

Group oblivious. This algorithm sorts the candidates according to the estimates $\hat{W}_i$'s and selects the candidates having the highest estimates, regardless of their group. The decision maker does not distinguish between A- and B-candidates and treats them equally. This corresponds to applying

the same threshold for the two groups, i.e., $\hat{\theta}_A = \hat{\theta}_B$. Due to different variances of estimation, this might lead to selecting more people from one group or the other.

Recall that we assume that $\sigma_A > \sigma_B$. This means that the distribution of $\hat{W}$ has more extremes for an A-candidate. Thus, if the selection size is small, more A-candidates will be selected compared to B-candidates because the probability to estimate an A-candidate as a "genius" is higher than for B-candidates. In contrast, if the selection size is large, the chance of estimating an A-candidate as bad is larger than for B-candidates, in which case the decision maker selects a lower fraction of A-candidates. This can be formally stated as follows.

LEMMA 3.1 (GROUP OBLIVIOUS SELECTION). *When using the group oblivious selection algorithm, the fractions x_G^{obl} of selected candidates from each group satisfy:*

(1) if $\alpha_1 < 1/2$, then $x_A^{\text{obl}} > x_B^{\text{obl}}$;
(2) if $\alpha_1 > 1/2$, then $x_A^{\text{obl}} < x_B^{\text{obl}}$;
(3) if $\alpha_1 = 1/2$, then $x_A^{\text{obl}} = x_B^{\text{obl}} = 1/2$.

PROOF SKETCH. The selection fraction x_G is by definition the probability that a G-candidate has a value of $\hat{W}$ larger than $\hat{\theta}_G$. As $\sigma_A > \sigma_B$, one has $P(\hat{W} \geq \mu_W + x|A) > P(\hat{W} \geq \mu_W + x|B)$ and $P(\hat{W} \geq \mu_W - x|A) < P(\hat{W} \geq \mu_W - x|B)$ for any $x > 0$. This implies that $x_A^{\text{obl}} > x_B^{\text{obl}}$ if $\alpha_1 < 1/2$ and $x_A^{\text{obl}} < x_B^{\text{obl}}$ if $\alpha_1 > 1/2$. Since the distribution of quality W is symmetric, then one has $x_A^{\text{obl}} = x_B^{\text{obl}} = 1/2$ when $\alpha_1 = 1/2$. The proof is detailed in Appendix B.1. □

γ-rule. An algorithm that satisfies the γ-rule is a variant of the group-oblivious algorithm that guarantees a minimum selection rate for each group. A straightforward computation shows that an algorithm satisfies both the γ-rule conditions of (2) and the size constraint $x_A p_A + x_B(1 - p_A) = \alpha_1$ if and only if $x_A \in [\frac{\alpha_1}{p_A + p_B/\gamma}, \frac{\alpha_1}{p_A + p_B\gamma}]$. By our definition of Section 2.3, the γ-rule selection algorithm is the algorithm that is the closest to the group-oblivious algorithm while respecting the γ-rule. Thus, the fraction of A-candidates selected at first stage of the γ-rule selection algorithm is

$$x_A^{\gamma\text{-rule}} = \min \left(\frac{\alpha_1}{p_A + p_B\gamma}, \max \left(x_A^{\text{obl}}, \frac{\alpha_1}{p_A + p_B/\gamma} \right) \right).$$

Demographic parity. Contrary to the group oblivious selection algorithm, this algorithm applies different thresholds for the two groups to preserve the demographics. Since the selection size is fixed to α_1 and both selection fractions are equal, then $x_A^{\text{dp}} = x_B^{\text{dp}} = \alpha_1$. Thus, if the selection size α_1 is small, the demographic parity selection algorithm increases the threshold $\hat{\theta}_A$ for A-candidates by removing extremes and filling other places by B-candidates. If the selection size is large, it lowers the threshold for A-candidates by supporting the ones that were removed from the selection.

Formally, using the properties of normal distributions, we can write that for a fixed α_1, a G-candidate with estimate $\hat{W}$ is selected if $\hat{W} \geq \sqrt{\sigma_W^2 + \sigma_G^2}\Phi^{-1}(1 - \alpha_1) + \mu_W$. Recall that in our regime with $n = \infty$, the thresholds $\hat{\theta}_A$ and $\hat{\theta}_B$ are the estimates of quality of the worst A- and B-candidates that are selected. Then, with the demographic parity selection algorithm, we have $(\hat{\theta}_A - \mu_W)/(\hat{\theta}_B - \mu_W) = \sqrt{\sigma_A^2 + \sigma_W^2}/\sqrt{\sigma_B^2 + \sigma_W^2}$.

Bayesian-optimal selection. Since $(W, \hat{W})$ is a bivariate normal, then using the property of conditional expectation, the expected quality of candidate given its estimate is

$$\mathrm{E}(W|\hat{W} = \hat{w}) = \frac{\sigma_W^2}{\sigma_G^2 + \sigma_W^2}\hat{w} + \left(1 - \frac{\sigma_W^2}{\sigma_G^2 + \sigma_W^2} \right)\mu_W. \tag{4}$$

The above expression corresponds to the expectation of the posterior distribution of W given $\hat{W}$. This expectation is $\hat{W}$ if $\sigma_G^2 = 0$ (i.e., there is no noise) and converges to μ_W when $\sigma_G^2 \to \infty$.

In one-stage selection, the Bayesian-optimal selection algorithm picks the candidates with the highest posterior quality expectation (4). This means that the thresholds $\hat{\theta}_G$, which are the estimations of quality of the worst selected A- and B-candidates, satisfy $(\hat{\theta}_A - \mu_W)/(\hat{\theta}_B - \mu_W) = (\sigma_A^2 + \sigma_W^2)/(\sigma_B^2 + \sigma_W^2)$. This shows that when the number of candidates to be selected is small ($\alpha_1 < 1/2$), the Bayesian-optimal selection is even more conservative than the demographic parity selection: it imposes more strict constraints on the high variance group A if selection is for candidates having estimated qualities higher than the median. We state the results in the following lemma.

LEMMA 3.2 (BAYESIAN-OPTIMAL SELECTION). *When using the Bayesian-optimal one-stage selection algorithm, the fraction of A- and B-candidates selected, x_A^{opt} and x_B^{opt}, satisfy*

(1) *if $\alpha_1 < 1/2$, then $x_A^{\text{opt}} < x_B^{\text{opt}}$;*
(2) *if $\alpha_1 > 1/2$, then $x_A^{\text{opt}} > x_B^{\text{opt}}$;*
(3) *if $\alpha_1 = 1/2$, then $x_A^{\text{opt}} = x_B^{\text{opt}} = 1/2$.*

PROOF SKETCH. To study the Bayesian-optimal selection, we express the utility $\mathcal{U}$ as a function of x_A. The key difficulty is to compute the first and second derivatives of $\mathcal{U}$ with respect to x_A. Then, using Harris inequality [14], we show that $\mathcal{U}$ is strictly concave. Hence the root of the equation $\frac{d\mathcal{U}}{dx_A} = 0$ gives the optimal threshold $\hat{\theta}_G$. The expression for the thresholds allows us to compare the values of x_A^{opt} and x_B^{opt} depending on selection size α_1. The full proof is given in Appendix B.2. □

The behavior of these algorithms is illustrated in Fig. 2. We plot the result for a small selection size ($\alpha_1 < 0.5$); the situation for $\alpha_1 > 0.5$ is symmetric. We observe that when the selection size is small, the group oblivious algorithm selects many more A-candidates than B-candidates. In order to select the same fraction of candidates from both groups, the demographic parity selection algorithm uses a higher threshold for the high-variance candidates. The γ-rule selection algorithm (for $\gamma = 0.8$, which corresponds to the $^4/_5$ rule) is between the group oblivious and demographic parity selection algorithms. We also observe that the Bayesian-optimal selection algorithm is even more conservative and selects fewer A-candidates than the other algorithms.
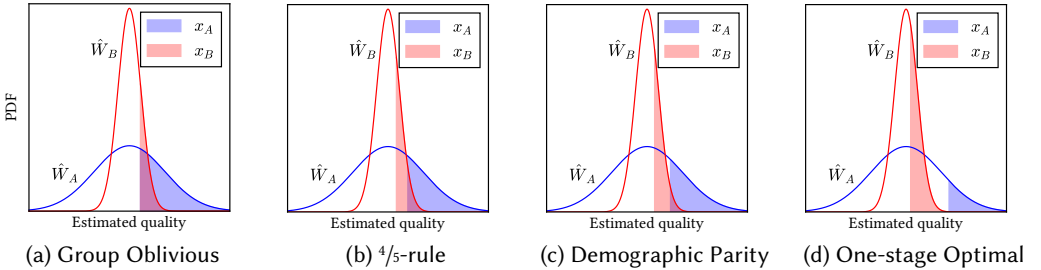

Fig. 2. Illustration of the fraction of selected candidates at first-stage ($\alpha_1 < 0.5$).

In Fig. 3, we plot the expected utility $\mathcal{U}(x_A)$ as a function of the fraction of selected A-candidates. We compare different values of selection sizes $\alpha_1 \in \{0.15, 0.35, 0.6, 0.8\}$. As stated in the proof of Lemma 3.2, this function is concave and its maximum is attained in x_A^{opt}. In this figure, we also plot the γ-fair regions for $\gamma = 0.8$ ($^4/_5$ rule). To satisfy the γ-rule (2), the fraction x_A^{obl} selected by the group-oblivious selection algorithm should suffer a correction such that the corresponding selection fraction will lie on the boundary of the γ-region, if it is outside of it.
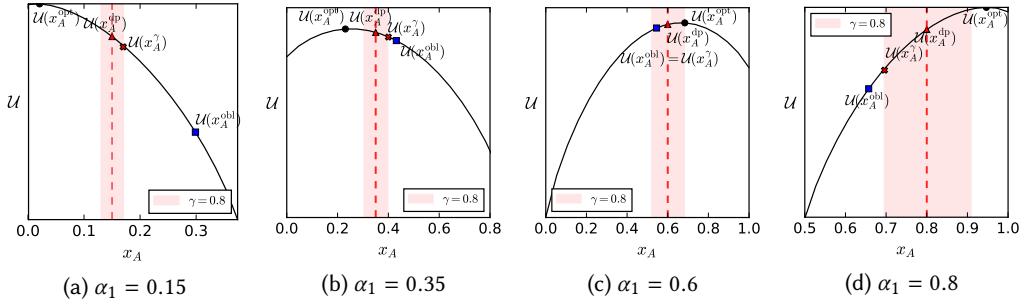
Fig. 3. **One-stage selection**: Utility $\mathcal{U}(x_A)$ as a function of selection probability x_A. The function is concave and attains its maximum in x_A^{opt}. The parameters are $\mu_W = 1$, $\sigma_W = 1$, $p_A = 0.4$, $\sigma_A = 3$, and $\sigma_B = 0.2$.

3.2 Fairness mechanisms improve selection quality

In our work we ask the following question: can fairness mechanisms be beneficial to the utility of a selection process? A positive answer is given in the following theorem where we show that, for one-stage selection, demographic parity always increases the average quality of a selection compared to the group oblivious algorithm. Our result also shows that the softer γ-rule lies between the group oblivious and demographic parity selection algorithms.

THEOREM 1 (FAIRNESS MECHANISMS IMPROVES SELECTION UTILITY). *For the one-stage selection problem and for any $\alpha_1 \neq 1/2$, the demographic parity selection algorithm provides a larger utility than a γ-rule selection algorithm with $\gamma < 1$, which in turns provides a larger utility than the group-oblivious selection algorithm:*

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

The above inequality is an equality when $\alpha_1 = 1/2$.

PROOF. Fig. 3 is a good illustration of this proof. As the selection size is equal to α_1, we have $x_A p_A + x_B p_B = \alpha_1$. By using the results in Lemma 3.1 and Lemma 3.2, this implies that:

- When $\alpha_1 < 1/2$, we have $x_A^{\text{opt}} < x_A^{\text{dp}} = \alpha_1 < x_A^{\gamma\text{-rule}} \leq x_A^{\text{obl}}$. We observe that in Fig. 3ab.
- When $\alpha_1 > 1/2$, we have $x_A^{\text{opt}} > x_A^{\text{dp}} = \alpha_1 > x_A^{\gamma\text{-rule}} \geq x_A^{\text{obl}}$. We observe that in Fig. 3cd.

The results then follow from the concavity of $\mathcal{U}$ proven earlier. □

Demographic parity helps the selection utility by reducing the effect of implicit variance, but it is also interesting to see how large this performance gap can be. In Fig. 4, we show the obtained utilities $\mathcal{U}$, the selection fractions x_A and the gap values $(\mathcal{U}^{\text{dp}} - \mathcal{U}^{\text{obl}})/\mathcal{U}^{\text{obl}}$ for different budgets α_1 from 0.01 to 0.99. Fig. 4a illustrates the utilities corresponding to different selection algorithms. We observe that demographic parity outperforms group oblivious selection, which corresponds to the result of Theorem 1. We also observe that the utilities of the Bayesian-optimal and demographic parity selections decrease with α_1. This is expected because this graph represents the average quality of a candidate: the average quality decreases with the number of selected candidates. 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_1 goes from 0.1 to 0.3. In fact, when $\alpha_1 < 0.1$, very few B-candidates are selected by the group oblivious algorithm. When $\alpha_1 > 0.1$, this algorithm selects a few good B-candidates which leads to an increased average performance.

In Fig. 4c we show the performance gap between group oblivious and demographic parity selection algorithms for different values of σ_A and fixed $\sigma_B = 0.2$, $\sigma_W = 1$. The values of σ_A are
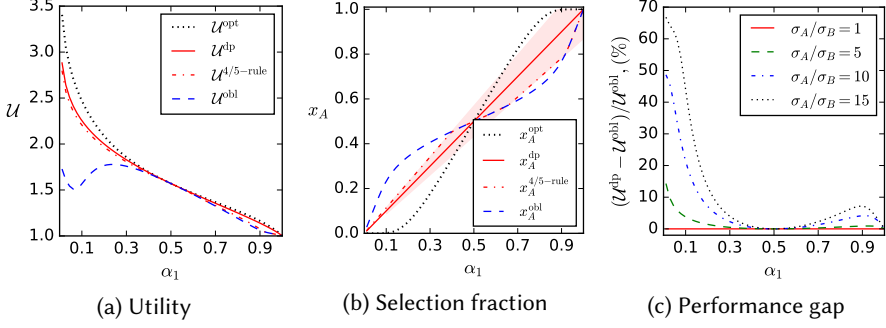
Fig. 4. **One-stage selection**: Utility $\mathcal{U}$, selection fraction x_A and performance gap $(\mathcal{U}^{\text{dp}} - \mathcal{U}^{\text{obl}})/\mathcal{U}^{\text{obl}}$ for different budgets α_1. The parameters are $\mu_W = 1$, $\sigma_W = 1$, $\sigma_B = 0.2$, and $p_A = 0.4$; $\sigma_A = 3$ for panels (a,b).

such that $\sigma_A/\sigma_B = k$, $k = 1, 5, 10, 15$. We see that the gap is in general larger when the selection size α_1 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 $\alpha_1 = 0.5$ 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_1. In addition, the larger the implicit variance ratio σ_A^2/σ_B^2, the larger the gain that demographic parity brings.

4 TWO-STAGE SELECTION

In the previous section, we considered one-stage selection. In reality many processes are in multiple stages, one typical example being hiring. Candidates are first evaluated using tests. After, the examination proceeds with higher accuracy, for instance by performing an interview for every candidate selected at first stage. In our model, first stage results are represented by low accuracy estimates $\hat{W}$ and second stage results are the true values of quality W. Let us recall that the first-stage fraction of selected candidates is denoted by α_1, and the second-stage fraction is α_2.

As for the one-stage process, we compare the performance obtained when using at first stage one of the algorithms introduced in Section 2.3. The selection properties of the demographic parity, γ-rule and group oblivious algorithms are the same as for the one-stage in Section 3.1: group oblivious tends to select a larger fraction of A-candidates if the selection size α_1 is small and a smaller fraction of A-candidates if the selection size is large. Demographic parity and γ-rule preserve the demographics of candidates during the selection.

For a first-stage selection size α_1 close to α_2, we expect the two-stage selection to have a behavior similar to one-stage. The behavior for large selection sizes is not obvious, but we expect that both the demographic parity and group oblivious selection algorithms should give similar performance, since there is always a chance to fix the selection at second stage because of the large first stage selection. In fact, the following theorem shows that both for α_1 close enough to α_2 and for α_1 large enough, the demographic parity algorithm leads to a higher average utility than the group-oblivious one. We believe that the case of α_1 close enough to α_2 is close to what would happen in a number of real selection problems: the proportion of candidates that the decision maker is able to preselect is not too much higher than the total amount of candidates needed.

THEOREM 2. *For any problem parameters, there exists α_1^* such that if $\alpha_1 < \alpha_1^*$ or $\alpha_1 > 1/2$, then imposing a fairness mechanism at first stage improves the utility $\mathcal{U}$ of the two-stage selection process:*

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

The above inequality is an equality when $\alpha_1 = 1/2$.

SKETCH OF PROOF. *Case $\alpha_1 < \alpha_1^*$.* We showed in Theorem 1 that for one-stage selection, demographic parity always leads to a better selection. Recall that the case $\alpha_1 = \alpha_2$ corresponds to one-stage selection. Thus, if $\alpha_1 = \alpha_2$, then $\mathcal{U}(x_A^{dp}) > \mathcal{U}(x_A^{\gamma\text{-rule}}) \geq \mathcal{U}(x_A^{obl})$. Since $\mathcal{U}(x_A^{dp})$, $\mathcal{U}(x_A^{\gamma\text{-rule}})$ and $\mathcal{U}(x_A^{obl})$ are continuous functions of α_1 (for a fixed α_2), we can always find α_1^* (that may depend on α_2), such that $\forall \alpha_1 < \alpha_1^*$, $\mathcal{U}(x_A^{dp}) > \mathcal{U}(x_A^{\gamma\text{-rule}}) \geq \mathcal{U}(x_A^{obl})$ holds.

Case $\alpha_1 > 1/2$. This second case is harder to study. The key difficulty is to compute the first and second derivatives of $\mathcal{U}$ with respect to x_A, where $\mathcal{U}$ is now the expected utility of the final selection (after selecting the candidates with the highest true quality at second stage). The expression for $\frac{d\mathcal{U}}{dx_A}$ allows us to check whether a particular algorithm selection fraction x_A is smaller or larger than the Bayesian-optimal selection fraction x_A^{opt}. We use the expression of the demographic parity selection fraction x_A^{dp} and substitute it into the expression of $\frac{d\mathcal{U}}{dx_A}$. We obtain that for $\alpha_1 > 1/2$, the derivative of $\mathcal{U}$ is positive, which means that $x_A^{dp} < x_A^{opt}$. At the same time, for $\alpha_1 > 1/2$, from Lemma 3.1, we have $x_A^{obl} < x_A^{dp}$. We prove that $\mathcal{U}$ is strictly concave, again using Harris inequality [14], and thus we conclude that $\mathcal{U}^{dp} > \mathcal{U}^{obl}$. A detailed proof is provided in Appendix B.3. □

Theorem 2 provides sufficient conditions under which the γ-rule and demographic parity algorithms will improve utility. To illustrate this gain, we plot in Fig. 5a the utility values obtained when using the Bayesian optimal, demographic parity, γ-rule or group oblivious selection algorithm. We fix the second stage selection size $\alpha_2 = 0.1$ and vary the first stage selection size $\alpha_1 \in [0.1, 1]$.

We observe that for this example, the value of α_1^* in Theorem 2 seems to be around 0.2. For $\alpha_1 < \alpha_1^*$, the demographic parity or $4/5$-rule algorithms provide a large gain. When $\alpha_1 \in (0.2, 0.5)$, the demographic parity and γ-rule algorithms are not as good as the group-oblivious one but the loss of quality is minimal. The γ-rule algorithm provides a compromise between demographic parity and group oblivious: it provides a smaller gain when demographic parity is better than group oblivious; but it provides a smaller loss when demographic parity is not as good as group oblivious.

In Fig. 5c, we show the performance gap for different cases of implicit variance values. We observe that in general the larger the implicit variance, the larger the gain that demographic parity brings. The gain can be up to 40% in our experiments. In general, as the budget α_1 grows, we observe a smaller performance gap between the demographic parity and group oblivious algorithms. When demographic parity harms the utility, the harm does not exceed 2%.
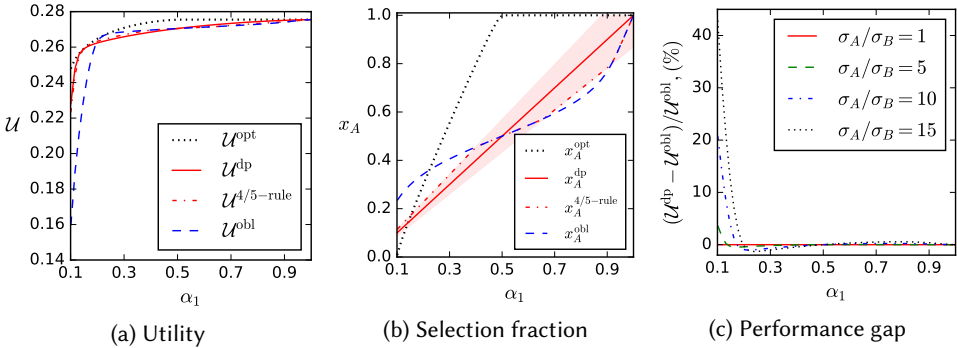

(a) Utility   (b) Selection fraction   (c) Performance gap

Fig. 5. Utility of **two-stage** selection $\mathcal{U}$, selection fraction x_A and performance gap $(\mathcal{U}^{dp} - \mathcal{U}^{obl})/\mathcal{U}^{obl}$ for different budgets α_1. The parameters are $p_A = 0.4$, $\sigma_B = 0.2$, $\alpha_2 = 0.1$; σ_A is set to $\sigma_A = 3$ in (a) and (b).

5 EXPERIMENTS

In this section,[6] 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 Pareto distribution. We also observe a similar behavior when considering in Section 5.2 a real dataset coming from the national Indian exam data. We conclude in Section 5.3 with experiments that show that a case with $n = 20$ candidates behaves similarly as with $n = \infty$.

5.1 Synthetic data with Pareto quality

Our assumption in the theoretical evaluation of Sections 3 and 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 [19], this for example the case for wealth, income or number of citations [9], meaning that a minority possesses a large fraction of the aggregate quality. In this experiment, we suppose that $W \sim \text{Pareto}(w_0, \kappa)$, where $w_0 > 0$ is a scale and $\kappa > 0$ is a shape parameter: the probability density function of W can be written as $p_W(w) = \frac{\kappa w_0^\kappa}{w^{\kappa+1}}$. We generate 100 datasets of size $n = 10,000$. For every dataset we perform a group oblivious and demographic parity selection. In Fig. 6, we report the average utilities $\langle \mathcal{U}_n \rangle$ over the 100 experiments.

In Fig. 6a, we show the performance gap between different selection algorithms for one-stage selection. We see that demographic parity improves the utility in most of the cases and that the largest gap corresponds to the smallest budget α_1. Note that contrary to Theorem 1, demographic parity does not always improve utility (for instance here when $\alpha_1 \in [.3, .5]$). Yet, the loss due to demographic parity is never larger than 0.1% while the gain can be up to 40%. In Fig. 6b, the two-stage case is shown. As expected, demographic parity helps utility for small budgets α_1 close to $\alpha_2 = 0.01$, since the selection is almost the same as in one stage. As the budget α_1 increases, both the demographic parity and group oblivious selection algorithms tend to perform close to each other due to large number of choices at the second stage. Finally, in Fig. 6c, we show how the selection fraction x_A^{obl} depends on α_1. We see that for small budgets α_1, the group oblivious algorithm tends to select more from group A, while for large budgets, the situation is opposite.
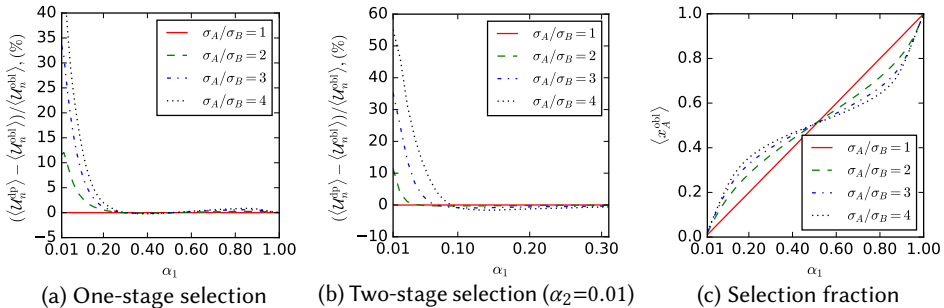

Fig. 6. **Synthetic data** with Pareto distribution $W \sim \text{Pareto}(1, 3)$: Gain of demographic parity over group oblivious for the one- and two-stage selections. The parameters are $p_A = 0.4$, $\sigma_B = 1$, $\sigma_A \in \{1, 2, 3, 4\}$.

Similar results are obtained for other distributions of quality W. We consider uniform, Gaussian mixture and beta distributions. The corresponding plots can be found in Appendix A.

5.2 IIT-JEE scores dataset

In this section, we consider a real dataset, the IIT-JEE dataset [1], with joint entrance exam results in India in 2009. These scores are used as an admission criteria to enter the high-rated universities.

---

[6]All codes are available at: https://gitlab.inria.fr/vemelian/implicit-variance-code [permalink on softwareheritage.org].

(a) Histogram of W (b) Histogram of $\hat{W}$ (c) 1-stage selection   (d) 2-stage selection
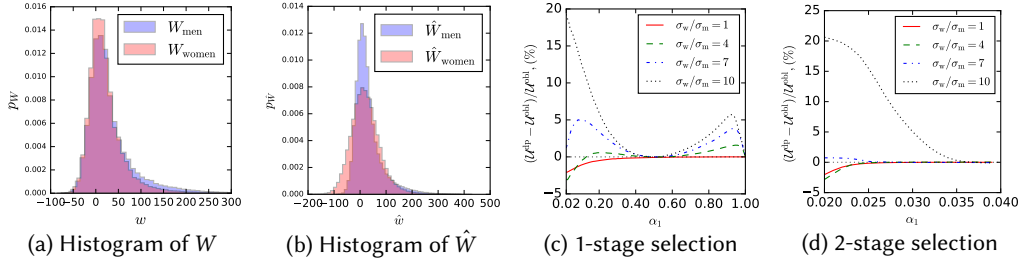
Fig. 7. Distribution of W and $\hat{W}$ given gender, and one- and two-stage selection for **IIT-JEE dataset** [1]. Mean values and standards deviations of W for two groups are: $\mu_{W_{\text{men}}} = 30.8$, $\sigma_{W_{\text{men}}} = 51.8$, $\mu_{W_{\text{women}}} = 21.2$, $\sigma_{W_{\text{women}}} = 39.3$. Added noise has standard deviation $\sigma_{\text{m}} = 10$ and $\sigma_{\text{w}} = k \cdot \sigma_{\text{m}}$; $k = 4$ in plot (b).

The dataset consists of 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 [7].

In order to construct a model of implicit 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: $\mu_{W_{\text{men}}} = 30.8$, $\sigma_{W_{\text{men}}} = 51.8$, $\mu_{W_{\text{women}}} = 21.2$, $\sigma_{W_{\text{women}}} = 39.3$. We then suppose that an unbiased estimator $\hat{W}$ of the grade is observed at the first stage. The standard deviation of estimation for male candidates is set to $\sigma_{\text{m}} = 10$. For the women group, which is the minority group, we consider different cases: $\sigma_{\text{w}} = k \cdot \sigma_{\text{m}}$, for $k = 1, 4, 7, 10$. The distribution of grades W and observed values $\hat{W}$ for $k = 4$ are shown in Fig. 7a and 7b.

We start our experiment with **one-stage** selection. For the dataset we perform a group oblivious (select best m_1) and demographic parity selection (select best m_1, but maintain the demographic parity condition $x_A = x_B$ up to one candidate). 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 [7].

The results for one-stage selection are given in Fig. 7c. We observe that for both small and large values of α_1 demographic parity helps utility, if the noise values of women evaluation σ_{w} are large. We see that the gain can be up to around 30% if the selection size is small and up to 5% if the selection size is large. For the case where σ_{w} and σ_{m} 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 now analyze **two-stage** selection. As for the one-stage case, we perform a group-oblivious selection (select best m_1; then select best m_2 out of m_1) and a demographic parity selection (select best m_1 but maintain demographic parity; then select best m_2 out of m_1). In Fig. 7d we show the case where the final-stage selection rate is $\alpha_2 = 2\%$, i.e., we select 7,700 candidates out of 384,977. As observed, the performance gain when using demographic parity can be up to 20%. However, as the selection size α_1 increases, $\mathcal{U}^{\text{dp}}$ and $\mathcal{U}^{\text{obl}}$ become close, since there will always be enough candidates among those selected at first stage to subselect a tiny proportion of good candidates.

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, we use an approximation that is exact as number of candidates n tends to infinity (Proposition 2.1). However, it is important to know how the approximation behaves for a small number of candidates n and small selection sizes m_1, m_2. For our experiment, we generate datasets of different sizes $n = 20, 50, 100$. For every size parameter n, we generate $10,000$ different datasets. For a population

(a) Average utility $\langle \mathcal{U}_n \rangle$, $n = 100$ (b) Gap of average performance   (c) Average performance gap
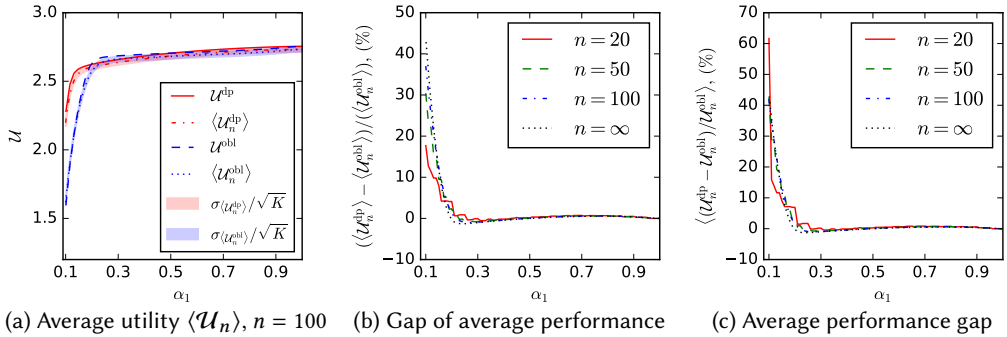
Fig. 8. Finite population size: quality of the two-stage selection and expected gain of demographic parity over group oblivious. The quality distribution W is Normal$(1, 1)$ and the noise parameters are $\sigma_A = 3$, $\sigma_B = 0.2$, $\alpha_2 = 0.1$. The number of experiments per set of parameters is $K = 10,000$. The shaded areas are the confidence intervals (corresponding to one standard deviation on the estimation of the empirical mean).

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 Normal$(1, 1)$.

In Fig. 8a we plot the average utilities $\langle \mathcal{U}_n \rangle$ for a population of $n = 100$, where we select $m_2 = 10$ individuals and where we vary m_1 from 10 to 100. The shaded region corresponds to a confidence interval. We consider two selection algorithms (demographic parity and group oblivious) and compare the performance for $n = 100$ with the limiting quantities $\mathcal{U}^{\mathrm{dp}}$ and $\mathcal{U}^{\mathrm{obl}}$. We observe that, even for $n = 100$, the average values of utility are close to the approximation. In Fig. 8b we compare the gap of average performances $(\langle \mathcal{U}_n^{\mathrm{dp}} \rangle - \langle \mathcal{U}_n^{\mathrm{obl}} \rangle)/\langle \mathcal{U}_n^{\mathrm{obl}} \rangle$ for different n. We observe that the approximation $n = +\infty$ is a good prediction of the average gain provided by the use of demographic parity. In order to distinguish more precisely over the various experiments, in Fig. 8c we compare the average gain of performance $\langle (\mathcal{U}_n^{\mathrm{dp}} - \mathcal{U}_n^{\mathrm{obl}})/\mathcal{U}_n^{\mathrm{obl}} \rangle$. Again, the curves for finite n are almost indistinguishable from the case where $n \to \infty$.

6 DISCUSSION AND EXTENSIONS

In this work, we study one- and two-stage selection problems in the presence of implicit variance. We propose a purposely simple model of the problem that captures the phenomenon of implicit variance and allows us to obtain clean mathematical results. In particular, we show that fairness mechanisms (a generalization of the $4/5$ rule) often lead to a higher selection utility compared to using a group oblivious algorithm. Our model is flexible and can be extended in several directions.

Different prior of the quality distribution. Our theoretical results are obtained under the assumption that the true latent quality W follows a group-independent distribution (to isolate the effect of implicit variance) that corresponds to a normal law (to allow for analytical derivations). Both assumptions can be relaxed. *First*, we can plug into the model any distribution of latent quality (e.g., Pareto, uniform, mixture of Gaussians, etc.). We show numerically in Section 5 and Appendix A that it does not change the flavor of the main results. *Second*, we can consider quality distributions dependent on the group. A natural extension in that direction would be to consider two different normal distributions. It is possible to extend our results to that case (at the expense of increased complexity). Our experiment on the ITT-JEE dataset (Section 5.2), however, gives a preview of how the results are modified: if the effect of implicit variance is small compared to the difference in the true quality distributions then demographic parity can decrease the selection quality for small selection budgets. If the effect of implicit variance is predominant then our results continue to hold.
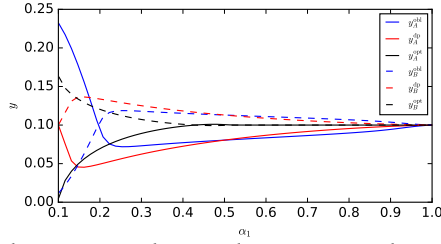
Fig. 9. Selection fractions y_A and y_B corresponding to selection at second stage. Parameters: $W \sim \text{Normal}(1, 1)$, $\sigma_A = 3$, $\sigma_B = 0.2$, $p_A = 0.4$ and $\alpha_2 = 0.1$

Combining implicit variance and implicit bias. Our model does not include implicit bias so as to better isolate the effect of implicit variance. It is easy, though, to incorporate implicit bias as in [7, 19]. The most natural in our model would be to consider $\hat{W}_i = W_i - \beta_{G_i} + \sigma_{G_i}\varepsilon_i$, where β_G is the implicit additive bias against group G (typically, $\beta > 0$ for the disadvantaged group and $\beta = 0$ for the other). In effect, the additive parameter β shifts the distribution of $\hat{W}_i$, while the additive noise widens it. If true qualities are normally distributed, we would then have $\hat{W}_A \sim \text{Normal}(\mu_W - \beta_A, \sigma_W^2 + \sigma_A^2)$ and $\hat{W}_B \sim \text{Normal}(\mu_W - \beta_B, \sigma_W^2 + \sigma_B^2)$. We leave as future work a detailed investigation of the group oblivious and fair selection utilities in that case.

Effect on global fairness in two-stage selection. Throughout the paper, we have studied the effect of imposing demographic parity at the first stage on the final selection *utility*. However, a natural question to ask is what is the effect of imposing fairness at the first stage on the fairness (or disparity) of the final selection (we term it global fairness following [13]). In Fig. 9, we plot the global selection ratios y_A and y_B for each group (at the second stage), for the different first-stage algorithms considered in the paper. For selection budgets α_1 close to α_2, we observe that the first-stage demographic parity algorithm leads to the smallest disparity in global selection fractions. It is natural since the selection fractions are close to ones obtained in the one-stage case (so for demographic parity $y_A \approx y_B$ since $x_A = x_B$). However, we observe that as α_1 grows, demographic parity can lead to a larger inequality in the global selection fractions than with the group oblivious and optimal algorithms. Thus, in a two-stage selection problem with implicit variance, imposing a fairness constraint at the first stage may lead to a degradation of the global fairness. We leave a detailed investigation of this counter-intuitive aspect as future work but note that this emphasizes the crucial importance of modeling implicit variance in multistage selection problems.

ACKNOWLEDGMENTS

REFERENCES

[1] 2019. IIT-JEE dataset. https://github.com/AnayMehrotra/Ranking-with-Implicit-Bias. [Online; accessed Jan 29, 2020].
[2] Dennis J. Aigner and Glen G. Cain. 1977. Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review* 30, 2 (1977), 175–187.
[3] Loukas Balafoutas and Matthias Sutter. 2012. Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. *Science* 335 (Feb. 2012), 579–82.
[4] Ariane Baye and Christian Monseur. 2016. Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education* 4 (Dec. 2016).
[5] Marianne Bertrand and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94, 4 (Sept. 2004), 991–1013.

[6] Marilyn Cavicchia. 2015. How to fight implicit bias? With conscious thought, diversity expert tells NABE. *American Bar Association: Bar Leader* 40, 1 (2015).

[7] L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi. 2020. Interventions for Ranking in the Presence of Implicit Bias. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)*. 369–380.

[8] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (2017), 153–163.

[9] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. Power-Law Distributions in Empirical Data. *SIAM Rev.* 51, 4 (2009), 661–703.

[10] Stephen Coate and Glenn Loury. 1993. Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review* 83 (Feb. 1993), 1220–40.

[11] Brian Collins. 2007. Tackling Unconscious Bias in Hiring Practices: The Plight of the Rooney Rule. *NYU Law Review* 82 (June 2007).

[12] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 797–806.

[13] Vitalii Emelianov, George Arvanitakis, Nicolas Gast, Krishna Gummadi, and Patrick Loiseau. 2019. The Price of Local Fairness in Multistage Selection. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*.

[14] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. 1971. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* 22, 2 (1971), 89–103.

[15] Anthony Greenwald and Linda Krieger. 2006. Implicit Bias: Scientific Foundations. *California Law Review* 94 (July 2006), 945.

[16] Moritz Hardt, Eric Price, and Nathan Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*. 3323–3331.

[17] Harry Holzer and David Neumark. 2000. Assessing Affirmative Action. *Journal of Economic Literature* 38, 3 (Sept. 2000), 483–568.

[18] Sampath Kannan, Aaron Roth, and Juba Ziani. 2019. Downstream Effects of Affirmative Action. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*.

[19] Jon M. Kleinberg and Manish Raghavan. 2018. Selection Problems in the Presence of Implicit Bias. In *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS)*. 33:1–33:17.

[20] Zachary Lipton, Julian McAuley, and Alexandra Chouldechova. 2018. Does mitigating ML's impact disparity require treatment disparity?. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS)*. 8125–8135.

[21] Francesco Locatello, Gabriele Abbati, Thomas Rainforth, Stefan Bauer, Bernhard Schölkopf, and Olivier Bachem. 2019. On the Fairness of Disentangled Representations. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*. 14584–14597.

[22] Michael Mathioudakis, Carlos Castillo, Giorgio Barnabo, and Sergio Celis. 2020. 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)*. 440–449.

[23] R. E. O'Dea, M. Lagisz, M. D. Jennions, and S. Nakagawa. 2018. Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications* 9, 1 (2018), 3777.

[24] Christina Passariello. 2016. Tech Firms Borrow Football Play to Increase Hiring of Women. *Wall Street Journal* (27 Sept. 2016).

[25] Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. 2008. Discrimination-aware Data Mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 560–568.

[26] Edmund Phelps. 1972. The Statistical Theory of Racism and Sexism. *American Economic Review* 62, 4 (1972), 659–61.

[27] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)*. 469–481.

[28] Muhammad B. Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi. 2017. Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification Without Disparate Mistreatment. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*. 1171–1180.

[29] Muhammad B. Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P. Gummadi. 2017. Fairness Constraints: Mechanisms for Fair Classification. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 962–970.

[30] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning Fair Representations. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*. 325–333.

A ADDITIONAL PLOTS

As was mentioned in Section 5, we perform our study on different prior distributions of W. In Section 5, we studied the Pareto case. In this section, we show the selection results for other distributions: Uniform, Gaussian mixture and Beta distribution.

The most interesting case is a two component Gaussian mixture, where with probability $\pi = 0.2$ there appears a "good" candidate, and with probability 0.8 appears a "bad" candidate. This is a typical situation where only small proportion of candidates are "suitable" for the selection. We set selection budget $\alpha_2 = 0.1$, which means that we aim to select only candidates from the second peak. Parameters of this mixture are as follows: $\mu_1 = 0.0$, $\mu_2 = 0.2$, $\sigma_1 = \sigma_2 = 0.05$.

While performing **one-stage** selection, we vary the selection budget α_1 from 0.1 to 1. In **two-stage** selection, we fix $\alpha_2 = 0.1$ and vary α_1 from α_2 to 1. The implicit variance for B-candidates is fixed to $\sigma_B = 0.1$, the implicit variance for A-candidates varies with $k = 1, 2, 3, 4$ as $\sigma_A = k\sigma_B$. Parameters $p_A = 0.4$ and $p_B = 0.6$.

On top row of Figure 10, we plot corresponding pdfs, on middle row we show the result for one-stage selection and on bottom row, the two-stage selection is displayed. We observe that the demographic parity selection outperforms the group oblivious algorithm in one-stage selection for small and large α_1. In two-stage selection, demographic parity is also better than the group oblivious algorithm for values α_1 close α_2, while for $\alpha_1 \gg \alpha_2$, both tend to perform similarly.

B OMMITED PROOFS

In this section we provide detailed proofs of the statements given before. Namely, these are proofs of Lemma 3.1, Lemma 3.2 and Theorem 2.

B.1 Proof of Lemma 3.1

The distribution of W is common for both groups and follows a normal law with parameters μ_W and σ_W^2. The noise ε is centered normal, thus, $\hat{W} = W + \varepsilon\sigma_G$ follows a normal law with parameters $\mu_{\hat{W}|G} = \mu_W$ and $\sigma_{\hat{W}|G}^2 = \sigma_W^2 + \sigma_G^2$.

The selection fraction x_G^{obl} is, by definition, the probability to observe a G-candidate with $\hat{W}$ larger than $\hat{\theta}^{\mathrm{obl}}$ (where $\hat{\theta}^{\mathrm{obl}} = \hat{\theta}_A^{\mathrm{obl}} = \hat{\theta}_B^{\mathrm{obl}}$ since the thresholds are the same in the group-oblivious selection algoritm). Thus, $x_G^{\mathrm{obl}} = \mathrm{P}(\hat{W} \geq \hat{\theta}^{\mathrm{obl}} \mid G) = \Phi^c\left(\frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_G^2 + \sigma_W^2}}\right)$. Recall that to shorten the notation, we write $\mathrm{P}(\hat{W} \geq \hat{\theta}^{\mathrm{obl}} \mid G)$ for $\mathrm{P}(\hat{W}_i \geq \hat{\theta}^{\mathrm{obl}} \mid G_i = G)$.

Now, we consider three different cases:

(1) if $\alpha_1 < 1/2$, then $\hat{\theta}^{\mathrm{obl}} > \mu_W$ and $\frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_A^2 + \sigma_W^2}} < \frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_B^2 + \sigma_W^2}}$;

(2) if $\alpha_1 = 1/2$, then $\hat{\theta}^{\mathrm{obl}} = \mu_W$ and $\frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_A^2 + \sigma_W^2}} = \frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_B^2 + \sigma_W^2}} = 0$;

(3) if $\alpha_1 > 1/2$, then $\hat{\theta}^{\mathrm{obl}} < \mu_W$ and $\frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_A^2 + \sigma_W^2}} > \frac{\hat{\theta}^{\mathrm{obl}} - \mu_W}{\sqrt{\sigma_B^2 + \sigma_W^2}}$.

The function Φ^c is decreasing with its argument, thus the statement of lemma is a direct consequence of the above inequalities.

B.2 Proof of Lemma 3.2

The following lemma is necessary to study the optimal one-stage selection. The proof of it is technical and we postpone it to Section B.4.

LEMMA B.1 (DERIVATIVES OF ONE-STAGE UTILITY $\mathcal{U}$). *For any α_1, where $0 < \alpha_1 \leq 1$, we have*
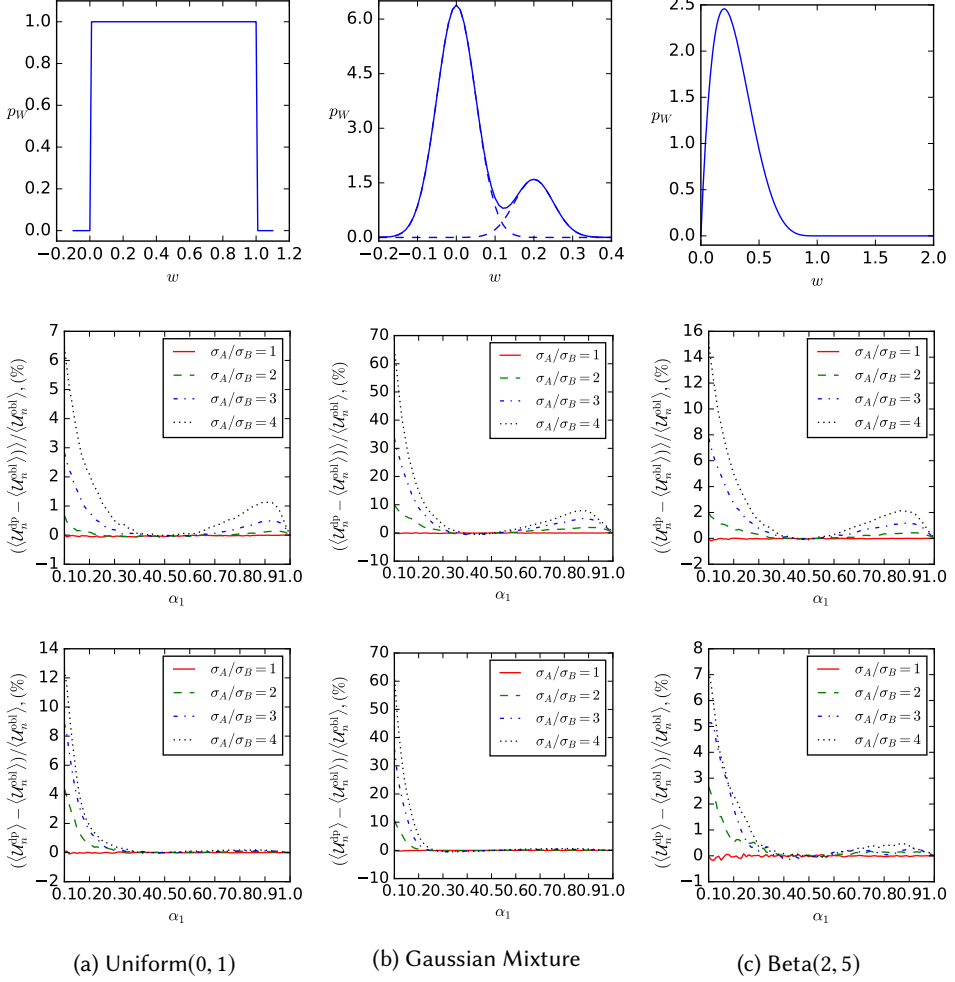
Fig. 10. One- and two-stage selection results for different prior distributions of quality W. The implicit variance for B-candidates is fixed to $\sigma_B = 0.1$, the implicit variance for A-candidates varies with $k = 1, 2, 3, 4$ as $\sigma_A = k\sigma_B$. We set $n = 1,000$ and average over 100 experiments.

(1) $\frac{d\,\mathcal{U}}{dx_A} = \frac{p_A}{\alpha_1} \left[\frac{\hat{\theta}_A \cdot \sigma_W^2 + \mu_W \cdot \sigma_A^2}{\sigma_W^2 + \sigma_A^2} - \frac{\hat{\theta}_B \cdot \sigma_W^2 + \mu_W \cdot \sigma_B^2}{\sigma_W^2 + \sigma_B^2} \right]$;

(2) $\frac{d^2\,\mathcal{U}}{dx_A^2} < 0,$

where $\hat{\theta}_A$ and $\hat{\theta}_B$ are such that

$$
\begin{cases}
\mathrm{P}(\hat{W}_A \geq \hat{\theta}_A \mid G = A) = x_A, \\
\sum_{G \in \{A,B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) \cdot \mathrm{P}(G) = \alpha_1.
\end{cases}
$$

The result in Lemma B.1 helps to obtain several insights about the optimal one-stage selection. From the condition on a maximum $\frac{d\,\mathcal{U}}{dx_A} = 0$, we obtain that optimal quantiles $\hat{\theta}_G^{\mathrm{opt}}$ should satisfy $\hat{\theta}_G^{\mathrm{opt}} = C \cdot (\sigma_W^2 + \sigma_G^2) + \mu_W$ for some C that does not depend on the group G. Since $\hat{W}|G$ follows a

normal law with parameters $\mu_{\hat{W}|G} = \mu_W$, $\sigma^2_{\hat{W}|G} = \sigma^2_G + \sigma^2_W$, the selection probability x^{opt}_G can be written as $x^{\text{opt}}_G = P(\hat{W} \geq \hat{\theta}^{\text{opt}}_G \mid G) = \Phi^c \left(\frac{\hat{\theta}^{\text{opt}}_G - \mu_W}{\sqrt{\sigma^2_W + \sigma^2_G}} \right)$. Using the expression for optimal quantiles, we obtain $x^{\text{opt}}_G = \Phi^c \left(C \cdot \sqrt{\sigma^2_W + \sigma^2_G} \right)$.

Now, three cases are possible:

(1) If $\hat{\theta}^{\text{opt}}_G < \mu_W$, then $C < 0$. As a result, using the expression for x^{opt}_G, we obtain that $x^{\text{opt}}_A > x^{\text{opt}}_B$.
(2) If $\hat{\theta}^{\text{opt}}_G > \mu_W$, then $C > 0$. Hence, $x^{\text{opt}}_A < x^{\text{opt}}_B$.
(3) If $\hat{\theta}^{\text{opt}}_G = \mu_W$, then $C = 0$ and $x^{\text{opt}}_A = x^{\text{opt}}_B = 1/2$.

Since $p_A x^{\text{opt}}_A + p_B x^{\text{opt}}_B = \alpha_1$, then x^{dp}_A is a convex combination of x^{opt}_A and x^{opt}_B. Thus, $x^{\text{dp}}_A < x^{\text{opt}}_A$ for $\hat{\theta}^{\text{opt}}_G < \mu_W$, $x^{\text{dp}}_A > x^{\text{opt}}_A$ for $\hat{\theta}^{\text{opt}}_G > \mu_W$ and $x^{\text{dp}}_A = x^{\text{opt}}_A$ for $\hat{\theta}^{\text{opt}}_G = \mu_W$.

B.3 Proof of Theorem 2

The following lemma is necessary to study the optimal two-stage selection. Due to the technicality of its proof, we postpone it to Section B.5.

LEMMA B.2 (DERIVATIVES OF TWO-STAGE UTILITY $\mathcal{U}$). *For any α_1, α_2, where $0 < \alpha_2 \leq \alpha_1$:*

(1) $\frac{d\,\mathcal{U}}{dx_A} = \frac{1}{\alpha_2} p_A \left[\hat{\sigma}_A \int_{-\infty}^{(\hat{\mu}_A - \theta)/\hat{\sigma}_A} \Phi(\tau)d\tau - \hat{\sigma}_B \int_{-\infty}^{(\hat{\mu}_B - \theta)/\hat{\sigma}_B} \Phi(\tau)d\tau \right]$, *where*

$$\hat{\mu}_G = \frac{\mu_W \sigma^2_G + \hat{\theta}_G \sigma^2_W}{\sigma^2_G + \sigma^2_W}, \quad \hat{\sigma}^2_G = \frac{\sigma^2_G \sigma^2_W}{\sigma^2_G + \sigma^2_W};$$

(2) $\frac{d^2\,\mathcal{U}}{dx^2_A} < 0$,

where $\hat{\theta}_A$, $\hat{\theta}_B$ and θ are such that

$$\begin{cases} P(\hat{W}_A \geq \hat{\theta}_A \mid G = A) = x_A, \\ \sum_{G \in \{A, B\}} P(\hat{W} \geq \hat{\theta}_G \mid G) \cdot P(G) = \alpha_1, \\ \sum_{G \in \{A, B\}} P(\hat{W} \geq \hat{\theta}_G, W \geq \theta \mid G) \cdot P(G) = \alpha_2. \end{cases}$$

The proof proceeds as follows. First, we show that for $\alpha_1 > 1/2$, we have $x^{\text{opt}}_A > x^{\text{dp}}_A$. Then, using Lemma 3.1, we have $x^{\text{dp}}_A > x^{\text{obl}}_A$ for $\alpha_1 > 1/2$. Using the property that $x^{\text{obl}}_A < x^{\text{dp}}_A < x^{\text{opt}}_A$, we conclude that there exists $\lambda \in (0, 1)$, such that $x^{\text{dp}}_A = \lambda x^{\text{obl}}_A + (1 - \lambda)x^{\text{opt}}_A$. By strict concavity of two-stage selection utility, we obtain

$$\mathcal{U}\left(x^{\text{dp}}_A \right) = \mathcal{U}\left(\lambda x^{\text{obl}}_A + (1 - \lambda)x^{\text{opt}}_A \right) > \lambda\, \mathcal{U}\left(x^{\text{obl}}_A \right) + (1 - \lambda)\, \mathcal{U}\left(x^{\text{opt}}_A \right) > \mathcal{U}\left(x^{\text{obl}}_A \right).$$

Hence, to complete the proof we need to show that $x^{\text{opt}}_A > x^{\text{dp}}_A$ for $\alpha_1 > 1/2$, which we do in the next few paragraphs.

The random variable $\hat{W}|G$ follows a normal law with parameters $\mu_{\hat{W}|G} = \mu_W$ and $\sigma^2_{\hat{W}|G} = \sigma^2_G + \sigma^2_W$. Hence, the threshold $\hat{\theta}^{\text{dp}}_G = F^{-1}_{\hat{W}|G}(1 - \alpha_1)$ can be simplified to $\hat{\theta}^{\text{dp}}_G = \mu_W + \sqrt{\sigma^2_W + \sigma^2_G} \cdot \Phi^{-1}(1 - \alpha_1)$. By using the expression for $\mathcal{U}'$ from Lemma B.2 and substituting $\hat{\theta}^{\text{dp}}_G$ into it, we obtain

$$\mathcal{U}'(x^{\text{dp}}_A) = \frac{1}{\alpha_2} p_A \left(\hat{\sigma}_A \int_{-\infty}^{(\hat{\mu}_A - \theta)/\hat{\sigma}_A} \Phi(\tau)d\tau - \hat{\sigma}_B \int_{-\infty}^{(\hat{\mu}_B - \theta)/\hat{\sigma}_B} \Phi(\tau)d\tau \right),$$

where $\hat{\mu}_G = \mu_W + \frac{\Phi^{-1}(1-\alpha_1)\cdot\sigma_W^2}{\sqrt{\sigma_W^2+\sigma_G^2}}$ and $\hat{\sigma}_G = \frac{\sigma_G^2\sigma_W^2}{\sigma_G^2+\sigma_W^2}$. As a result, $\hat{\mu}_A - \theta > \hat{\mu}_B - \theta$ for $\Phi^{-1}(1-\alpha_1) < 0 \iff \alpha_1 > 1/2$. Thus, $\mathcal{U}'(x_A^{\mathrm{dp}}) \geq I(\hat{\sigma}_A) - I(\hat{\sigma}_B)$, where $I(\sigma) = \sigma \int_{-\infty}^{I_0/\sigma} \Phi(\tau)d\tau$ for some $I_0 \in \mathbb{R}$.

We can show that the function $I(\sigma)$ is increasing for any value of I_0 by looking at its first derivative:

$$I'_\sigma = \int_{-\infty}^{I_0/\sigma} \Phi(\tau)d\tau + \sigma\Phi\left(\frac{I_0}{\sigma}\right)\left(-\frac{I_0}{\sigma^2}\right) = \frac{I_0}{\sigma}\Phi\left(\frac{I_0}{\sigma}\right) + \phi\left(\frac{I_0}{\sigma}\right) - \frac{I_0}{\sigma}\Phi\left(\frac{I_0}{\sigma}\right) = \phi\left(\frac{I_0}{\sigma}\right) > 0 .$$

Thus, for demographic parity selection, $\mathcal{U}'(x_A^{\mathrm{dp}}) > 0$. Due to concavity of $\mathcal{U}$, we obtain $x_A^{\mathrm{opt}} > x_A^{\mathrm{dp}}$, which concludes the proof.

B.4 Proof of Lemma B.1

In this section, we provide a proof of Lemma B.1. We prove it in general setting, i.e. quality W follows any general distribution having a probability density function $p_W(w)$.

The one-stage selection can be viewed as a two-stage selection where we select all candidates at second stage. This corresponds to setting the second-stage threshold to $\theta = -\infty$. Writing $\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B) = \mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, -\infty)$ by abuse of notation, the one-stage utility $\mathcal{U}$ is defined as

$$\mathcal{U}(x_A) = \mathcal{V}(\hat{\theta}_A, \hat{\theta}_B),$$

$$\text{with } \hat{\theta}_A, \hat{\theta}_B \text{ such that} \begin{cases} \mathrm{P}(\hat{W} \geq \hat{\theta}_A \mid G = A) = x_A, \\ \sum_{G \in \{A, B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) \cdot \mathrm{P}(G) = \alpha_1. \end{cases}$$

For a general prior distribution of the quality W, the quantities we study can be expressed as

$$\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B) = \frac{1}{\alpha_1} \sum_G p_G \int_{\hat{\theta}_G} d\hat{w} \int dw\, w \cdot p_W(w) \cdot \frac{1}{\sigma_G}\phi\left(\frac{w-\hat{w}}{\sigma_G}\right),$$

$$\mathrm{P}(\hat{W}_G \geq \hat{\theta}_G \mid G) = \int_{\hat{\theta}_G} d\hat{w} \int dw\, p_W(w) \cdot \frac{1}{\sigma_G}\phi\left(\frac{w-\hat{w}}{\sigma_G}\right) = x_G(\hat{\theta}_G),$$

where $p_W(w)$ is a probability density function of the distribution of quality W. The symbol $\int$ here represents the integration over all support of the corresponding probability density function $p_W(w)$, unless the limits are specified. We prove the following statement about the derivatives of the one-stage utility $\mathcal{U}$.

LEMMA (DERIVATIVES OF $\mathcal{U}$). *For the one-stage selection utility $\mathcal{U}$, we have*

$$\frac{d\mathcal{U}}{dx_A} = \frac{p_A}{\alpha_1}\left[\frac{\int w \cdot p_W(w) \cdot \frac{1}{\sigma_A}\phi\left(\frac{w-\hat{\theta}_A}{\sigma_A}\right)dw}{\int p_W(w) \cdot \frac{1}{\sigma_A}\phi\left(\frac{w-\hat{\theta}_A}{\sigma_A}\right)dw} - \frac{\int w \cdot p_W(w) \cdot \frac{1}{\sigma_B}\phi\left(\frac{w-\hat{\theta}_B}{\sigma_B}\right)dw}{\int p_W(w) \cdot \frac{1}{\sigma_B}\phi\left(\frac{w-\hat{\theta}_B}{\sigma_B}\right)dw}\right]; \qquad (5)$$

$$\frac{d^2\mathcal{U}}{dx_A^2} < 0, \qquad (6)$$

where $\hat{\theta}_A$ and $\hat{\theta}_B$ are such that

$$\begin{cases} \mathrm{P}(\hat{W}_A \geq \hat{\theta}_A \mid G = A) = x_A, \\ \sum_{G \in \{A, B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) \cdot \mathrm{P}(G) = \alpha_1. \end{cases}$$

First Derivative of $\mathcal{U}$. Using the parameterization by x_A, we can write

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

From the first stage constraint $p_A x_A + p_B x_B = \alpha_1$,

$$p_A\frac{dx_A}{dx_A} + p_B\frac{\partial x_B}{\partial\hat{\theta}_B}\frac{d\hat{\theta}_B}{dx_A} = 0 \implies \frac{d\hat{\theta}_B}{dx_A} = -\frac{p_A}{p_B}\frac{\partial\hat{\theta}_B}{\partial x_B}.$$

Then $\frac{d\,\mathcal{U}}{dx_A} = p_A\left(\frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_A}\frac{\partial\hat{\theta}_A}{\partial x_A} - \frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_B}\frac{\partial\hat{\theta}_B}{\partial x_B}\right)$ from which (5) follows directly.

Second Derivative of $\mathcal{U}$. The second derivative of $\mathcal{U}$ with respect to x_A, then can be written as $\frac{d^2\,\mathcal{U}}{dx_A^2} = \sum_G \frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_G}\frac{d\hat{\theta}_G}{dx_A}$, where by $\mathcal{Z}$ we denote the expression (5). Using (5), we conclude that

$$\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_G} = \frac{-\int w\cdot p_W(w)\cdot\left(\frac{\hat{\theta}_G-w}{\sigma_G^2}\right)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw}{\alpha_1\left(\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw\right)^2}$$

$$+ \frac{\int w\cdot p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw\int p_W(w)\cdot\left(\frac{\hat{\theta}_G-w}{\sigma_G^2}\right)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw}{\alpha_1\left(\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw\right)^2}.$$

If we use the notation $dP = p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw$, then

$$\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_G} = \frac{-\int w(\hat{\theta}_G - w)dP\cdot\int dP + \int wdP\cdot\int(\hat{\theta}_G - w)dP}{\alpha_1\sigma_G^2\left(\int dP\right)^2}$$

$$= \frac{\int w^2 dP\cdot\int dP - \int wdP\cdot\int wdP}{\alpha_1\sigma_G^2\left(\int dP\right)^2} = \frac{1}{\alpha_1\sigma_G^2}\left[\frac{\int w^2 dP}{\int dP} - \frac{\int wdP}{\int dP}\cdot\frac{\int wdP}{\int dP}\right].$$

Since $f(w) = w^2$ and $g(w) = w$ are increasing functions of w, then by applying Harris inequality [14] to f, g and probability measure $d\mu = dP/\int dP$, we obtain

$$\frac{\int w^2 dP}{\int dP} - \frac{\int wdP}{\int dP}\cdot\frac{\int wdP}{\int dP} > 0\;.$$

Hence, the second derivative of $\mathcal{U}$ with respect to x_A can be written as

$$\frac{d^2\,\mathcal{U}}{dx_A^2} = p_A\frac{d\hat{\theta}_A}{dx_A}\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_A} - p_A\frac{d\hat{\theta}_B}{x_A}\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_B} = p_A\frac{\partial\hat{\theta}_A}{\partial x_A}\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_A} + \frac{p_A^2}{p_B}\frac{\partial\hat{\theta}_B}{\partial x_B}\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_B} < 0.$$

B.5 Proof of Lemma B.2

In this section, we provide a proof of Lemma B.2. We also prove it in general setting, i.e. quality W follows any general distribution having a probability density function $p_W(w)$.

The two-stage utility is

$$\mathcal{U}(x_A) = \mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, \theta),$$

$$\text{with } \hat{\theta}_A, \hat{\theta}_B, \theta \text{ such that } \begin{cases} \mathrm{P}(\hat{W}_A \geq \hat{\theta}_A \mid G = A) = x_A, \\ \sum_{G \in \{A,B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) \cdot \mathrm{P}(G) = \alpha_1, \\ \sum_{G \in \{A,B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G, W \geq \theta \mid G) \cdot \mathrm{P}(G) = \alpha_2. \end{cases}$$

For a general prior distribution of the quality W, the quantities we study can be written as

$$\mathcal{V}(\hat{\theta}_A, \hat{\theta}_B, \theta) = \frac{1}{\alpha_2} \sum_G p_G \int_{\hat{\theta}_G} d\hat{w} \int_\theta dw\, w \cdot p_W(w) \cdot \frac{1}{\sigma_G} \phi\left(\frac{w - \hat{w}}{\sigma_G}\right),$$

$$\mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) = \int_{\hat{\theta}_G} d\hat{w} \int dw\, p_W(w) \cdot \frac{1}{\sigma_G} \phi\left(\frac{w - \hat{w}}{\sigma_G}\right) = x_G(\hat{\theta}_G),$$

$$\mathrm{P}(\hat{W} \geq \hat{\theta}_G, W \geq \theta \mid G) = \int_{\hat{\theta}_G} d\hat{w} \int_\theta dw\, p_W(w) \cdot \frac{1}{\sigma_G} \phi\left(\frac{w - \hat{w}}{\sigma_G}\right) = y_G(\hat{\theta}_G, \theta),$$

where $p_W(w)$ is a probability density function of the distribution of the quality W. The symbol $\int$ here represents the integration over all support of the corresponding probability density function $p_W(w)$, unless the limits are specified. We want to prove the following statement about the derivatives of the two-stage utility $\mathcal{U}$.

LEMMA (DERIVATIVES OF $\mathcal{U}$). *For the two-stage utility $\mathcal{U}$, we have*

$$\frac{d\mathcal{U}}{dx_A} = \frac{p_A}{\alpha_2}\left(\frac{\int_\theta (w - \theta) \cdot p_W(w) \cdot \frac{1}{\sigma_A}\phi\left(\frac{w - \hat{\theta}_A}{\sigma_A}\right) dw}{\int p_W(w) \cdot \frac{1}{\sigma_A}\phi\left(\frac{w - \hat{\theta}_A}{\sigma_A}\right) dw} - \frac{\int_\theta (w - \theta) \cdot p_W(w) \cdot \frac{1}{\sigma_B}\phi\left(\frac{w - \hat{\theta}_B}{\sigma_B}\right) dw}{\int p_W(w) \cdot \frac{1}{\sigma_B}\phi\left(\frac{w - \hat{\theta}_B}{\sigma_B}\right) dw} \right); \quad (7)$$

$$\frac{d^2\mathcal{U}}{dx_A^2} < 0, \tag{8}$$

where $\hat{\theta}_A$, $\hat{\theta}_B$ and θ are such that

$$\begin{cases} \mathrm{P}(\hat{W}_A \geq \hat{\theta}_A \mid G = A) = x_A, \\ \sum_{G \in \{A,B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G \mid G) \cdot \mathrm{P}(G) = \alpha_1, \\ \sum_{G \in \{A,B\}} \mathrm{P}(\hat{W} \geq \hat{\theta}_G, W \geq \theta \mid G) \cdot \mathrm{P}(G) = \alpha_2. \end{cases}$$

First Derivative of $\mathcal{U}$. Using the parameterization by x_A, $\frac{d\mathcal{U}}{dx_A}$ can be written as

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

From the first stage constraint $p_A x_A + p_B x_B = \alpha_1$:

$$p_A \frac{dx_A}{dx_A} + p_B \frac{\partial x_B}{\partial \hat{\theta}_B} \frac{d\hat{\theta}_B}{dx_A} = 0 \implies \frac{d\hat{\theta}_B}{dx_A} = -\frac{p_A}{p_B} \frac{\partial \hat{\theta}_B}{\partial x_B}.$$

From second stage budget constraint $y := p_A y_A + p_B y_B = \alpha_2$, we get

$$\sum_G p_G \frac{\partial y_G}{\partial \hat{\theta}_G} \frac{d\hat{\theta}_G}{dx_A} + \frac{\partial y}{\partial \theta} \frac{d\theta}{dx_A} = 0 \implies \frac{d\theta}{dx_A} = -\left(\frac{\partial y}{\partial \theta}\right)^{-1} \sum_G p_G \frac{\partial y_G}{\partial \hat{\theta}_G} \frac{d\hat{\theta}_G}{dx_A}.$$

Hence,

$$
\begin{aligned}
\frac{d\,\mathcal{U}}{dx_A} &= p_A \left(\frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_A}\frac{\partial\hat{\theta}_A}{\partial x_A} - \frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_B}\frac{\partial\hat{\theta}_B}{\partial x_B} \right) - \left(\frac{\partial y}{\partial \theta} \right)^{-1} p_A \left(\frac{\partial y_A}{\partial\hat{\theta}_A}\frac{\partial\hat{\theta}_A}{\partial x_A} - \frac{\partial y_B}{\partial\hat{\theta}_B}\frac{\partial\hat{\theta}_B}{\partial x_B} \right) \frac{\partial\,\mathcal{V}}{\partial\theta} \\
&= p_A \left[\frac{\partial\hat{\theta}_A}{\partial x_A} \left(\frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_A} - \frac{\partial\,\mathcal{V}}{\partial\theta}\frac{\partial\theta}{\partial y}\frac{\partial y_A}{\partial\hat{\theta}_A} \right) - \frac{\partial\hat{\theta}_B}{\partial x_B} \left(\frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_B} - \frac{\partial\,\mathcal{V}}{\partial\theta}\frac{\partial\theta}{\partial y}\frac{\partial y_B}{\partial\hat{\theta}_B} \right) \right].
\end{aligned}
$$

Let us consider the following quantity:

$$
\begin{aligned}
\frac{\partial\hat{\theta}_G}{\partial x_G}\left(\frac{\partial\,\mathcal{V}}{\partial\hat{\theta}_G} - \frac{\partial\,\mathcal{V}}{\partial\theta}\frac{\partial\theta}{\partial y}\frac{\partial y_G}{\partial\hat{\theta}_G} \right) &= \frac{\frac{1}{\alpha_2}\int_\theta w p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{\hat{\theta}_G-w}{\sigma_G}\right) dw}{\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{\hat{\theta}_G-w}{\sigma_G}\right) dw} \\
&\quad - \frac{\int_\theta p_W(w)\frac{1}{\sigma_g}\phi\left(\frac{\hat{\theta}_G-w}{\sigma_G}\right) dw}{\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{\hat{\theta}_G-w}{\sigma_G}\right) dw} \cdot \underbrace{\frac{\frac{1}{\alpha_2}\sum_g p_G \int_{\hat{\theta}_G}\theta p_W(\theta)\frac{1}{\sigma_G}\phi\left(\frac{\hat{w}-\theta}{\sigma_G}\right) d\hat{w}}{\sum_g p_G \int_{\hat{\theta}_G} p_W(\theta)\frac{1}{\sigma_G}\phi\left(\frac{\hat{w}-\theta}{\sigma_G}\right) d\hat{w}}}_{=\theta} \\
&= \frac{1}{\alpha_2}\frac{\int_\theta (w-\theta)\cdot p_W(w)\cdot\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw}{\int p_W(w)\cdot\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw}.
\end{aligned}
$$

Equation (7) follows directly from the expression developed above.

Second Derivative of $\mathcal{U}$. The second derivative of $\mathcal{U}$ with respect to x_A, then can be written as:

$$
\frac{d^2\,\mathcal{U}}{dx_A^2} = \sum_G \frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_G}\frac{d\hat{\theta}_G}{dx_A} + \frac{\partial\,\mathcal{Z}}{\partial\theta}\frac{d\theta}{dx_A},
$$

where by $\mathcal{Z}$ we denote the expression (7). Then, by using (7), we can calculate

$$
\begin{aligned}
\frac{\partial\,\mathcal{Z}}{\partial\theta} &= 0; \\
\frac{\partial\,\mathcal{Z}}{\partial\hat{\theta}_G} &= \frac{-\int_\theta (w-\theta)\cdot p_W(w)\cdot\left(\frac{\hat{\theta}_G-w}{\sigma_G^2}\right)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw \int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw}{\alpha_2 \left(\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw\right)^2} \\
&\quad + \frac{\int_\theta (w-\theta)\cdot p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw \int p_W(w)\cdot\left(\frac{\hat{\theta}_G-w}{\sigma_G^2}\right)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw}{\alpha_2 \left(\int p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right) dw\right)^2}.
\end{aligned}
$$

Using the notation $dP = p_W(w)\frac{1}{\sigma_G}\phi\left(\frac{w-\hat{\theta}_G}{\sigma_G}\right)dw$, we get

$$\frac{\partial \mathcal{Z}}{\partial \hat{\theta}_G} = \frac{-\int_\theta (w-\theta)(\hat{\theta}_G - w)dP \cdot \int dP + \int_\theta (w-\theta)dP \cdot \int(\hat{\theta}_G - w)dP}{\alpha_2 \sigma_G^2 \left(\int dP\right)^2}$$

$$= \frac{\int_\theta (w^2 - \theta w)dP \cdot \int dP - \int_\theta (w-\theta)dP \cdot \int wdP}{\alpha_2 \sigma_G^2 \left(\int dP\right)^2}$$

$$= \frac{1}{\alpha_2 \sigma_G^2}\left[\frac{\int w \cdot (w-\theta)\,\mathbb{1}_{w\geq\theta}\,dP}{\int dP} - \frac{\int(w-\theta)\,\mathbb{1}_{w\geq\theta}\,dP}{\int dP} \cdot \frac{\int wdP}{\int dP}\right].$$

Since $f(w) = w$ and $g(w) = (w-\theta)\cdot\mathbb{1}_{w\geq\theta}$ are increasing functions of w, then by applying Harris inequality [14] to f, g and probability measure $d\mu = dP/\int dP$, we obtain:

$$\frac{\int w \cdot (w-\theta)\,\mathbb{1}_{w\geq\theta}\,dP}{\int dP} - \frac{\int(w-\theta)\,\mathbb{1}_{w\geq\theta}\,dP}{\int dP} \cdot \frac{\int wdP}{\int dP} > 0\,.$$

Then, the second derivative of $\mathcal{U}$ with respect to x_A can be written as

$$\frac{d^2\mathcal{U}}{dx_A^2} = p_A \frac{d\hat{\theta}_A}{dx_A}\frac{\partial \mathcal{Z}}{\partial \hat{\theta}_A} - p_A \frac{d\hat{\theta}_B}{x_A}\frac{\partial \mathcal{Z}}{\partial \hat{\theta}_B} = p_A \frac{\partial \hat{\theta}_A}{\partial x_A}\frac{\partial \mathcal{Z}}{\partial \hat{\theta}_A} + \frac{p_A^2}{p_B}\frac{\partial \hat{\theta}_B}{\partial x_B}\frac{\partial \mathcal{Z}}{\partial \hat{\theta}_B} < 0.$$