On Statistical Discrimination as a Failure of Social Learning: A Multi-Armed Bandit Approach*

Junpei Komiyama† Shunya Noda‡

First Draft: October 2, 2020 Current Version: December 7, 2020

Abstract

We analyze statistical discrimination using a multi-armed bandit model where myopic firms face candidate workers arriving with heterogeneous observable characteristics. The association between the worker's skill and characteristics is unknown ex ante; thus, firms need to learn it. In such an environment, laissez-faire may result in a highly unfair and inefficient outcome—myopic firms are reluctant to hire minority workers because the lack of data about minority workers prevents accurate estimation of their performance. Consequently, minority groups could be *perpetually underestimated*—they are never hired, and therefore, data about them is never accumulated. We proved that this problem becomes more serious when the population ratio is imbalanced, as is the case in many extant discrimination problems. We consider two affirmative-action policies for solving this dilemma: One is a subsidy rule that is based on the popular upper confidence bound algorithm, and another is the Rooney Rule, which requires firms to interview at least one minority worker for each hiring opportunity. Our results indicate temporary affirmative actions are effective for statistical discrimination caused by data insufficiency.

JEL Codes: C44, D82, D83, J71
Keywords: Statistical Discrimination, Affirmative Action, Multi-Armed Bandit, Social Learning, Strategic Experimentation

*We are grateful to Itai Ashlagi, Tomohiro Hara, Yoko Okuyama, Masayuki Yagasaki, and all the seminar participants at Happy Hour Seminar!, Tokyo Keizai University, and the University of British Columbia for helpful comments. All remaining errors are our own.

†Leonard N. Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012, United States. E-mail: junpei.komiyama@gmail.com.

‡Vancouver School of Economics, University of British Columbia, 6000 Iona Dr, Vancouver, BC V6T 1L4 Canada. E-mail: shunya.noda@gmail.com. Noda has been supported by the Social Sciences and Humanities Research Council of Canada.

1 Introduction

Statistical discrimination refers to discrimination against minority people, taken by fully rational and non-prejudiced agents.[1] In contrast to taste-based discrimination (Becker, 1957), which regards agents' preferences (e.g., racism, sexism) as the primary source of discrimination, the model of statistical discrimination does not assume any preferences towards specific groups. Previous studies have shown that even in the absence of prejudice, discrimination could occur persistently because of various reasons, such as the discouragement of human capital investment (Arrow, 1973; Foster and Vohra, 1992; Coate and Loury, 1993; Moro and Norman, 2004), information friction (Phelps, 1972; Cornell and Welch, 1996), and search friction (Mailath, Samuelson, and Shaked, 2000). The literature has proposed a variety of affirmative-action policies to solve statistical discrimination, and many of them are being implemented in practice.

The contribution of this paper is to articulate a new channel of statistical discrimination—underestimation of minority workers that appears as a consequence of social learning. Most of the extant literature focuses on behaviors of rational agents under an equilibrium where agents have a correct belief about the relationship between observable characteristics and unobservable skills. However, several empirical studies have shown that real-world people often have a biased belief towards minority groups.[2] The aim of this study is to endogenize the evolution of the biased belief and analyze its consequence. In our model, (i) all firms (decision makers) are fully rational and non-prejudiced (i.e., attempt to hire the most productive worker), and (ii) all workers are ex ante symmetric. We show that, even in such an environment, a biased belief could be generated endogenously and persist in the long run.

For instruction, we use the terminology of hiring markets (while our model is applicable to a broader class of situations). In our model, firms make hiring decisions based on observable characteristics, sequentially.[3] However, the true statistical relationship between characteris-

[1] According to Moro's (2009) definition,

> Statistical discrimination is a theory of inequality between demographic groups based on stereotypes that do not arise from prejudice or racial and gender bias.

Although some previous studies in statistical discrimination consider the consequence of exogenously endowed biased beliefs (e.g., Bohren, Haggag, Imas, and Pope, 2019a; Bohren, Imas, and Rosenberg, 2019b; Monachou and Ashlagi, 2019), we additionally require that agents are fully rational.

[2] De Paola, Scoppa, and Lombardo (2010) analyzed an Italian local administration record where a gender quota is introduced in a short period of 1993-1995, which resulted in increased representation of women politicians even after the quota is terminated. Battaglini, Harris, and Patacchini (2020) showed that increased professional exposure of women judges promotes hiring of other women judges and hypothesized that it is due to the reinforced belief of women's professional capabilities. Bohren et al. (2019a) showed that there is a widespread misconception on the mathematical competence of American people.

[3] Observable characteristics can be very informative in one's career. Wang, Zhang, Posse, and Bhasin (2013) showed that the CV is very informative for predicting whether a software engineer switches another

tics and actual skills is not observed directly; thus, firms need to learn it based on the data about past hiring cases. Firms tend to have insufficient data about minority groups because (i) minority groups are literally a "minority" (in terms of the population), and (ii) they have been discriminated against and not hired in history. The lack of data makes it difficult to assess the skills of minority workers. Hence, it tends to be safer and more profitable to hire a majority candidate, whose skill is accurately estimable. This situation persists because no firm is willing to "experiment" with hiring minorities for the purpose of data diversity. Consequently, minority workers may never be hired, and firms may miss many skillful workers from the minority group. We also show that some temporary affirmative-action policies can effectively prevent this form of discrimination.

We develop a *multi-armed bandit model* of social learning, in which many myopic and short-lived firms sequentially make hiring decisions. In each round, a firm faces multiple candidate workers. Each firm wants to hire only one person. Each firm's utility is determined by the hired worker's skill, which cannot be observed directly until employment. However, as in the standard statistical discrimination model, each worker also has an observable characteristic that is associated with the worker's hidden skill. In the beginning, no one knows the precise way to interpret the worker's observable characteristic for predicting that skill. Hence, firms first need to learn the relationship between the characteristic and skill, and then apply the statistical model to evaluate the predicted skill of workers. We assume that, firms submit all the information about their hiring cases to a public database, and therefore, each firm can observe all the past hiring cases (the characteristics and skills of all the workers *actually hired* in the past).

Each worker belongs to a group that represents the worker's gender, race, and ethnicity. We assume that the characteristics of workers who belong to different groups should be interpreted differently. This assumption is realistic. First, previous studies have revealed that the underrepresented groups receive unfairly low evaluations in many places.[4] When the observable characteristic is an evaluation provided by an outside rater, then the characteristic information itself could be biased because of the prejudice of the rater. Second, evaluations may also reflect differences in the culture, living environment, and social system. For example, firms must be familiar with the custom of writing recommendation letters to inter-

senior position in three years.

[4]For example, Trix and Psenka (2003) study letters of recommendation for medical faculty and find that letters written for female applicants differ systematically from those written for male applicants. Hanna and Linden (2012) suggest that students who belong to lower caste (in India) tend receive unfairly lower exam scores. Conversely, as for teaching evaluation, MacNell, Driscoll, and Hunt (2015) and Mitchell and Martin (2018) demonstrate that students rated the male identity significantly higher than the female identity. Hannák, Wagner, Garcia, Mislove, Strohmaier, and Wilson (2017) study online freelance marketplaces and find that gender and race are significantly correlated with worker evaluations.

pret letters correctly.[5] Hence, the observable characteristics (curriculum vitae, exam score, grading report, recommendation letter, teaching evaluation, etc.) may provide very different implications even when their appearances are similar. If firms are impartial and aware of these biases, they should adjust the way they interpret the characteristics, by applying different statistical models for different groups.[6]

Firms are typically less knowledgeable about minority workers. In many cases, discriminated groups are literally "minorities," and therefore, the number of candidate workers itself tends to be smaller. Furthermore, even when discriminated groups are demographically a majority, they might not have been hired in the past due to a historical reason. Hence, compared with majority workers, the data about minority workers are often insufficient.

The lack of data results in inaccurate prediction of minority workers' skills, and the inaccuracy discourages firms from hiring the minorities. Many workers apply for each job opening. To get hired, a worker must have the highest predicted skill. Once the minority group is underestimated, it is difficult for a minority worker to appear to be the best candidate worker—even if the true skill is the highest, the firm will not be convinced. Underestimation rarely happens once society acquires a sufficiently rich data set. However, in the very beginning of the hiring process, the minority group is underestimated due to bad realizations of the unpredictable component.

The structure described above causes *perpetual underestimation*. Firms tend to hire majority workers because of the imbalance of data richness. However, as long as firms only hire majority workers, society cannot learn about the minority group; thus, the imbalance remains even in the long run. Here, the minority group is perpetually underestimated: the lack of data prevents hiring, and therefore, minority workers are never hired. We prove that, perpetual underestimation may happen with a significantly large probability in our model. Importantly, perpetual underestimation becomes more frequent when the population ratio is imbalanced, as observed in many real-world discrimination problems.

The social discrimination triggered by perpetual underestimation is not only unfair but also socially inefficient. From the welfare perspective, if the time horizon (the total number of hiring opportunities) is sufficiently long, then it is not very costly to "experimenting" with a small fraction of minority workers for learning. However, because firms are selfish and myopic, they are not willing to bear the cost of experiments on their own. Here, *laissez-faire*

[5] Precht (1998) and Al-Ali (2004) report cross-cultural differences in letters of recommendations (that do not originate from discrimination).

[6] Through a randomized experiment, Williams and Ceci (2015) demonstrate that as for the STEM tenure track hiring, female applicants are favored over male applicants. This result is consistent with our assumption here: If the observed characteristics are systematically biased, an impartial employer would debias the data before interpreting it. This may lead to reversal discrimination.

results in the underprovision of a public good—the information about minority groups. By enforcing or incentivizing early movers (firms) to review the minority groups, late movers can refer to a more useful data set of hiring cases, leading to improvement of social welfare. Note that the policy intervention need not be persistent: once sufficiently rich data are collected, the government can terminate the affirmative action and return to laissez-faire.

We analyze the equilibrium consequence of laissez-faire and study desirable policy interventions. Multi-armed bandit models are useful for quantifying the value of information as the width of confidence bounds. We use a linear contextual bandit model to study whether a policy can lead society to achieve "no regret" in the long run. The regret is one of the most popular criteria for evaluating the performance of algorithms in multi-armed bandit problems. The regret measures the welfare loss compared with the first-best decision rule (which firms would take when they had perfect information about the statistical model). When the regret grows sublinearly in N, it means that firms make fair and efficient decisions after a certain time.

In our theoretical analyses of laissez-faire, we first prove that it achieves no regret in the long run: When the groups are ex ante symmetric and the population ratio is equal, the expected regret of laissez-faire is shown to be $\tilde{O}(\sqrt{N})$, where $\tilde{O}$ is a Landau notation that ignores a logarithmic factor. In contrast, when the population ratio is imbalanced (i.e., the number of majority workers is larger than the number of minority workers), this result no longer holds. In such a case, the expected regret is proven to be $\tilde{\Omega}(N)$, which implies that efficiency is not attained even in the long run.

This paper studies two policy interventions towards fair and efficient social learning. The first policy is a subsidy rule, based on the idea of the *upper confidence interval algorithm* (UCB). UCB is an effective solution for balancing exploration and exploitation in the (standard single-agent) multi-armed bandit problem (Lai and Robbins, 1985; Auer, Cesa-Bianchi, and Fischer, 2002). By incentivizing firms to take actions that are consistent with the recommendation of UCB, we can lead the social learning to no regret in the long run. We achieve it by providing firms subsidies when they hire a worker who belongs to an underexplored group. The subsidy is adjusted to the degree of information externality; thus, its total amount shrinks as time goes. Formally, we show that the UCB mechanism has expected regret of $\tilde{O}(\sqrt{N})$. The subsidy amount required for implementing the UCB mechanism is also $\tilde{O}(\sqrt{N})$.

This paper further proposes a *hybrid mechanism*, which terminates affirmative actions once a sufficiently rich data set is collected and returns to laissez-faire. In our setting, once firms obtain a certain amount of data, the diversity of workers' characteristics naturally promotes learning about the minority group. Hence, even if we terminate the policy intervention

earlier than a standard UCB algorithm would do, society rarely falls into perpetual underestimation. We prove that our hybrid mechanism achieves $\tilde{O}(\sqrt{N})$ regret with $\tilde{O}(1)$ subsidy in N rounds. Furthermore, in our simulation, the hybrid mechanism achieved smaller regret than the UCB mechanism.

The second policy is the Rooney Rule.[7] The Rooney Rule is a "soft" affirmative action (in that no hiring quota is required) and it does not require monetary compensation. Instead, the Rooney Rule requires each firm to select at least one minority candidate as a finalist for each job opening. In the final selection, firms can obtain additional signals, besides the observable characteristics shown as an application document. The Rooney Rule leaves minority workers an opportunity to be hired. Even when a firm underestimates a minority worker's skill in the beginning (due to the prediction inaccuracy), the worker may turn out to be the most attractive candidate once the interview is done. As long as minority workers have a chance to be hired, perpetual underestimation will not occur. However, our analysis also shows that the Rooney Rule may hinder hiring of skilled majority candidates, and therefore, should not be adopted as a permanent policy.

The remainder of this paper is organized as follows. Section 2 reviews the literature. Section 3 introduces the model. Section 4 studies the equilibrium consequence of laissez-faire. Section 5 develops the upper-confidence-bound subsidy rules, and Section 6 improves it further. Section 7 studies the two-stage model and analyze the performance of laissez-faire and the Rooney Rule. Section 8 exhibits the simulation results. Section 9 discusses the connection to a Bayesian setting. Section 10 describes this paper's contribution to the multi-armed bandit literature. In Section 11, we make concluding remarks.

2 Related Literature

A survey by Fang and Moro (2011) classifies the literature of statistical discrimination broadly into two strands. The first strand, originates from Arrow (1973), assumes that groups are ex ante identical and analyzes how statistical discrimination occurs as an asymmetric equilibrium (e.g., Foster and Vohra; Coate and Loury, 1992; 1993; Mailath et al., 2000; Moro and Norman, 2003, 2004; Gu and Norman, 2020). This strand interprets statistical discrimination as a random selection of multiple equilibria and does not explain why demographic minorities tend to be discriminated against. The second strand of the literature, originates from Phelps (1972), studies discrimination triggered by unexplained exogenous

[7]The Rooney Rule is originally introduced to the National Football League, and the original version of the rule required league teams to interview ethnic-minority candidates for head coaching and senior football operation jobs. The rule is named after Dan Rooney, the former chairman of the league's diversity committee (Eddo-Lodge, 2017).

differences between groups, coupled with incomplete information about workers' skills (e.g., Aigner and Cain, 1977; Lundberg and Startz, 1983; Cornell and Welch, 1996). The difference in the signal distribution of workers' skill is one of the most popular assumptions in this strand. This paper unifies these two strands in that we endogenize the difference in the signal distribution. We consider otherwise ex ante identical individuals from different groups.[8] Using a social learning model, we demonstrate how the difference in the prediction of skills is generated and persists. We find that when the population ratio of a group is small, the group tends to be statistically discriminated against. Hence, in contrast to most papers in the first strand, our result indicates that a minority group tends to suffer as an inevitable consequence under laissez-faire.

More recently, several works (e.g., Bohren et al., 2019a, 2019b; Monachou and Ashlagi, 2019) demonstrate how misspecified beliefs about groups will result in discrimination. Thus far, this literature has attributed the belief misspecification to psychological bias (e.g., Judd and Park, 1993; Hilton and Von Hippel, 1996) and bounded rationality (e.g., Fryer and Jackson, 2008; Schwartzstein, 2014; Bordalo, Coffman, Gennaioli, and Shleifer, 2019). In contrast, we develop a model of fully rational agents and show that a misspecified belief persists (i.e., a minority group is perpetually underestimated) even in the long run. Our result supports a fundamental assumption of the belief-based statistical discrimination literature.

We model statistical discrimination as a consequence of social learning. The previous studies on social learning, herding, and information cascades have discussed how a sequence of myopic agents reaches an incorrect conclusion with various settings (e.g., Bikhchandani, Hirshleifer, and Welch, 1992; Banerjee, 1992; Smith and Sørensen, 2000). A number of papers have studied the improvement of social welfare through subsidy for exploration (e.g., Frazier, Kempe, Kleinberg, and Kleinberg, 2014; Kannan, Kearns, Morgenstern, Pai, Roth, Vohra, and Wu, 2017) and selective information disclosure (e.g., Kremer, Mansour, and Perry, 2014; Che and Hörner, 2018; Papanastasiou, Bimpikis, and Savva, 2018; Immorlica, Mao, Slivkins, and Wu, 2020; Mansour, Slivkins, and Syrgkanis, 2020). Bohren et al. (2019b) and Monachou and Ashlagi (2019), which are discussed above, study discrimination using social learning models. We develop a novel social learning model for analyzing statistical discrimination and find that, under laissez-faire, society tends to perpetually underestimate workers form a minority group, even when all firms are fully rational and non-prejudiced.

Che, Kim, and Zhong (2019) consider a repeated two-sided matching model where buyers search sellers with the help of rating information about previous transactions. In particular, they analyzed a discriminating equilibrium due to the reinforced inter-group asymmetry of the amount of information due to the search mechanism. This is related to our notion of

[8]The population imbalance is not an "unexplained" difference.

perpetual underestimation in the sense that statistical discrimination occurs in the endogenous process of sampling data. They attribute it to the search process, whereas we attribute it to the imbalanced population. Note also that our bandit-based model is able to evaluate the price of information by balancing exploration and exploitation, and we propose two ways to mitigate perpetual underestimation.

A multi-armed bandit problem stems from the literature of statistics (Thompson, 1933; Robbins, 1952). The theme of this problem is how one long-lived decision maker can maximize his payoff by balancing exploration and exploitation. More recently, the machine learning community has proposed the contextual bandit framework, in which payoffs associated with "arms" (actions) not only depend on the hidden state but are also influenced by additional information, called "contexts" (Abe and Long, 1999; Langford and Zhang, 2008). To study statistical discrimination in labor markets, we adopt the contextual bandit framework because it allows us to capture the diversity of worker characteristics. For readers' convenience, we summarize our technological contribution to the contextual bandit literature in Section 10.

While most of the literature in multi-armed bandit has studied abstract models, Joseph, Kearns, Morgenstern, and Roth (2016) and Kannan et al. (2017) apply contextual bandit frameworks to human-related decision making. These two papers study contextual fairness, which is a stronger concept than our notion of no-regret learning (while these two fairness notions are asymptotically equivalent). Although Kannan et al. (2017) propose a contextually fair UCB-based subsidy rule, it requires an impractically large budget (because contextual fairness is a too stringent requirement with finite horizon). Our subsidy rules also originate from the idea of UCB. However, as we show theoretically and by simulation, our subsidy rules require a much smaller budget. Bardhi, Guo, and Strulovici (2020) study a Bayesian bandit model in which a long-lived employer allocates a task to one of the two workers from different groups.[9] Bardhi et al. (2020) show that a small difference in the prior belief about each worker's type (associated with the group the worker belongs to) could generate a significant difference in payoffs of workers[10]. In contrast, the focus of this paper is on how society acquires a persistent misspecified belief about the minority group endogenously.

Some previous studies consider a linear contextual bandit problem and study the performance of a "greedy" algorithm, which myopically makes decision in accordance with the current information (Bastani, Bayati, and Khosravi, 2020; Kannan, Morgenstern, Roth, Waggoner, and Wu, 2018). As firms take greedy actions under laissez-faire, their results are

[9] As a decision maker is a long-lived employer, Bardhi et al. (2020) belongs to the literature on dynamic employer learning (Farber and Gibbons, 1996; Altonji and Pierret, 2001), not to the *social* learning literature.

[10] Bardhi et al. (2020) also described the effect of population imbalance in an independent section.

also relevant to our model. They show that the greedy algorithm could lead to no regret in the long run, if (i) the contexts (corresponding to workers' characteristics in our model) are diverse enough, and (ii) the decision maker acquires sufficiently many uniform samples in the beginning. While their results suggest that laissez-faire performs well, uniform sampling is not adaptive, and thus does not adequately quantify the value of information. Subsection 8.6 provides a detailed analysis on this point—we show that, our hybrid mechanism performs better than uniform sampling followed by laissez-faire.

There is also increasingly large literature on algorithmic fairness in the field of machine learning. Although most of the papers of this literature were focused on batch learning (i.e., learning from past data to support future hiring), bandit-based sequential learning with algorithmic fairness has been considered in several papers such as Joseph et al. (2016); Raghavan, Slivkins, Vaughan, and Wu (2018); Schumann, Counts, Foster, and Dickerson (2019a); Bechavod, Ligett, Roth, Waggoner, and Wu (2019); Chen, Cuellar, Luo, Modi, Nemlekar, and Nikolaidis (2020). The literature of algorithmic fairness has implicitly assumed exogenous asymmetry in the workers' skill and has been sought affirmative action policies. To this aim, "discrimination-aware" constraints such as demographic parity (Pedreschi, Ruggieri, and Turini, 2008; Calders and Verwer, 2010) and equal opportunity (Hardt, Price, and Srebro, 2016) have been proposed. In contrast, we mainly consider the case where there is no difference in the expected skills among groups. Under such symmetric skill distribution, most of these algorithmic fairness tools are not very meaningful: For example, the demographic parity, that requires the firm to hire workers of each group with equal ratio, is (asymptotically) achieved by any sublinear regret decision rule even in the absence of fairness constraints. Note also that, insights into the long-term effects of these algorithmic fairness policies are still limited (Liu, Dean, Rolf, Simchowitz, and Hardt, 2018).

Theoretical analyses for the Rooney Rule are relatively scarce.[11] Kleinberg and Raghavan (2018) show that, when the recruiter has an unconscious bias against the discriminated group, the Rooney Rule not only helps the representation of the discriminated group but also leads to a higher payoff for the recruiter. To the best of our knowledge, our study is the first attempt to demonstrate the practical advantage of the Rooney Rule without unconscious bias. Our result indicates that even when no unconscious bias exists, the Rooney Rule may improve social welfare by preventing perpetual underestimation.[12]

[11] De Paola et al. (2010) empirically show that a gender quota imposed on an Italian local administration broke down negative stereotypes towards women even after it was terminated. This quota can be regarded as a version of the Rooney Rule.

[12] While there is literature on the tiered interviewing problem (such as Schumann, Lang, Foster, and Dickerson, 2019b), their focus is on an optimization of a single firm's profit.

3 Model

Basic Setting We develop a linear contextual bandit problem with myopic agents (firms). We consider a situation where N firms (indexed by $n = 1, \dots, N$) sequentially hire one worker for each. In each round n, a set of workers $I(n)$ arrives. Each worker $i \in I(n)$ takes no action, and firm n selects one worker $\iota(n) \in I(n)$. We denote the set of all workers by $I \coloneqq \bigcup_{n=1}^{N} I(n)$. Both firms and workers are short-lived. Once round n is finished, firm n's payoff is finalized, and all the workers not hired leave the market.

Each worker i belongs to a group $g \in G$. We assume that the population ratio is fixed: for every round n, the number of arrived workers who belong to group g is $K_g \in \mathbb{N}$ and $K = \sum_{g \in G} K_g$. Slightly abusing the notation, we denote the group worker i belongs to by $g(i)$. Each worker $i \in I$ also has an observable characteristic $\boldsymbol{x}_i \in \mathbb{R}^d$, where $d \in \mathbb{N}$ is its dimension. Finally, each worker i also has a skill $y_i \in \mathbb{R}$, which is not observable until worker i is hired. The characteristics and skills are random variables.

Because each firm's payoff is equal to the hired worker's skill y_i (plus the subsidy assigned to worker i as an affirmative action, if any), firms want to predict the skill y_i based on the characteristics $\boldsymbol{x}_i$. We assume that the characteristics and skills are associated in the following way:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)} + \epsilon_i,$$

where $\boldsymbol{\theta}_g \in \mathbb{R}^d$ is a *coefficient parameter*, and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d. is an unpredictable error term. We assume $||\boldsymbol{\theta}_g|| \leq S$ for some $S \in \mathbb{R}_+$, where $||\cdot||$ is the standard L2-norm. Since ϵ_i is unpredictable,

$$q_i \coloneqq \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)} \tag{1}$$

is the best predictor of worker i's skill y_i.

The coefficient parameters $(\boldsymbol{\theta}_g)_{g \in G}$ are unknown in the beginning. Hence, unless firms share the information about past hiring cases, firms are unable to predict each worker's skill y_i. We assume that all firms share information about hiring cases. Accordingly, when firm n makes a decision, besides current workers' characteristics and groups $(\boldsymbol{x}_i, g(i))_{i \in I(n)}$, firm n can observe all the past candidate workers' characteristics and groups, $(\boldsymbol{x}_i, g(i))$ for all $i \in \bigcup_{n'=1}^{n-1} I(n')$, the past firms' decisions $(\iota(n'))_{n'=1}^{n-1}$, and the past hired workers' skills $(y_{\iota(n')})_{n'=1}^{n-1}$. We refer to all of the realizations of these variables as the *history* in round n, and denote it by $h(n)$. Formally, $h(n)$ is given by

$$h(n) = \left((\boldsymbol{x}_i, g(i))_{i \in I(n)}, (\boldsymbol{x}_i, g(i))_{i \in \bigcup_{n'=1}^{n-1} I(n')}, (\iota(n'))_{n'=1}^{n-1}, (y_{\iota(n')})_{n'=1}^{n-1} \right).$$

Note that, $h(n)$ does not include the information about (i) the worker hired in firm n, and (ii) that worker's actual skill. This is because $h(n)$ represents the information set firm n faces when it makes a hiring decision. We define the set of all histories in round n as $H(n)$. We define the set of all histories as $H \coloneqq \bigcup_{n=1}^{N} H(n)$. The firm's decision rule for hiring and the government's subsidy rule will be defined as a function that maps a history to a hiring decision and the subsidy amount (described later). For notational convenience, we often drop $h(n)$.

Prediction We assume that firms are not Bayesian but *frequentist*. Hence, firms have no prior distribution but they estimate the true parameter $\boldsymbol{\theta}$ using the available data set. (We would obtain a similar result in a Bayesian setting. See the discussion in Section 9.)

We assume that each firm predicts the skill by using *ridge regression* (also known as $l2$-regularized least square) to stabilize the small-sample inference. Let $N_g(n)$ be the number of rounds at which group-g workers are hired by round n. Let $\boldsymbol{X}_g(n) \in \mathbb{R}^{N_g(n)\times d}$ be a matrix that lists the characteristics of group-g workers hired by round n: each row of $\boldsymbol{X}_g(n)$ corresponds to $\{\boldsymbol{x}_{\iota(n')} : \iota(n') = g\}_{n'=1}^{n-1}$. Likewise, let $Y_g(n) \in \mathbb{R}^{N_g(n)}$ be a vector that lists the skills of group-g workers hired by round n: each element of $Y_g(n)$ corresponds to $\{y_{\iota(n')} : \iota(n') = g\}_{n'=1}^{n-1}$. We define $\boldsymbol{V}_g(n) \coloneqq (\boldsymbol{X}_g(n))'\boldsymbol{X}_g(n)$. For a parameter $\lambda > 0$, we define $\bar{\boldsymbol{V}}_g(n) = \boldsymbol{V}_g(n) + \lambda \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ denotes the $d \times d$ identity matrix. Firm n estimates the parameter as follows:

$$\hat{\boldsymbol{\theta}}_g(n) \coloneqq (\bar{\boldsymbol{V}}_g(n))^{-1}(\boldsymbol{X}_g(n))'Y_g(n). \tag{2}$$

Unlike ordinary least square (OLS), for $\lambda > 0$ the inverse $(\bar{\boldsymbol{V}}_g(n))^{-1}$ is always well-defined. Firm n predicts worker i's skill by (1), while substituting $\boldsymbol{\theta}_g$ with $\hat{\boldsymbol{\theta}}_g(n)$:

$$\hat{q}_i(n) \coloneqq \boldsymbol{x}_i'\hat{\boldsymbol{\theta}}_{g(i)}(n).$$

Note that, both $\hat{q}_i(n)$ and $\hat{\boldsymbol{\theta}}_g(n)$ depend on the history $h(n)$. We often drop $h(n)$ for notational simplicity.

Mechanism Besides the (predicted) skill of workers, firms also take the subsidies provided as affirmative actions into consideration. We assume that firms' preferences are risk-neutral and quasi-linear. Hence, if firm n hires worker i, firm n's payoff (von-Neumann–Morgenstern utility) is given by $y_i + s_i$, where where $s_i \in \mathbb{R}_+$ denotes the amount of the subsidy assigned to worker i.

In the beginning of the game, the government commits to a *subsidy rule* $s_i(n, \cdot) : H \to \mathbb{R}_+$,

which maps a history to a subsidy amount. Hence, once a history $h(n)$ is specified, firm n can identify the subsidy assigned to each worker $i \in I(n)$. Firm n attempts to maximize

$$\mathbb{E}\left[y_i + s_i(n; h(n)) \middle| h(n)\right] = \hat{q}_i(n; h(n)) + s_i(n; h(n)).$$

Firm n's *decision rule* $\iota(n, \cdot) : H(n) \to I(n)$ specifies the worker firm n hires given a history $h(n)$. We say that, a decision rule ι is *implemented* by a subsidy rule s_i if for all n, for all $h(n)$, we have

$$\iota(n; h(n)) = \arg\max_{i \in I(n)} \left\{\hat{q}_i(n; h(n)) + s_i(n; h(n))\right\}. \tag{3}$$

We call a pair of a decision rule and subsidy rule a *mechanism*.

Throughout this paper, any ties are broken in an arbitrary way. Again, we often drop $h(n)$ from the input of decision rule ι when it does not cause confusion.

Remark 1 (Observability of the Past Hiring Data)**.** While we assume that firms share the entire history of past hiring data for simplicity, practically, each firm may have limited access to the database. Even if we make such an assumption, our analysis and results will not require qualitative changes. Rational firms estimate $\boldsymbol{\theta}_g$ based on the available data and use it to predict workers' skills. The smaller the sample size of the available data is, the severer the data insufficiency of minority workers is.

Social Welfare We measure social welfare by the smallness of *regret*, which is the standard measure to evaluate the performance of algorithms in multi-armed bandit models. The regret is defined as follows:

$$\mathrm{Reg}(N) \coloneqq \sum_{n=1}^{N} \left\{\max_{i \in I(n)} q_i - q_{\iota(n)}\right\}.$$

Since ϵ_i is unpredictable, it is natural to evaluate the performance of the algorithm (or the equilibrium consequence of the policy intervention) by checking the value of predictors q_i. If the parameter $(\boldsymbol{\theta}_g)_{g \in G}$ were known, each firm could easily calculate q_i for each worker i and choose $\iota(n) = \arg\max_{i \in I(n)} q_i$. In this case, the regret would become zero. However, since $(\boldsymbol{\theta}_g)_{g \in G}$ is unknown, it is too demanding to aim at zero regret. The goal of the policy design is to set up a mechanism that minimizes the expected regret $\mathbb{E}[\mathrm{Reg}(N)]$, where the expectation is taken on a random draw of the workers.[13] This aim is equivalent to maximizing the sum of the skills of the hired workers.

[13] All the mechanisms proposed in this paper are deterministic.

Algorithm 1 Initial Sampling Phase

$\{g_n\}_{n=1}^{N^{(0)}}$ is allocated such that $\sum_{n=1}^{N^{(0)}} \mathbf{1}[g_n = g] = N_g^{(0)}$.
for $n = 1, \cdots, N^{(0)}$ **do**
  Hire $\iota(n; h(n)) = \min_{i \in I(n): g(i) = g_n} i$. ▷ Firm n blindly hires a group-g_n candidate.
end for

Following the literature, we mainly evaluate the performance by the limiting behavior (order) of expected regrets. One useful benchmark is whether the expected regret is linear (i.e., $\mathbb{E}[\mathrm{Reg}(N)] = \Omega(N)$) or sublinear (i.e., $\mathbb{E}[\mathrm{Reg}(N)] = o(N)$).[14] As we described above, once $(\boldsymbol{\theta}_g)_{g \in G}$ is known, firms can use the best predictor q_i to evaluate workers. After that point, regret does not increase. Although $(\boldsymbol{\theta}_g)_{g \in G}$ is unknown ex ante, firms can learn it from the data. A linear regret means that society fails to learn the underlying parameter $(\boldsymbol{\theta}_g)_{g \in G}$, and therefore, firms are hiring less-skilled workers even in the long run. In our model, perpetual underestimation is often a consequence of statistical discrimination—typically, minority workers are more likely to be underexplored, and therefore, they are unfairly rejected.

Budget Some of the policies we study incentivize exploration by subsidization. The total budget required by a subsidy rule is also an important policy concern. The total amount of the subsidy is given by

$$\mathrm{Sub}(N) \coloneqq \sum_{n=1}^{N} s_{\iota(n)}(n).$$

Initial Sampling Phase For analytical tractability, we assume that for the first $N^{(0)}$ rounds, each firm n is forced to hire from a pre-specified group, g_n. We refer to the first $N^{(0)}$ rounds as the *initial sampling phase* (Algorithm 1). Namely, for all $n = 1, \ldots, N^{(0)}$, firm n hires a group-g_n candidate who has the smallest agent number:

$$\iota(n; h(n)) = \min_{i \in I(n)} i \qquad \text{subject to } g(i) = g_n. \tag{4}$$

Choosing an agent who has the smallest number is just a random choice. Whenever agents belong to the same group, their characteristics and skill distributions are the same. Accordingly, (4) is equivalent to choosing a group-g_n worker blindly (i.e., uniformly at random without looking at workers' predicted skills). We define $N_g^{(0)} \coloneqq \sum_{n=1}^{N^{(0)}} \mathbf{1}[g_n = g]$ as the data size of initial sampling for group g. The initial sampling phase is exogenous and not regarded

[14] In the literature of the multi-armed bandit problem, sublinear regret is also referred as *no regret* since the regret per round approaches zero as $N \to \infty$.

Algorithm 2 Laissez-Faire

Complete the initial sampling phase by running Algorithm 1.
for $n = N^{(0)} + 1, \cdots, N$ **do** ▷ Laissez-Faire starts.
  Offer $s_i(n) = 0$ for all $i \in I(n)$. ▷ No Subsidy is provided.
  Firm n hires $\iota(n) = \arg\max_i \boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_{g(i)}(n)$ as an equilibrium consequence.
end for

as a part of the mechanism. Hence, we ignore the incentives and payoffs of firms hiring in the initial sampling phase.

4 Laissez-Faire

This section analyzes the equilibrium under *laissez-faire*, that is, the consequence of social learning when policy intervention is absent. Subsection 4.1 introduces a basic fact: laissez-faire has linear regret in a general domain. However, a general domain is not suitable for the analysis of statistical discrimination. Hence, in Subsection 4.2, we define a symmetric and diverse environment, with which we can discuss how statistical discrimination grows. In Subsection 4.3, we formally define perpetual underestimation and discuss its implications. Subsection 4.4 describes the case where (i) both of the groups have sufficient variation, and (ii) the population ratio is balanced. In this case, the underestimation of minority groups is spontaneously resolved, and therefore, laissez-faire performs well. However, as shown in Subsection 4.5, when the population ratio is imbalanced, laissez-faire tends to result in perpetual underestimation, and therefore, performs poorly.

4.1 Preliminary: Failure in a General Domain

We first define laissez-faire.

Definition 1 (Laissez-Faire)**.** The *laissez-faire decision rule* always selects the worker who has the highest predicted skill, i.e.,

$$\iota(n) = \arg\max_{i \in I(n)} \hat{q}_i(n).$$

Clearly, the laissez-faire decision rule is implemented by the *laissez-faire subsidy rule*, which provide no subsidy $s_i = 0$ after any history (Algorithm 2).

Laissez-faire makes no intervention, and therefore, each firm hires the worker whose expected skill, predicted by the current data set, is the highest. In the multi-armed bandit literature, the laissez-faire decision rule is referred to as the *greedy algorithm*. The greedy algorithm often results in a catastrophic outcome due to insufficient exploration. Since information is a public good, its supply is inefficiently low if the government makes no policy intervention. This well-known result applies to our environment if no structure is assumed. We state this basic result as a benchmark.

Theorem 1 (Failure of Laissez-Faire in General Domain)**.** Let $\mathrm{Reg}^{\mathrm{LF}}$ be the regret under the laissez-faire decision rule. There exists an instance with which

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{LF}}(N)] = \Omega(N).$$

Proof. See Appendix B.2. □

The analysis in Appendix B.2 is essentially the same as the analysis of greedy algorithm in the standard K-armed bandit problem, which is well-known to be $\Omega(N)$. We show Theorem 1 by constructing an instance explicitly. By assuming that the distribution of the characteristics $(\boldsymbol{x}_i)$ to be degenerate, our linear contextual bandit problem reduces to a basic K-armed bandit problem, where the expected skill (reward) of each group (arm) is fixed. We assume that one group is more productive than another, and therefore, the first-best decision rule would always hire from the better group. With a constant probability, firms happen to underestimate the more productive group in the beginning. When a less productive group constantly performs better than the underestimated predicted skill of the better group, firms never want to investigate the better group further. Consequently, with a significant probability, a worker from the better group is never hired again, implying linear expected regret. Once an underestimation of the minority group occurs, it tends to persist: When the "context" of the majority group is fixed, there is a constant probability that the minority group is never chosen throughout all the rounds.

4.2 Symmetry and Diverse Characteristics

It is too naive to conclude from Theorem 1 that the laissez-faire decision rule may cause statistical discrimination. First, the instance constructed in the proof of Theorem 1 assumes an unexplained exogenous difference (in expected skills) between groups, while our aim is to endogenize the difference. Second, we assumed that one group has higher expected skill than the other. With this assumption, it is efficient to always hire a worker from one group. Under such an assumption, when social learning is successful, workers from the inferior group

are never hired. Third, we reduced a contextual bandit model to a K-armed bandit model by assuming that the distribution of characteristics is degenerate. However, in the real word, candidate workers have diverse characteristics, even when they belong to the same group.

To provide better analysis for the laissez-faire decision rule, we make the following three assumptions. First, we focus on the case of two groups.

Assumption 1 (Two Groups)**.** The population consists of two groups $G = \{1, 2\}$.

When we consider asymmetric groups and equilibria, we refer to group 1 as a majority (dominant) group and group 2 as a minority (discriminated) group. The two-group assumption helps us to elucidate how the minority group is discriminated against by the majority group.

Second, we assume that groups are symmetric.

Assumption 2 (Symmetric Groups)**.** The characteristics of all groups are identically distributed, and the coefficient parameters are the same across all groups. Namely, a probability distribution F such that for all $i \in I$,

$$\boldsymbol{x}_i \sim F,$$

and there exists $\boldsymbol{\theta} \in \mathbb{R}^d$ such that for all $g \in G$,

$$\boldsymbol{\theta}_g = \boldsymbol{\theta}.$$

Note that although we assume that groups are symmetric, firms do not see them as symmetric, and therefore, apply different statistical models for different groups. In other words, even though the *true* coefficients are identical ($\boldsymbol{\theta}_g = \boldsymbol{\theta}'_g$ for all $g, g' \in G$), firms estimate them separately; thus, the values of the *estimated* coefficients are typically different ($\hat{\boldsymbol{\theta}}_g(n) \neq \hat{\boldsymbol{\theta}}_{g'}(n)$ for $g \neq g'$).

Although Assumption 2 is unrealistic (as it is evident that the characteristics should be interpreted differently), it is useful for elucidating how laissez-faire nourishes statistical discrimination. Under Assumption 2, there is no ex ante difference between groups (as assumed in Arrow, 1973; Foster and Vohra, 1992; Coate and Loury, 1993; Moro and Norman, 2004, etc.). Hence, all the differences we observe in the equilibrium consequence are purely due to the property of the equilibrium learning process.

Under Assumption 2, statistical discrimination implies inefficiency: although a best candidate belongs to a minority group with substantial probability (K_2/K), that candidate is not hired due to underexploration. Hence, when the groups are symmetric, the resolution of statistical discrimination make the hiring process not only fair but also efficient. By contrast, when there is exogenous asymmetry between groups, fairness and efficiency are often

conflicting. For example, *demographic parity* is one of the most popular fairness notions studied in machine learning (or supervised learning) literature. In our model, the demographic parity requires that the probability of hiring from the minority group is equal to the population ratio; i.e., K_2/K. Clearly, when the groups are asymmetric, the "first-best decision rule" does not satisfy this condition—it hires more from a "more productive group," while it is arguable that such a decision rule is socially desirable. As long as we assume the group symmetry, our argument avoids this controversy: The first-best decision rule is fair and efficient. Thus, we should attempt to approximate it.

Third, we assume that characteristics are normally distributed, and therefore, the distribution is non-degenerate. This assumption captures the diversity of workers, which is the nature of the real-world labor market.

Assumption 3 (Normally Distributed Characteristics)**.** For every candidate i,

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_{xg(i)}, \sigma^2_{xg(i)} \boldsymbol{I}_d),$$

where $\boldsymbol{\mu}_{xg} \in \mathbb{R}^d$ and $\sigma_{xg} \in \mathbb{R}_{++}$ for every $g \in G$. We also denote $\boldsymbol{x}_i = \boldsymbol{\mu}_{xg(i)} + \boldsymbol{e}_{xi}$ to highlight the noise term $\boldsymbol{e}_{xi}$.

Note that when we have both Assumptions 2 and 3, then there exist $\boldsymbol{\mu}_x, \sigma_x$ such that

$$\boldsymbol{\mu}_{xg} = \boldsymbol{\mu}_x,$$
$$\sigma_{xg} = \sigma_x,$$

for all $g \in G$. Hence, $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_x, \sigma_x^2 \boldsymbol{I}_d)$ for all $i \in I$.

4.3 Perpetual Underestimation

To determine whether social learning incurs linear expected regret or not, it is useful to check whether it results in *perpetual underestimation* with a significant probability.

Definition 2 (Perpetual Underestimation)**.** A group g_0 is *perpetually underestimated* if, for all $n > N^{(0)}$, we have $g(\iota(n)) \neq g_0$.

Namely, when group g_0 is perpetually underestimated, no worker from group g_0 is hired after the initial sampling phase.

If social learning results in perpetual underestimation with a significant probability, then it often incurs linear expected regret. In particular, under Assumptions 2, perpetual under-

estimation against any group $g \in G$ implies that firms fail to hire at least

$$\frac{K_g}{K}\left(N - N^{(0)}\right)$$

best candidate workers, which is linear in N. Hence, if the probability of perpetual underestimation is constant (independent of N), then we have linear expected regret.

In our model, perpetual underestimation is also closely related to social discrimination. When perpetual underestimation occurs, a candidate who belongs to an underestimated group is not hired, while groups are symmetric. This outcome happens because society cannot accurately predict the skills of minority workers due to the lack of data. Hence, in our model, perpetual discrimination can be regarded as a form of statistical discrimination.

4.4 Sublinear Regret with Balanced Population

This section analyzes the case where is only one candidate arrives at each round for both groups. In this case, the variation of context implicitly urges the firms to explore all the groups with some frequency. Consequently, laissez-faire has sublinear regret, implying that statistical discrimination is eventually resolved, spontaneously.

Theorem 2 (Sublinear Regret with Balanced Population)**.** Suppose Assumptions 1, 2, and 3. Suppose also that $K_g = 1$ for $g = 1, 2$. Then, the expected regret is bounded as

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{LF}}(N)] \leq C_{\mathrm{bal}}\sqrt{N}$$

where C_{bal} is a $\tilde{O}(1)$ factor that depends on model parameters. Here, $\tilde{O}(1)$ is a Landau notation that ignores polylogarithmic factors.[15] Letting $\mu_x = ||\boldsymbol{\mu}_x||$, the factor C_{bal} is inverse proportional to $\Phi^c(\mu_x/\sigma_x)$, which approximately scales as $\exp(-(\mu_x/\sigma_x)^2/2)$.

Proof. See Appendix B.3. The explicit form of C_{bal} is found at the end of the Appendix B.3. □

The crux of the analysis here is whether the perpetual underestimation is prevented. Assume that group 2 is underestimated, which happens with some constant probability. The ratio μ_x/σ_x represents the stability of characteristics. The larger this value is, the more stable the skill of candidates. If μ_x/σ_x is small, there is some probability such that the skill

[15] Namely, there exists $N_0 \in \mathbb{N}$ and a function $f(N)$ that is finite-order polynomial of $\log N$ such that $\mathbb{E}[\mathrm{Reg}^{\mathrm{LF}}(N)] \leq f(N)$ for all $N \geq N_0$. In this and subsequent theorems, we often ignore polylogarithmic factors (factors that are finite-order polynomial of the logarithm) of N because they grow very slowly as N grows large. We remark on the important dependence on model parameters and refer to the equation of explicit formulae of each factor.

of the group-1 candidate is predicted to be bad. In such a case, the candidate from group 2 might be chosen, which updates the belief about group 2 to resolve underestimation. As is expected by the theory of least squares, the standard deviation of $\hat{\boldsymbol{\theta}}_g(n)$ is proportional to $(\bar{\boldsymbol{V}}_g(n))^{-1/2}$, and we show that its diameter $(\lambda_{\min}(\bar{\boldsymbol{V}}_g(n)))^{-1/2}$ shrinks as $\tilde{O}(1/\sqrt{n})$. The regret per error is defined by this quantity, and the total regret is $\tilde{O}(\sum_{n\leq N}(1/\sqrt{n})) = \tilde{O}(\sqrt{N})$.

Theorem 3 shows that statistical discrimination is resolved spontaneously when the candidate variation is large. At a glance, this appears to be in contradiction with widely known results that states laissez-faire may lead to suboptimal results in bandit problems due to underexploration. Since selfish firms do not want to experiment with underrepresented groups at their own risk, laissez-faire perpetually underestimates the skill of the minority group (as demonstrated in Theorem 1). However, the variation in characteristics naturally incentivizes selfish agents to explore the underestimated group, and therefore, with some additional conditions, we can bound the probability of perpetual underestimation.

Theorem 2 shares some intuitions with the previous results (Kannan et al., 2018; Bastani et al., 2020), which have shown that the variation in contexts (characteristics) improves the performance of the greedy algorithm (laissez-faire) in contextual multi-armed bandit problems. Kannan et al. (2018) assume that there is a sufficiently long initial sampling phase, in which society can collect the uniform-sample data until the model parameters are stabilized. Theorem 1 in Bastani et al. (2020) corresponds to Theorem 2 in our paper, and we further attributes it to the stability μ_x/σ_x rather than the diameter of the characteristics.

More importantly, in the next subsection, we prove that these positive results are "special cases"—we will show that, even when there is a variation in characteristics, when the population ratio imbalanced, the laissez-faire decision rule may cause perpetual underestimation with a substantial probability.

4.5 Large Regret with Imbalanced Population

While Theorem 2 implies that statistical discrimination might be spontaneously resolved in the long run (if we admit that workers' characteristics are diverse enough), it crucially relies on one unrealistic assumption—the balanced population ratio. In many real-world problems, the population ratio is imbalanced. The dominant group is often the majority of the population, and the discriminated is minority. Even when the population demographically balanced, if we look at a specific labor market, the population ratio could be imbalanced due to an imbalanced wealth distribution or discouragement of human capital investments.

We indeed find that the population ratio between groups has a crucial role for the welfare under laissez-faire. In the following theorem, we assume that, in each round, while only one

minority worker arrives (i.e., $K_2 = 1$), while many majority workers ($K_1 > 1$) arrive. When the population is imbalanced, perpetual underestimation becomes more likely, and therefore, society suffers from a large expected regret.

Theorem 3 (Large Regret with Imbalanced Population)**.** Suppose Assumptions 1, 2, and 3. Suppose also that $K_2 = 1$ and $d = 1$. Let $K_1 > \log_2 N$. Then, under the laissez-faire decision rule, group 2 is perpetually underestimated with the probability at least $C_{\text{imb}} = \tilde{\Theta}(1)$. Accordingly, the expected regret of the laissez-faire decision rule is

$$\mathbb{E}\left[\text{Reg}^{\text{LF}}(N)\right] \geq \frac{C_{\text{imb}}(N - N^{(0)})}{K} = \tilde{\Omega}(N).$$

Proof. See Appendix B.4. The explicit form of C_{imb} is found at Eq. (39). □

In the proof of Theorem 3, we evaluate the probability of the following two events occuring. (i) The coefficient parameter for the minority candidates, $\boldsymbol{\theta}_2$ is underestimated. (ii) The characteristics and skills of the hired majority workers are consistently good throughout the rounds. The probability of (i) is constant (independent of N) and the probability of (ii) is constant if $K_1 > \log_2 N$. When both (i) and (ii) occur, minority workers are perpetually underestimated, and therefore, we have a large regret.

Theorem 3 indicates that we should not be too optimistic about the consequence of laissez-faire. The imbalance in the population ratio naturally favors the majority group by helping society to collect a richer data set about them, leading to statistical discrimination. This insight applies to many real-world problems because an imbalanced population is commonplace.

Remark 2. In the proof of Theorem 3, we explicitly bound the probability that each event happens. Hence, the effect of initial sample size is revealed. The probability of underestimating the minority group is exponentially small to the number initial samples for minorities, $N_2^{(0)}$, which implies that small number of initiators in the minority group can prevent the underestimation to be perpetuated.[16] Note also that this is consistent with the prior results by Kannan et al. (2018), which state a sufficiently large initial samples prevent perpetual underestimation because it alleviates the underestimation of $\hat{\boldsymbol{\theta}}_2$. In Subsection 8.6, we demonstrate that this solution is not desirable because uniform sampling is costly and difficult to implement. According to our simulation, the UCB-based subsidy rule (the hybrid mechanism, proposed in Section 6) outperforms uniform sampling followed by laissez-faire.

Remark 3. In our framework, the statistical distribution purely attributes to a failure of social learning and the resultant misinformation. Hence, even when perpetual underesti-

[16] See Lemma 24 (in the Appendix) for the full detail.

mation occurs, the true skills of minority workers (the distribution of y_i) are not lowered. However, if we additionally incorporate the choice of the education level and human capital investments (as in Foster and Vohra, 1992; Coate and Loury, 1993), the misinformation naturally discourages minority workers from improving their skills. Therefore, if education is endogenous, the welfare loss and inequality raised by social learning would be more serious.

5 The Upper Confidence Bound Mechanism

Section 4 has discussed the equilibrium consequence under laissez-faire. We observed that, when the population ratio is imbalanced (as in the real-world job market), there is a substantial probability that the underestimation is perpetuated. This result indicates that a policy intervention (affirmative action) is effective for improving social welfare and fairness of the hiring market.

This section proposes a subsidy rule to resolve such a perpetual underestimation. We use the idea of the *upper confidence bound* (UCB) algorithm, which is widely used in the literature of the bandit problem.[17] The UCB algorithm balances exploration and exploitation by allocating handicaps to less explored arms (groups), whose rewards (skills) cannot be predicted accurately. The UCB algorithm develops a confidence interval for the true reward and evaluate each arm's performance by its upper confidence bound to achieve this balance. Although firms are not willing to follow the UCB's recommendation under laissez-faire, the government can provide a subsidy to promote a candidate worker who has the highest UCB. In this section, we establish a UCB-based subsidy rule and evaluate its performance.

5.1 The UCB Decision Rule

To establish the UCB-based subsidy rule, we first define the hiring decision suggested by the UCB algorithm. After that, we construct a subsidy rule that incentivizes firms to hire workers based on UCB. A challenge is that the adaptive selection of the candidates based on history can induce some bias, and the standard confidence bound no longer applies to our case. To overcome this issue, we use martingale inequalities (Peña, Lai, and Shao, 2008; Rusmevichientong and Tsitsiklis, 2010; Abbasi-Yadkori, Pál, and Szepesvári, 2011). We here introduce the confidence interval for the true coefficient parameter, $(\boldsymbol{\theta}_g)_{g\in G}$.

[17] The idea of UCB goes back to at least in 1980s. The seminal paper by Lai and Robbins (1985) analyzed a version of UCB. More recently, Auer et al. (2002) introduced UCB1, which is widely known in the machine learning literature.

Definition 3 (Confidence Interval, Abbasi-Yadkori et al., 2011)**.** Given the group g's collected data matrix $\bar{\boldsymbol{V}}_g(n)$, the *confidence interval* of group g's coefficient parameter $\boldsymbol{\theta}_g$ is given by

$$\mathcal{C}_g(n) = \left\{ \bar{\boldsymbol{\theta}}_g \in \mathbb{R}^d : \left\| \bar{\boldsymbol{\theta}}_g - \hat{\boldsymbol{\theta}}_g(n) \right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \sigma_\epsilon \sqrt{d \log \left(\frac{\det(\bar{\boldsymbol{V}}_g(n))^{1/2} \det(\lambda \boldsymbol{I}_d)^{-1/2}}{\delta} \right)} + \lambda^{1/2} S \right\}$$

where $||\boldsymbol{v}||_{\boldsymbol{A}} = \sqrt{\boldsymbol{v}' \boldsymbol{A} \boldsymbol{v}}$ for a d-dimensional vector $\boldsymbol{v}$ and $d \times d$ matrix $\boldsymbol{A}$.

The standard confidence interval, $\mathcal{C}_g(n)$, shrinks as firm n has a richer set of data about group g. Abbasi-Yadkori et al. (2011) study the property of this confidence interval, and they prove that the true parameter $\boldsymbol{\theta}_g$ lies in $\mathcal{C}_g(n)$ with a probability $1 - \delta$ (Lemma 19). If we choose sufficiently small δ,[18] it is "safe" to assess that worker i's skill is *at most*

$$\tilde{q}_i(n) \coloneqq \max_{\bar{\boldsymbol{\theta}}_{g(i)} \in \mathcal{C}_{g(i)}(n)} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}_{g(i)}.$$

We call $\tilde{q}_i(n)$ the *upper confidence bound index* (UCB index) of worker i's skill. Intuitively, $\tilde{q}_i(n)$ is worker i's skill in the most optimistic scenario. The UCB decision rule makes a decision based on this UCB index.

Definition 4 (UCB Decision Rule)**.** The UCB decision rule selects the worker who has the highest UCB index; i.e.,

$$\iota(n) = \arg\max_{i \in I(n)} \tilde{q}_i(n). \tag{5}$$

The UCB index $\tilde{q}_i(n)$ is close to the pointwise estimate $\hat{q}_i(n)$ when society has a rich data about group $g(i)$, because $\mathcal{C}_{g(i)}(n)$ is small in such a case. However, when the information about group $g(i)$ is insufficient, $\tilde{q}_i(n)$ is much larger than $\hat{q}_i(n)$, because the firm is not sure about the true skill of worker i and $\mathcal{C}_{g(i)}(n)$ is large. In this sense, the UCB decision rule offers affirmative actions to underexplored groups. In contrast to the greedy algorithm (laissez-faire), the UCB algorithm appropriately balances exploration and exploitation, and therefore, it has a sublinear expected regret in general environments.

The UCB decision rule recommends the exploration of majority candidates as well as minority candidates. The amount of the subsidy is proportional to the uncertainty of the candidate's characteristic, which is represented by the confidence interval $\mathcal{C}_g(n)$. The confidence interval $\mathcal{C}_g(n)$ is inverse proportional to $\bar{\boldsymbol{V}}_g(n) = (\boldsymbol{X}_g(n))' \boldsymbol{X}_g(n) + \lambda \boldsymbol{I}_d$.[19] Hence, if the data $\boldsymbol{V}_g(n)$ do not have a large variation in a particular dimension of $\boldsymbol{x}_i$, then the prediction

[18] We typically choose $\delta = 1/N$ so that the confidence interval is asymptotically correct in the limit of $N \to \infty$.

[19] The standard ordinary least square has a confidence bound of the form $\boldsymbol{\theta}_g - \hat{\boldsymbol{\theta}}_g(n) \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{V}_g^{-1}(n))$

from that dimension can be inaccurate. In such a case, the UCB decision rule recommends hiring a candidate who contributes to increasing the data variation for that dimension. For example, when a candidate has some skills that previous candidates do not have, then the candidate's UCB index tends to become large.

As the UCB decision rule nicely balances exploration and experimentation, it has a sublinear regret for a general environment.

Theorem 4 (Sublinear Regret of UCB)**.** Suppose Assumptions 3. Let $\mathrm{Reg}^{\mathrm{UCB}}$ be the regret from the UCB decision rule. Let $\lambda \geq \max(1, L^2)$. Then, by choosing sufficiently small δ, the regret under the UCB decision rule is bounded as

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{UCB}}(N)] \leq C_{\mathrm{ucb}}\sqrt{N},$$

where C_{ucb} is a $\tilde{O}(1)$ factor to N that depends on model parameters.

Proof. See Appendix B.5. The explicit form of C_{ucb} is found in Eq. (43) therein. □

Note that, $\tilde{O}(\sqrt{N})$ regret is the optimal rate for these sequential optimization problems under partial feedback (Chu, Li, Reyzin, and Schapire, 2011). Hence, Theorem 4 states that the UCB decision rule effectively prevents perpetual underestimation and is asymptotically efficient. The analysis here does not depend on the size of candidate pool K, and thus effective regardless of the population ratio.

Remark 4. Although we have made several strong assumptions for the analysis of laissez-faire (e.g., two groups, symmetry), Theorem 4 does not rely on them, and therefore, it is applicable to a very general environment. The groups need not be symmetric. The normal characteristic assumption (Assumption 3) can be relaxed to a weaker condition that guarantees that the distributions are light-tailed, or the characteristics can even be arbitrary as long as they are bounded with high probability.

5.2 The UCB Index Subsidy Rule

To implement the UCB decision rule, we need to satisfy the firms' obedience condition (3) along with the UCB's decision rule (5). In this paper, we focus on two types of subsidy rules. One is the *UCB index subsidy rule*, and another is the *UCB cost-saving subsidy rule*.

and thus $|\boldsymbol{\theta}_g - \hat{\boldsymbol{\theta}}_g(n)| \sim \sigma_\epsilon \boldsymbol{V}_g^{-1/2}(n)$. The martingale confidence bound $\mathcal{C}_g(n)$ is larger than OLS confidence bound in two factors because of the price of adaptivity. Namely, (1) $\sqrt{d}$ factor and (2) $\sqrt{\log(\det(\bar{\boldsymbol{V}}_g(n)))}$ factor. As discussed in Xu, Honda, and Sugiyama (2018), the first $\sqrt{d}$ factor unnecessarily overestimates the confidence bound for most cases.

Algorithm 3 The UCB Index Subsidy Rule

Complete the initial sampling phase by running Algorithm 1.
for $n = N^{(0)} + 1, \cdots, N$ **do** ▷ The UCB index subsidy rule starts.
  for i **do**
    Compute $\tilde{q}_i(n) = \max_{\bar{\boldsymbol{\theta}}_{g(i)} \in \mathcal{C}_{g(i)}(n)} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}_{g(i)}$. ▷ Obtain UCB indices.
    Offer $s_i = \tilde{q}_i(n) - \hat{q}_i(n)$ for all $i \in I(n)$. ▷ Align firm n's payoff to the UCB index.
    Firm n hires $\iota(n) = \arg\max_{i \in I(n)} \tilde{q}_i(n)$ as an equilibrium consequence.
  end for
end for

First, we formally define the UCB index subsidy rule. The UCB index subsidy rule induces firms to hire a candidate with the largest UCB index by aligning each firm's profit with the UCB index.

Definition 5 (UCB Index Subsidy Rule)**.** The *UCB index subsidy rule* s subsidizes worker i who arrives in round n by

$$s_i(n; h(n)) = \tilde{q}_i(n; h(n)) - \hat{q}_i(n; h(n)).$$

The formal algorithm is shown as Algorithm 3.

The UCB index subsidy rule is named "index" because it belongs to an *index policy* (Gittins, 1979) in the terminology of the multi-armed bandit literature.

Definition 6 (Index Policy)**.** A subsidy rule s is an *index policy* if for all n and $i \in I(n)$, $s_i(n; \cdot)$ only depends on $\boldsymbol{X}_{g(i)}(n), Y_{g(i)}(n), \boldsymbol{x}_i$.

To be more precise, our definition of the index rule is slightly weaker than the standard definition. A standard definition requires that the index of an arm only depends on the data generated by the arm. However, since we regard a set of arms as a group, it does not make sense to focus on the data generated by "an arm." Hence, we utilize all the data about group $g(i)$. Having said that, our definition requires that the subsidy for worker i is independent of (i) the other agents' characteristics $\boldsymbol{x}_j$ for any $j \in I(n) \setminus \{i\}$ and (ii) the data about other groups, $\boldsymbol{X}_{g'}(n)$ for any $g' \neq g(i)$.

If a subsidy rule is an index policy, the government need not observe the characteristics of $I(n) \setminus \{i\}$ to determine the subsidy assigned to the employment of worker i. This is a practically desirable property: In many real-world problems, it is difficult for the government to observe the characteristics of candidate workers who are not hired.

The following theorem states the property of the UCB index subsidy rule. Among all index subsidy rules that implement the UCB decision rule, the UCB index subsidy rule requires the minimum amount of the subsidy. Its expected amount is proven to be $\tilde{O}(\sqrt{N})$.

Theorem 5 (Sublinear Subsidy of the UCB Index Rule)**.**

1. The UCB index subsidy rule implements the UCB decision rule.

2. The UCB index subsidy rule needs a minimum amount of the subsidies among all subsidy rules that (i) implement the UCB decision rule, and (ii) str an index policy. Formally, let $s^{\text{U-I}}$ be the UCB cost-saving subsidy rule and s be an arbitrary subsidy rule that satisfies (i) and (ii). Then, for all i, n and $h(n)$, we have

$$s_i^{\text{U-I}}(n; h(n)) \leq s_i(n; h(n)).$$

3. Under the same assumptions as Theorem 4, the amount of the subsidy required by the UCB index subsidy rule is bounded as

$$\mathbb{E}[\text{Sub}^{\text{UCB-I}}(N)] \leq C_{\text{ucb}}\sqrt{N}.$$

   where C_{ucb} is an $\tilde{O}(1)$ factor that is the same as Theorem 4.

Proof. See Appendix B.6. □

The square-root subsidy implies that the government can eventually end the subsidy because $\text{Sub}^{\text{UCB-I}}(N)/N \to 0$ as $N \to \infty$. Alternatively, Theorem 5 implies that society can terminate affirmative actions once a sufficiently rich data set about the minority groups is obtained.

5.3 The UCB Cost-Saving Subsidy Rule

If the mechanism does not have to be an index policy (i.e., the subsidy for worker $i \in I(n)$ may depend on $(\boldsymbol{x}_j)_{j \in I(n)}$ of the other candidates), then we can save the budget without modifying the decision rule. To achieve it, we can subsidize the minimum amount such that candidate ι is more profitable than the other candidates. Formally, the UCB cost-saving subsidy rule is defined as follows.

Definition 7 (UCB Cost-Saving Subsidy Rule)**.** For every round n, the *UCB cost-saving subsidy rule* chooses $s_i(n) = 0$ for every $i \in I(n) \setminus \{\iota(n)\}$, where $\iota(n)$ is the candidate worker

Algorithm 4 The UCB Cost-Saving Subsidy Rule

Complete the initial sampling phase by running Algorithm 1.
for $n = N^{(0)} + 1, \cdots, N$ **do** ▷ The UCB cost-saving subsidy rule starts.
  for i **do**
    Compute $\hat{q}_i(n) = \boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_{g(i)}(n)$.
    Compute $\tilde{q}_i(n) = \max_{\bar{\boldsymbol{\theta}}_{g(i)} \in \mathcal{C}_{g(i)}(n)} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}_{g(i)}$.
    Compute $\iota(n) = \arg\max_{i \in I(n)} \tilde{q}_i(n)$. ▷ $\iota(n)$ is the UCB winner.
    Offer $s_{\iota(n)}(n) = \max_{j \in I(n)} \hat{q}_j(n) - \hat{q}_{\iota(n)}(n)$. ▷ Make $\iota(n)$ becomes most profitable.
    Offer $s_j(n) = 0$ for all $j \in I(n) \setminus \{\iota(n)\}$.
    Firm n hires $\iota(n)$ as an equilibrium consequence.
  end for
end for

selected by the UCB algorithm, (5). For $i = \iota(n)$, the subsidy s_i is given by

$$s_i(n; h(n)) = \max_{j \in I(n)} \hat{q}_j(n; h(n)) - \hat{q}_i(n; h(n)).$$

The formal algorithm is shown as Algorithm 3.

The UCB cost-saving subsidy rule subsidizes only the targeted worker, $\iota(n)$. Hence, for other workers $j \neq \iota(n)$, the payoff from the employment is $\hat{q}_j(n)$. The UCB cost-saving subsidy rule sets the subsidy amount $s_{\iota(n)}$ in such a way that the payoff from hiring worker $\iota(n)$, which is $\tilde{q}_{\iota(n)}(n) + s_{\iota(n)}$, is equal to (or slightly larger than) the payoff from hiring the worker who has the highest predicted skill, $\max_{j \in I(n)} \tilde{q}_j(n)$.

Clearly, the UCB cost-saving subsidy rule is the subsidy rule that requires the minimum budget to implement the UCB decision rule. As fines (negative subsidies) are not allowed in our model, the government cannot discourage the employment of the other candidate workers, $j \in I(n) \setminus \{\iota(n)\}$, further. Hence, the UCB cost-saving subsidy rule requires the smallest budget among all subsidy rules that implements the decision rule (5).

Combining this observation with Theorem 5, we obtain the following theorem.

Theorem 6 (Sublinear Subsidy of the UCB Cost-Saving Rule)**.**

1. The UCB cost-saving subsidy rule implements the UCB decision rule.

2. The UCB cost-saving subsidy requires the minimum budget for implementing the UCB decision rule. Formally, let $s^{\text{U-CS}}$ be the UCB cost-saving subsidy rule and s be an arbitrary subsidy rule that implements the UCB decision rule. Then, for all i, n and

$h(n)$, we have

$$s_i^{\text{U-CS}}(n; h(n)) \leq s_i(n; h(n)).$$

3. The amount of the subsidy required by the UCB cost-saving subsidy rule is bounded as

$$\mathbb{E}[\text{Sub}^{\text{UCB-CS}}(N)] \leq \mathbb{E}[\text{Sub}^{\text{UCB-I}}(N)] \leq C_{\text{ucb}}\sqrt{N}.$$

Proof. The first two statements straightforwardly follow from the argument above. The last statement follows from the first two and Theorem 5. □

The cost-saving subsidy rule has some drawbacks. It depends on the characteristics of all the potential candidates. Hence, the government must have precise knowledge about candidates who appeared in each round but were not hired by the firm. Still, as a theoretical benchmark, it is useful to study the minimum subsidy amount incurred. In Subsection 8.4, we compare the index rule and cost-saving rule numerically. Our simulation results indicate that the cost-saving rule outperform the index rule by much in terms of the total amount of subsidy.

6 The Hybrid Mechanism

In the previous section, we showed that the UCB mechanism effectively prevents perpetual underestimation and achieves sublinear regret for general environments. However, the UCB mechanism has one draw back: it assigns subsidies forever. Although the confidence interval $\mathcal{C}_g(n)$ shrinks as n grows large, it does not degenerate to a singleton for any finite n. Accordingly, even for a large n, there remains a gap between expected skill $\hat{q}_i(n)$ and the UCB index $\tilde{q}_i(n)$ (though small in size). This feature is not desirable for the following reasons. First, introducing a permanent policy is often more politically difficult than introducing a temporary policy. If the government declares that hiring of minority workers is *permanently* subsidized, the policy may look quite unfair to the majority group. The appearance of unfairness would cause significant opposition. Second, if we keep distributing subsidies over the long run, the required budget tends to grow. Third, besides the subsidy itself, the permanent allocation of the subsidy comes with (unmodeled) administration costs.

To overcome these limitations of the UCB mechanism, we propose the *hybrid mechanism*, which starts with the UCB mechanism and turns to laissez-faire by terminating the subsidy at some point. We terminate the UCB-phase once the amount of data of the minority group is enough to induce spontaneous exploration. We prove that, our hybrid mechanism has

$\tilde{O}(\sqrt{N})$ regret (as the UCB mechanism does), and its expected total subsidy amount is $\tilde{O}(1)$ (as opposed to $\tilde{O}(\sqrt{N})$ subsidy of UCB).

The construction of the hybrid mechanism is as follows. Let $s_i^{\text{U-I}}(n) = \tilde{q}_i(n) - \hat{q}_i(n)$ be the size of confidence bound. Note that, $s_i^{\text{U-I}}(n)$ corresponds to the amount of the subsidy allocated by the UCB index subsidy rule (Definition 5). The *hybrid index* $\tilde{q}_i^{\text{H}}$ is defined as

$$\tilde{q}_i^{\text{H}}(n; h(n)) := \begin{cases} \tilde{q}_i(n; h(n)) & \text{if } s_i^{\text{U-I}}(n; h(n)) > a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}(n; h(n))||, \\ \hat{q}_i(n; h(n)) & \text{otherwise,} \end{cases} \tag{6}$$

where $a \geq 0$ is the mechanism's parameter.

The hybrid index is literally a "hybrid" of the predicted skill $\hat{q}_i(n)$ and the UCB index $\tilde{q}_i(n)$. If the difference between the UCB index and the predicted skill is larger than the threshold (i.e., $s_i^{\text{U-I}}(n) > a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}(n)||$), the hybrid index is equal to the UCB index $\tilde{q}_i(n)$. The confidence bound $|\tilde{q}_i(n) - \hat{q}_i(n)|$ is large while we have insufficient knowledge about group $g(i)$; this is typically the case in an early stage of the game. Once this gap becomes smaller than the threshold (i.e., $s_i^{\text{U-I}}(n) \leq a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}(n)||$), then the hybrid index becomes equal to the predicted skill $\hat{q}_i(n)$.

Naturally, the hybrid decision rule is defined as the rule that hires the highest hybrid index.

Definition 8 (The Hybrid Decision Rule)**.** The *hybrid decision rule* selects the worker who has the highest hybrid index; i.e.,

$$\iota^{\text{H}}(n; h(n)) = \arg\max_{i \in I(n)} \tilde{q}_i^{\text{H}}(n; h(n)).$$

As the hybrid decision rule is a hybrid of the UCB decision rule and the laissez-faire decision rule, it can be implemented by mixing the laissez-faire subsidy rule and either the UCB index subsidy rule or the UCB cost-saving subsidy rule.

Definition 9 (The Hybrid Index Subsidy Rule)**.** Let $s_i^{\text{U-I}}$ be the UCB index subsidy rule. The *hybrid index subsidy rule* $s^{\text{H-I}}$ is defined by

$$s_i^{\text{H-I}}(n; h(n)) := \begin{cases} s_i^{\text{U-I}}(n; h(n)) & \text{if } s_i^{\text{U-I}}(n; h(n)) > a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}(n; h(n))||, \\ 0 & \text{otherwise.} \end{cases}$$

Or, equivalently, the hybrid index subsidy rule can be defined by

$$s_i^{\text{H-I}}(n; h(n)) = \tilde{q}_i^{\text{H}}(n; h(n)) - \hat{q}_i(n; h(n)).$$

Algorithm 5 The Hybrid Index Subsidy Rule

Complete the initial sampling phase by running Algorithm 1.
for $n = N_0 + 1, \cdots, N$ **do** ▷ The hybrid index subsidy rule starts.
  for i **do**
    Compute $\hat{q}_i(n) = \boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_{g(i)}(n)$.
    Compute $\tilde{q}_i(n) = \max_{\bar{\boldsymbol{\theta}} \in \mathcal{C}_{g(n)}} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}$.
    Compute $s_i^{\text{U-I}}(n) = \tilde{q}_i(n) - \hat{q}_i(n)$.
    Offer $s_i(n) = \begin{cases} 0 & \text{if } s_i^{\text{U-I}} \leq a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}||, \\ s_i^{\text{U-I}}(n) & \text{otherwise.} \end{cases}$ ▷ The hybrid index subsidy.
    Firm n hires $\iota(n) = \arg\max_{i \in I(n)} \left\{ \hat{q}_i(n) + s_i^{\text{U-I}}(n) \right\}$ as an equilibrium consequence.
  end for
end for

Definition 10 (The Hybrid Cost-Saving Subsidy Rule)**.** Let $s_i^{\text{U-CS}}$ be the UCB cost-saving subsidy rule. The *hybrid cost-saving subsidy rule* $s^{\text{H-CS}}$ is defined by

$$s_i^{\text{H-CS}}(n; h(n)) := \begin{cases} s_i^{\text{U-CS}}(n; h(n)) & \text{if } s_i^{\text{U-I}}(n; h(n)) > a\sigma_x ||\hat{\boldsymbol{\theta}}_{g(i)}(n; h(n))||, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 7 (The Properties of the Hybrid Subsidy Rules)**.**

1. The hybrid index subsidy rule $s^{\text{H-I}}$ and the hybrid cost-saving subsidy rule $s^{\text{H-CS}}$ implement the hybrid decision rule ι^{H}.

2. The hybrid index subsidy rule $s^{\text{H-I}}$ requires the minimum subsidy among all index subsidy rules that implement ι^{H}.

3. The hybrid cost-saving subsidy rule $s^{\text{H-CS}}$ requires the minimum subsidy among all subsidy rules that implement ι^{H}.

The proof of Theorem 7 is analogous to that of Theorems 5 and 6, and thus is omitted.

The algorithm of the hybrid index subsidy rule and its equilibrium consequence is stated as Algorithm 5. As it is straightforward to modify Algorithm 5 to construct a hybrid cost-saving subsidy rule, we omit the algorithm for the hybrid cost-saving subsidy rule here.

The following two theorems characterize the regret and amount of subsidy of the hybrid decision rule.

Theorem 8 (Regret Bound for the Hybrid Decision Rule)**.** Suppose Assumptions 1, 2, and 3. Then, by choosing sufficiently small δ, the regret under the hybrid decision rule ι^{H} is

bounded as

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{H}}(N)] \leq C_{\mathrm{hyb}}\sqrt{N}$$

where C_{hyb} is a factor that is $\tilde{O}(1)$ to N.

Theorem 9 (Subsidy Bound for the Hybrid Subsidy Rules)**.** Suppose Assumptions 1, 2, and 3. By choosing sufficiently small δ, for any $a > 0$, the total amount of the subsidy under the hybrid index subsidy rule ($\mathrm{Sub}^{\text{H-I}}$) and the hybrid cost-saving subsidy rule ($\mathrm{Sub}^{\text{H-CS}}$) is bounded as

$$\mathrm{Sub}^{\text{H-CS}}(N) \leq \mathrm{Sub}^{\text{H-I}}(N) = C_{\text{hyb-sub}}.$$

where $C_{\text{hyb-sub}}$ is a factor that is $\tilde{O}(1)$ to N.

Proof. See Appendix B.7. The explicit form of C_{hyb} is found at Eq. (54). The explicit form of $C_{\text{hyb-sub}}$ is found at Eq. (61) therein. □

Theorem 8 states that the order of the regret under the hybrid decision rule is $\tilde{O}(\sqrt{N})$, which is the same as the original UCB decision rule. Theorem 9 states that the amount of the subsidy is polylogarithmic to N, which is a substantial improvement from the standard UCB where $\tilde{O}(\sqrt{N})$ subsidy is required.

The threshold of switching from the UCB mechanism to laissez-faire is crucial for guaranteeing the performance of the hybrid mechanism. Our threshold, $a\sigma_x||\hat{\boldsymbol{\theta}}(n)||$, is determined in such a way that the hybrid decision rule ι^{H} satisfies *proportionality*, which is a new concept established in this paper. The formal statement appears in Lemma 28 in Appendix B.7 but it requires additional notations that do not appear in the main body of this paper. In what follows, we provide a high-level intuition regarding the concept of proportionality.

We evaluate the expected regret of the hybrid decision rule by comparing it with the expected regret of the UCB decision rule. However, since different decision rules generate different histories and data, neither decision rule dominates the other. This is why the comparison is challenging. We overcome this problem by proving that the hybrid decision rule ι^{H} is proportional to ι^{U} in the sense that there exists a constant $c > 0$ such that when the UCB rule ι^{U} hires worker i with probability p_i, then the hybrid rule ι^{H} hires worker i with probability at least cp_i given the same history. This property guarantees that the hybrid rule escapes from underexploring the minority group and secures expected regret of $\tilde{O}(\sqrt{N})$.

The timing of switching to laissez-faire is crucial for the proportionality. When the data about the minority group are insufficient, firms rarely hire minority workers under laissez-faire. We prove that, when the threshold is set to $a\sigma_x||\hat{\boldsymbol{\theta}}_g(n)||$, then firms keep hiring minority

workers with sufficiently high frequency, and therefore, statistical discrimination is resolved eventually.

Remark 5 (Dependence on Parameter a)**.** There is a tradeoff between the regret and subsidy. The constant on the top of regret (Theorem 8) is $\exp(a^2/4)$, which is increasing in a. By contrast, the constant on the top of subsidy (Theorem 9) is $\exp(3a^2/4)/a^2$, which goes to infinity as $a \to 0$. Theorem 9 guarantees that the subsidy is $\tilde{O}(1)$ whenever $a > 0$. However, when a is small, the bound provided by Theorem 9 becomes large and may not be insightful. To balance the tradeoff, the government should select a "right-size" value for a. In our simulations (Section 8) we adopt $a = 1$.

For small a, because the hybrid mechanism is close to the UCB mechanism, we can divert our analysis for the UCB mechanism (Theorem 5). When a is small and N is finite, the square-root subsidy bound established in Theorem 5 may provide a tighter characterization of the total subsidy.

7 Interviews and the Rooney Rule

7.1 Two-Stage Model

Although the UCB-based subsidy rule is a powerful policy intervention to resolve statistical discrimination, the subsidy rule is sometimes difficult to implement in practice. This section articulates the advantages and disadvantage of the Rooney Rule, which requires each firm to invite at least one candidate of each group to an on-site interview. The Rooney Rule is relatively easy to implement because it requires neither the subsidy nor hard hiring quota.

To incorporate the additional information the firms acquire through the interview, we make the following modification to the model. In the modified model, each round n consists of two stages. In the first stage, firm n observes the characteristics of each arriving agent $i \in I(n)$, $\boldsymbol{x}_i$. Based on $\boldsymbol{x}_i$, firm n selects a shortlist of *finalists* $I^F(n) \subseteq I(n)$, where $|I^F(n)| = K^F$ for some $K^F \in \mathbb{N}$. In the second stage, by interviewing finalists, firm n observes an additional signal η_i for each finalist i (as assumed in Kleinberg and Raghavan, 2018). Firm n predicts each finalist i's skill from the characteristics $\boldsymbol{x}_i$ and the additional signal η_i, and hires one worker from the set of finalists, $\iota(n) \in I^F(n)$. Firms are not allowed to hire a worker who was not selected as a finalist. After the firm makes a decision, the skill of the hired worker $y_{\iota(n)}$ is publicly disclosed.

We assume the following linear relationship between the skill y_i and the observable variables $\boldsymbol{x}_i$ and ι_i:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)} + \eta_i + \epsilon_i$$

The "noise" term comprises two variables: η_i and ϵ_i. η_i is revealed as an additional signal when the firm chooses i as a finalist. However, ϵ_i remains to be unpredictable even after the hiring—firms only observe y_i after worker i is hired.

For analytical tractability, besides Assumptions 1, 2, and 3, we make the following two assumptions.

Assumption 4 (Two Finalists)**.** Each firm can invite only two finalists; i.e., $K^F = 2$.

Assumption 4 generates a minimal environment to consider the performance of the Rooney Rule.

Assumption 5 (Normal Additional Signals)**.** The signal that the finalist reveals is the independent and identically distributed normal random variable:

$$\eta_i \sim \mathcal{N}(0, \sigma_\eta^2).$$

Remark 6. If $\sigma_\eta = 0$, then the two-stage model is the same as the one-stage model that we have considered in the previous sections.

7.2 Failure of Laissez-Faire in the Two-Stage Model

This subsection analyzes the performance of laissez-faire in this two-stage setting. The result is analogous to the one-stage case (Theorem 3) : laissez-faire often falls in perpetual underestimation, and therefore, has linear regret.

First, we formally define the regret. As in the one-stage model, the benchmark is the first-best decision rule, which is the rule firms would take if the coefficient parameter $\boldsymbol{\theta}$ were known. Clearly, the first-best decision rule would greedily invite top-K^F workers in terms of q_i to the final interview. We denote this set of finalists chosen by the first-best decision rule in round n by $\bar{I}^F(n)$. Formally, $\bar{I}^F(n)$ is obtained by solving the following problem:

$$\bar{I}^F(n) = \operatorname*{arg\,max}_{I' \subseteq I(n)} \sum_{i \in I'} q_i \quad \text{s.t. } |I'| = K^F.$$

After that, the first-best decision rule would observe the realization of η_i for $i \in \bar{I}^F(n)$, and then hires the worker i who has the highest skill predictor: $q_i + \eta_i$. The *unconstrained two-stage regret* is defined as the loss compared with this first-best decision rule. (This regret is named "unconstrained" because we introduce an alternative definition of regret later.)

Definition 11 (Unconstrained Two-Stage Regret)**.** In the two-stage hiring model, the *un-*

constrained two-stage regret U2S-Reg of decision rule ι is defined as follows:

$$\text{U2S-Reg}(N) = \sum_{n=1}^{N} \left\{ \max_{i \in I^F(n)} (q_i + \eta_i) - \left(q_{\iota(n)} + \eta_{\iota(n)}\right) \right\}.$$

Under laissez-faire, firm n's optimal strategy is to greedily choose their candidates based on the belief, i.e.,

$$I^F(n) = \arg\max_{I' \subseteq I(n)} \sum_{i \in I'} \hat{q}_i(n) \quad \text{s.t. } |I'| = K^F.$$

After observing the realization of the additional signals η_i, firm n again selects the candidate with the highest predicted skill:

$$\iota(n) = \arg\max_{i \in I^F(n)} \{\hat{q}_i(n) + \eta_i\}.$$

Even in the two-stage model, laissez-faire has linear regret when the population ratio is imbalanced.

Theorem 10 (Failure of Laissez-Faire in the Two-Stage model)**.** Suppose Assumptions 1, 3, 2, 4, and 5. Suppose also that $K_2 = 1$ and $d = 1$. Let $K_1 - \log_2(K_1 + 1) > \log_2 N$. Then, under the laissez-faire decision rule, group 2 is perpetually underestimated with the probability at least $C_{\text{imb}} = \tilde{\Theta}(1)$. Accordingly, the expected regret of the laissez-faire decision rule is

$$\mathbb{E}\left[\text{U2S-Reg}^{\text{LF}}\right] = \tilde{\Omega}(N).$$

Proof. See Appendix B.8. □

The proof idea of Theorem 10 is as follows. Under laissez-faire, each firm n interviews two candidates who have the highest expected skills, $\hat{q}_i(n)$. If both of these two workers are majorities, then minority workers are never hired no matter what the η_i for each finalist is. By evaluating the probability that both finalists are majorities, we derive the probability that perpetual underestimation occurs. Note that, to meet $K_1 - \log_2(K_1 + 1) \geq \log_2 N$, K_1 should be $\Omega(\log N) = \tilde{\Omega}(1)$. Hence, Theorems 3 and 10 require the same rate of the imbalanced population ratio.

To summarize, even in a two-stage setting, the laissez-faire decision has linear regret (when the population ratio is imbalanced). This is because the laissez-faire decision rule results in perpetual underestimation with a significant probability.

7.3 The Rooney Rule and Exploration

As laissez fair does not perform well, we need to seek for desirable policy intervention. The Rooney Rule, which requires each firm to invite at least one minority finalist to the final interview, is one of the natural affirmative actions in this setting, and is widely implemented in real-world problems.

Definition 12 (The Rooney Rule)**.** In the two stage hiring model, the *Rooney Rule* requires each firm n to select at least one finalist from every group $g \in G$; i.e., for every n and every $g \in G$, $I^F(n)$ must satisfy

$$\left|\left\{i \in I^F(n) \mid g(i) = g\right\}\right| \geq 1. \tag{7}$$

The Rooney Rule is relatively easy to implement because it imposes no hiring quota or hiring preference given to minorities. The Rooney Rule of originally introduced as the National Football League policy to promote hiring of ethnic-minority candidates for head coaching positions, but variations of the Rooney Rule are now implemented in many industries.[20] Although the Rooney Rule has been used in many places, its theoretical performance has not been studied intensively.

To understand the fact that the Rooney Rule resolves statistical discrimination, we introduce an alternative (weaker) notion of regret, *constrained two-stage regret*.

Definition 13 (Constrained Two-Stage Regret)**.** In the two-stage hiring model, the *constrained two-stage regret* (C2S-Reg) of decision rule ι is defined as follows:

$$\text{C2S-Reg}(N) = \sum_{n=1}^{N} \left\{ \max_{i \in \breve{I}^F(n)} (q_i + \eta_i) - \left(q_{\iota(n)} + \eta_{\iota(n)}\right) \right\}.$$

where $\breve{I}^F(n)$ is given by

$$\begin{aligned} \breve{I}^F(n) = \underset{I' \subseteq I(n)}{\arg\max} & \sum_{i \in I} q_i \\ \text{s.t. } & |I'| = K^F, \\ & \forall g \in G, \ \left|\left\{i \in \breve{I}^F(n) \mid g(i) = g\right\}\right| \geq 1. \end{aligned} \tag{8}$$

[20] For example, in a securities and exchange commission filing posted on 2018, Amazon declares that "*The Amazon Board of Directors has adopted a policy that the Nominating and Corporate Governance Committee include a slate of diverse candidates, including women and minorities, for all director openings. This policy formalizes a practice already in place*" (https://www.sec.gov/Archives/edgar/data/1018724/000119312518162552/d588714ddefa14a.htm). In addition, according to O'Brien (2018), Facebook COO Sheryl Sandberg said that "*The company's 'diverse slate approach' is a sort of 'Rooney Rule,' the National Football League policy that requires teams to consider minority candidates.*"

In plain words, $\breve{I}^F(n)$ is the best list of finalists who satisfy the constraint (7). If (7) is imposed as an "exogenous constraint" (rather than a policy), the first-best decision rule would interview $\breve{I}^F(n)$ to maximize social welfare. Clearly, the unconstrained regret is larger than the constrained regret.

The constrained regret is useful in that it enables us to identify whether the Rooney Rule prevents perpetual underestimation—if perpetual underestimation occurs under the Rooney Rule, then the constrained regret is linear in N. To the contrary, if the social learning is successful (i.e., $\hat{q}_i$ is very close to q_i for all the workers), the constrained regret would be zero.

Under Rooney Rule, myopic firm n would greedily choose candidates based on estimator $\hat{q}_i(n)$ subject to the constraints:

$$\begin{aligned} I^F(n) = \underset{I' \subseteq I(n)}{\arg\max} \sum_{i \in I} \hat{q}_i(n) \qquad (9) \\ \text{s.t. } |I'| = K^F, \\ \forall g \in G,\ \left|\left\{i \in I^F(n) \mid g(i) = g\right\}\right| \geq 1. \end{aligned}$$

and $\iota = \arg\max_{i \in I^F(n)} \{\hat{q}_i(n) + \eta_i\}$. Note that the only difference between Eq. (8) and (9) is that q_i is replaced by $\hat{q}_i(n)$.

The following theorem states that the Rooney Rule is able to resolve perpetual underestimation with sufficiently revealing signal η_i.

Theorem 11 (Sublinear Constrained Regret under the Rooney Rule)**.** Suppose Assumptions 1, 2, 3, 4, and 5. Then, the regret under the Rooney Rule is bounded as

$$\mathbb{E}\left[\text{C2S-Reg}^{\text{Rooney}}(N)\right] \leq C_{\text{2SR}}\sqrt{N}$$

where C_{2SR} is $\tilde{O}(1)$ to N.

Proof. See Appendix B.9. The explicit form of C_{2SR} is found at Eq. (68) therein. Note that C_{2SR} is exponentially dependent on signal variance σ_η (see the definition of C_6 in Eq. (64)), which implies that a sufficiently large value of σ_η is required to obtain a reasonable bound. □

The proof idea is as follows. When a group is underrepresented, no candidates from the group is regarded as the most promising finalist with a significant probability. Hence, laissez-faire may result in perpetual underestimation. The Rooney Rule mitigates this problem by securing a finalist seat for each group. If the additional signal is informative enough (i.e., σ_η is large), there is some probability that the minority finalist beats the majority finalist

and is hired. In other words, additional signal naturally induces exploration for the minority group and prevents perpetual underestimation.

Remark 7. The Rooney Rule is analogous to the ϵ-greedy algorithm that is widely studied in the multi-armed bandit and reinforcement learning literature. The ϵ-greedy algorithm usually makes a decision based on the greedy algorithm (equivalent to laissez-faire in our model), but there is a small probability (ϵ) that the algorithm chooses a worker uniformly at random. In the bandit literature, the ϵ- greedy algorithm is known to be robust to the choice of the exploration probability ϵ: In fact, one can prove that the regret of the ϵ-greedy algorithm is sub-linear for any value $\epsilon > 0$. In our model, the Rooney Rule successfully resolve underexploration because of the randomness in the additional signal η_i induces ϵ-experiments.

7.4 The Rooney Rule and Exploitation

This subsection shows that, although the Rooney Rule successfully prevents statistical discrimination, it may worsen social welfare evaluated by the original unconstrained regret.

When the population ratio is imbalanced (i.e., K_1/K_2 is large), there is a significant probability that more than one majority worker has high skills. In that case, the *true* predicted skill of the second-best majority worker (q_i) is likely to be higher than that of the minority champion. This feature raises constant regret per round: when η_i is normally distributed, any finalist has a positive probability of being hired. Hence, the skills of all candidates matter, and therefore, firms want to interview top-K^F candidates who have the highest skills. The Rooney Rule prevents this outcome. This effect would present even when firms had perfect information about coefficients $\boldsymbol{\theta}$. Furthermore, the loss from the constraint (7) is constant per round, and therefore, results in the unconstrained regret of $\Omega(N)$ in total.

Theorem 12 (Linear Unconstrained Regret under the Rooney Rule)**.** Suppose Assumptions 1, 2, 3, 4, 5. Then, the regret under the Rooney Rule is bounded as

$$\mathbb{E}\left[\text{U2S-Reg}^{\text{Rooney}}(N)\right] = \Omega(N).$$

The proof is straightforward from the argument above, and therefore, is omitted.

In summary, both laissez-faire and the Rooney Rule have linear unconstrained regret. However, the structure behind these results are different. Laissez-faire has linear regret due to underexploration. In contrast, the Rooney Rule has linear regret due to underexploitation.

One way to resolve this trade-off is to mix the Rooney Rule and laissez-faire (as the hybrid mechanism does). By starting with the Rooney Rule and abolishing it after sufficiently rich

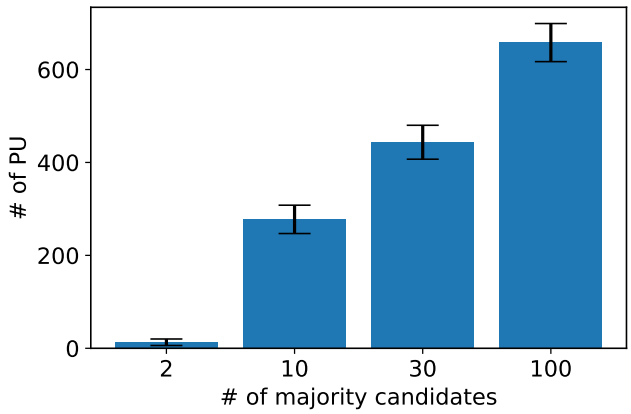


Figure 1: The number of perpetual underestimation among 2000 runs under laissez-faire. The error bars are the two-sigma binomial confidence intervals.

data is obtained, we could mitigate the disadvantage of the Rooney Rule. In Section 8, we also testify the performance of such a mechanism.

8 Simulation

8.1 Setting

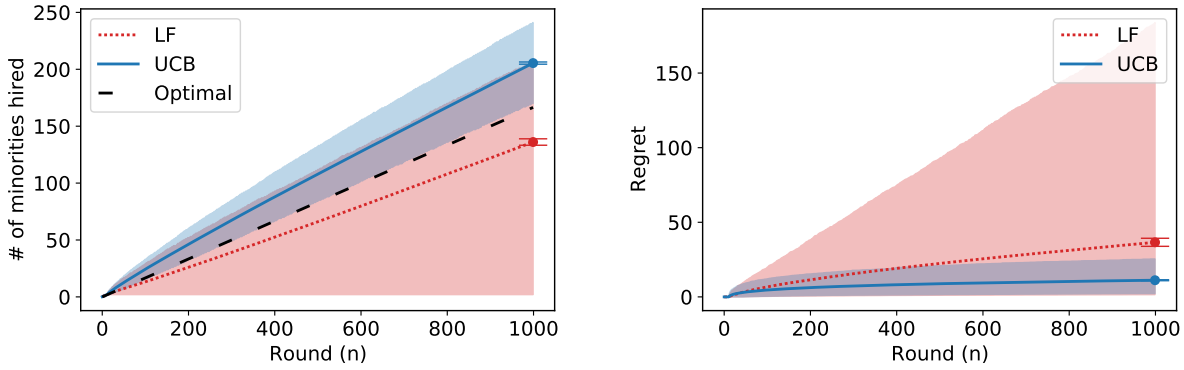
This section reports the results of the simulations that we run to support our theoretical findings[21]. Unless specified, the model parameters are set as follows: $d = 1, \mu_x = 3, \sigma_x = 2$, $\sigma_\epsilon = 2$. The regularizer of regression is set to be $\lambda = 1$. The group sizes are set to be $(K_1, K_2) = (10, 2)$. The initial sample size is $N^{(0)} = K_1 + K_2$, and the sample size for each group is equal to its population ratio: $N_1^{(0)} = K_1, N_2^{(0)} = K_2$. All the results are averaged over 2000 runs.

The value of δ in the confidence bound is set to 0.1.

8.2 The Effects of the Population Ratio

We first testify the population ratio effects to the frequency of perpetual underestimation (i.e., group 2 is never hired after the initial sampling phase). The decision rule is fixed to laissez-faire (LF). We fix the number of minority candidates in each round to two (i.e., $K_2 = 2$) and vary the number of majority candidates ($K_1 = 2, 10, 30, 100$).

[21] The source code of the simulations is available at https://github.com/jkomiyama/FairSocialLearning/



(a) Number of minority candidates hired

(b) Regret

Figure 2: The comparison between the LF and UCB decision rules. The lines are the average over sample paths, and the areas cover between 5% and 95% percentile of runs. The error bars at $N = 1000$ are the two-sigma confidence intervals.

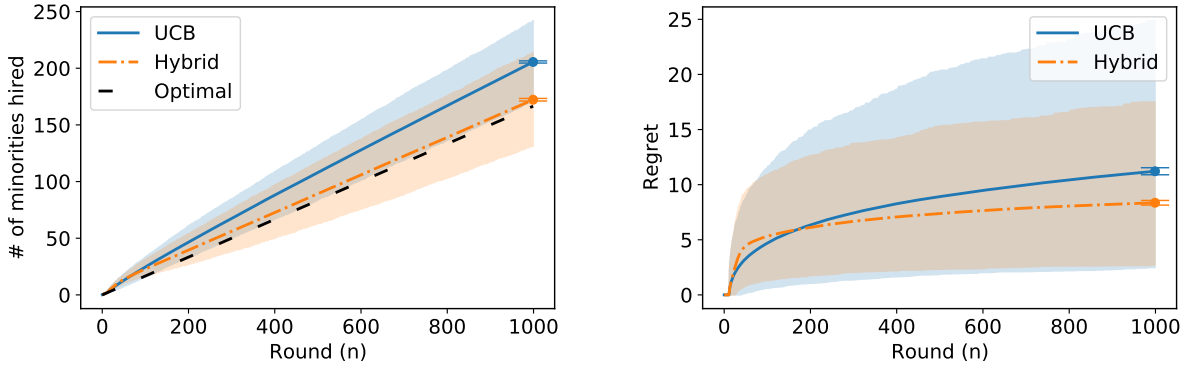
Figure 1 exhibits the simulation result. Consistent with our theoretical analyses, we observe that (i) as indicated by Theorem 2 laissez-faire rarely results in perpetual underestimation if the population is balanced (i.e., K_1 is close to $K_2 = 2$), and (ii) as indicated by Theorem 3, perpetual underestimation becomes more frequent as the population of majority workers increases (i.e., K_1 increases).

8.3 Laissez-Faire vs The UCB Decision Rule

Figure 2a compares the number of minority workers hired by the laissez-faire (LF) and UCB decision rules. Figure 2b compares the regret under these two rules. The horizontal axis represents the round (where the number of total rounds is fixed to $N = 1000$), and the vertical axis represents the number of minority workers hired and the regret, respectively. The subsidy required by the UCB mechanism will be shown later (in Figure 4).

As indicated by Theorem 3, our simulation shows that laissez-faire has a significant probability of underestimating the minority group. Consequently, we observe the following two facts. First, the number of minority workers hired on average is lower than the first-best decision rule would hire (hire a minority worker with probability $K_2/(K_1+K_2) = 2/(10+2) \approx 17\%$ for each round). Second, laissez-faire sometimes causes perpetual underestimation, and therefore, the number of minority workers hired could be zero, and the regret grows linearly in n even after 1000 rounds. Due to the possibility of perpetual underestimation, the confidence intervals of the sample paths (denoted by the read area) is very large, indicating that the performance of laissez-faire is highly uncertain.

In contrast, consistent with Theorem 4, the performance of the UCB decision rule is



(a) Number of minority candidates hired

(b) Regret

Figure 3: The comparison between the UCB and hybrid decision rules. The lines are average over sample paths, and the areas cover between 5% and 95% percentile of runs. The error bars at $N = 1000$ show the two-sigma confidence intervals of the expected regret.

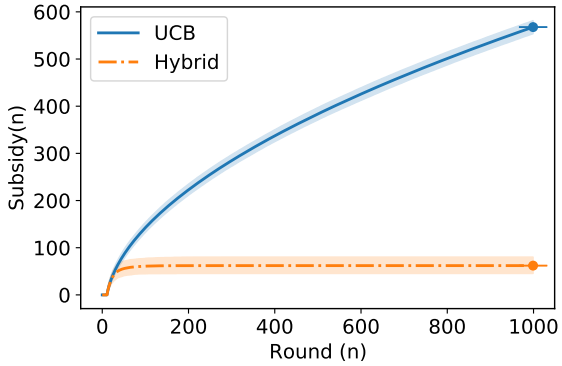
shown to be much more stable. As the UCB rule avoids underexploration, it does not cause perpetual underestimation. Consequently, the regret of UCB is not only lower than laissez-faire on average but also has smaller variance. Note that the UCB decision rule tends to hire more minority workers than the first-best decision rule. This outcome happens because society is typically less knowledgeable about the minority group (due to an uneven population ratio), and therefore, the confidence interval for minority workers is typically larger than that for the majority.

8.4 The UCB Mechanism vs the Hybrid Mechanism

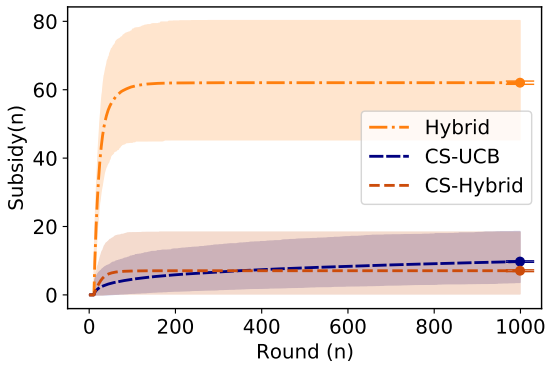
Next, we compare the performance of the UCB and hybrid mechanisms. The parameter of the hybrid mechanism is set to be $a = 1$. Figure 3 compares the performance of these decision rules: Figure 3a shows the number of minority workers hired, and Figure 3b shows the regret.

We observe that the number of the minority hired on average becomes closer to the first-best decision rule (Figure 3a). Furthermore, as expected by Theorems 4 and 8, as for efficiency (regret), the performance of these two decision rules grows in the same order. However, we find that the hybrid decision rule outperforms UCB in our simulation setting (Figure 3b). We consider that these results happen because the hybrid decision rule stops overexploration of the minority group in an early stage.

Figure 4 compares the total budgets required by (i) the UCB index subsidy rule (UCB), (ii) the hybrid index subsidy rule (Hybrid), (iii) the UCB cost-saving subsidy rule (CS-UCB), and (iv) the hybrid cost-saving subsidy rule (CS-Hybrid).



(a) The UCB index subsidy rule vs the hybrid index subsidy rule.



(b) The UCB cost-saving subsidy rule vs the hybrid index subsidy rule and the hybrid cost-saving subsidy rule.

Figure 4: The comparison of the budget required by subsidy rules. The lines are average over sample paths, and the areas cover between 5% and 95% percentile of runs. The error bars at $N = 1000$ show the two-sigma confidence intervals of the expected regret.

Figure 4a compares the index subsidy rules. As predicted by Theorems 5 and 9, the hybrid index subsidy rule requires a much smaller budget than the UCB index subsidy rule. Furthermore, the subsidy distributed by the UCB rule seems still growing, even after 1000 rounds are finished. This is also consistent with our theory because the UCB rule requires $\tilde{O}(\sqrt{N})$ subsidy (while the hybrid rule only requires $\tilde{O}(1)$ subsidy).

Figure 4b compares the subsidy amount of the UCB cost-saving subsidy rule and the hybrid subsidy rules. The UCB index subsidy rule is excluded because it requires a much larger subsidy amount. We observe that (i) two cost-saving subsidy rules require a similar amount of the subsidy (while the hybrid cost-saving subsidy rule performs slightly better), and (ii) the cost-saving method is very effective, even when it is compared with the hybrid index rule.

Note that, although the subsidy amounts required by these two cost-saving rules are similar, when we have more rounds, the hybrid cost-saving subsidy rule outperforms. Figure 5 articulates this result. While the subsidy required by the hybrid cost-saving rule remains constant after a few (about 100) rounds, the subsidy by the UCB cost-saving rule gradually grows. This result is also consistent with our theory: While the subsidy required by the hybrid rule is $\tilde{O}(1)$ (Theorem 9), the subsidy required by the UCB cost-saving rule is $\tilde{O}(\sqrt{N})$ (Theorem 6).

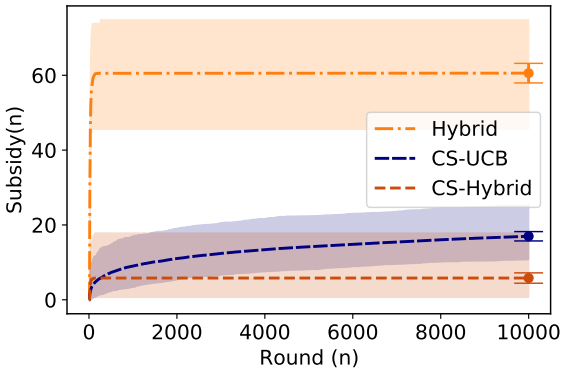


Figure 5: The UCB cost-saving subsidy rule vs the hybrid index subsidy rule and the hybrid cost-saving subsidy rule where $N = 10000$. Each line is an average over sample paths, and the areas cover between 5% and 95% percentile of runs. Due to computational limitation, we only did 50 runs of this simulation. The error bars at $N = 10000$ show the two-sigma confidence intervals of the expected regret.

8.5 The Rooney Rule

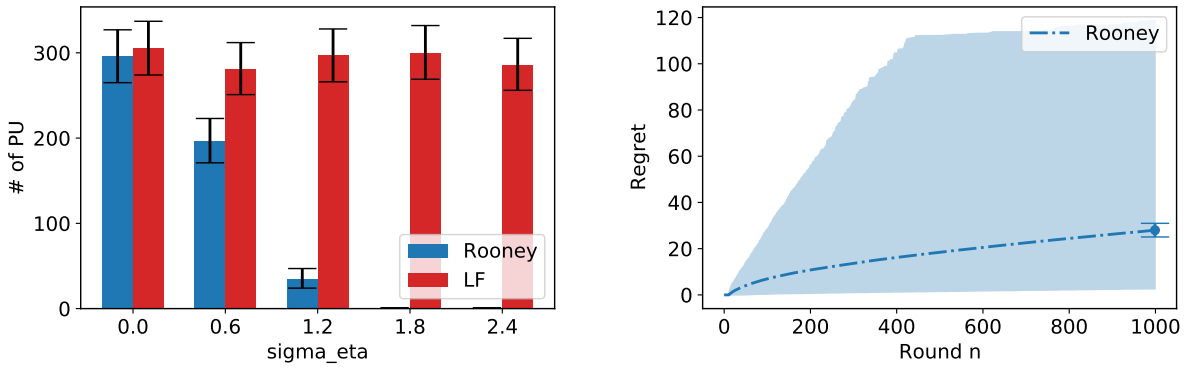
This subsection describes the performance of the Rooney Rule compared with laissez-faire. Figure 6a depicts the relationship between the frequency of perpetual underestimation and the informativeness of the signal obtained at the second stage (measured by σ_η^2, which is the variance of η_i) under laissez-faire and the Rooney Rule.

For the Rooney Rule, we observe that when the second-stage signal η_i is more informative, perpetual underestimation occurs less often. This outcome happens because, even when the minority finalist is underestimated (the predicted skill $\hat{q}_i$ is small while the true skill q_i is large), when σ_i^2 is large, the minority finalist has a significant probability of overturning the situation. If this happens often enough, society can learn about the minority group, and statistical discrimination can be spontaneously resolved.

As for laissez-faire, we observe that laissez-faire falls in perpetual underestimation with a significant probability for any σ_η adopted in the simulation. This outcome is consistent with our analysis (Theorem 10). Since minority workers are rarely chosen as finalists, they have no opportunity to be hired even when σ_η is large. These results imply that that, even in a two-stage model, laissez-faire frequently results in statistical discrimination.

Figure 6b shows the constrained regret of the Rooney Rule. We can observe that the constrained regret grows sublinearly in n, implying that the Rooney Rule resolves perpetual underestimation. Hence, under the Rooney Rule, society does not suffer from underexploration of the minority group.

However, this does not imply that the Rooney Rule arrives come without cost. As we



(a) The number of perpetual underestimation among 2000 runs. The error bars show binomial confidence intervals.

(b) The constrained two-stage regret. σ_η is fixed to 1.2.

Figure 6: The Rooney Rule's performance for exploration.

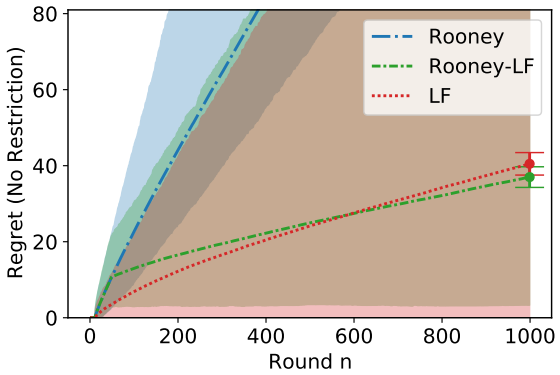


Figure 7: The unconstrained two-stage regret under laissez-faire (LF), the Rooney Rule, and their hybrid (Rooney-LF).

discussed in Subsection 7.4, once the coefficient parameter $\boldsymbol{\theta}$ is learned, the Rooney Rule may prevent society from making a fair and efficient decision. To testify this, we also examine the growth of unconstrained regret. Figure 7 exhibits the results of this simulation. We find that the performance of the Rooney Rule is worse than laissez-faire because the cost of underexploitation (of Rooney) exceeds the cost of underexploitation (of laissez-faire).

As we indicated in Subsection 7.4, the performance of the Rooney Rule could be improved if we terminate it after "learning is completed." In this simulation, we also test this rule, the *Rooney-LF rule*—impose the Rooney Rule to first 100 firms and then turn to laissez-faire. We find that the Rooney-LF rule avoids perpetual underestimation, and therefore, has a similar performance to laissez-faire. This result indicates that, if we select the transition timing appropriately, then we can resolve statistical discrimination without compromising

the quality of the finalists.

8.6 The Hybrid Mechanism vs Uniform Sampling

Kannan et al. (2018) show that when we have sufficiently large initial samples (i.e., $N^{(0)}$ is large), the greedy algorithm (corresponding to laissez-faire in this paper) has sublinear regret.[22] As stated in Remark 2, our analysis also indicates that the probability of perpetual underestimation is small when $N^{(0)}$ is large (see Lemma 24 for the full detail).

One may think that this "warm-start" version of laissez-faire is efficient. However, the warm-start approach has several disadvantages. First, while we have ignored the cost of acquiring initial samples thus far for analytical tractability, we need to take into account of the cost of acquiring uniform samples if we want to *take* a sufficiently long warm-start period. As uniform sampling ignores firms' incentives for hiring workers, we need a large budget to implement it in practice. Second, uniform sampling does not maximize any index. Accordingly, it cannot be implemented by any index policy. Third, uniform sampling is inefficient in terms of information acquisition because it is not adaptive to current estimated parameters.

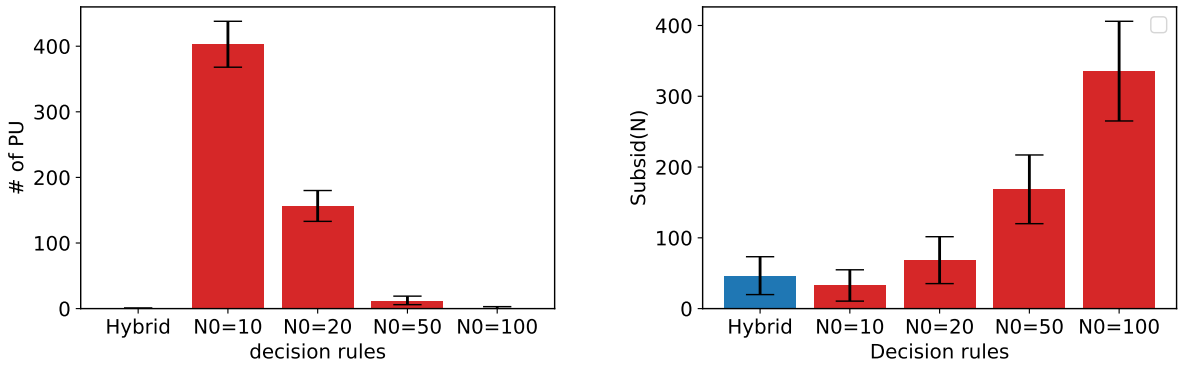
We argue that our hybrid mechanism (Section 6) is a more sophisticated version of laissez-faire with a warm start—it initially samples the data adaptively and then switch to laissez-faire at an efficient timing. Hence, we can naturally expect that the hybrid mechanism outperforms laissez-faire with initial uniform sampling.

Figure 8 exhibits the simulation results that compare the hybrid mechanism with laissez-faire with various initial samples. In this simulation, the number of initial samples for each group is proportional to the population ratio; i.e., $N_g^{(0)} = (K_g/K) \cdot N^{(0)}$.

Figure 8a measures the number of perpetual underestimations. As indicated by our theory, the larger the initial sample, the less frequently perpetual underestimation occurs. In addition, we observed no perpetual underestimation under the hybrid mechanism, as it solidly incentivizes hiring from an underexplored group.

Figure 8b depicts the subsidy amount required by the cost-saving subsidy rules (recall that uniform sampling cannot be acquired by any index subsidy rule). Here, we can observe that the hybrid cost-saving subsidy rule outperforms laissez-faire with uniform sampling. Laissez-faire requires at least $N^{(0)} > 20$ samples to mitigate perpetual underestimation. However, when $N^{(0)} \geq 20$, the hybrid cost-saving subsidy rule requires a smaller budget than uniform sampling. This result indicates that the hybrid mechanism is more efficient in

[22] We also note that the number of initial samples required by the relevant theorem ($n_{\min}$ of Lemma 4.3 therein) is very large and cannot be satisfied in our simulation setting: Letting $R = \sigma_x\sqrt{2\log(N)}$, we have $n_{\min} \geq 320R^2\log(R^2dK/\delta)/\lambda_0 \geq 10^3$.



(a) The number of perpetual underestimations among 2000 runs.

(b) The total amount of subsidies. "Hybrid" denotes the hybrid cost-saving subsidy rule.

Figure 8: The comparison between the hybrid mechanism and uniform sampling. N0 ($= N^{(0)}$) denotes the number of initial samples taken prior to laissez-faire. The error bars are the two-sigma binomial confidence intervals.

compensating firms.

9 Bayesian Approach

Thus far, we have assumed that firms and the regulator are frequentists, implying that they estimate the underlying parameter $(\boldsymbol{\theta}_g)_{g\in G}$ only based on the data, without forming a "prior belief" about the distribution of the parameter.

The frequentist approach is widely used in the multi-armed bandit literature because selecting a "prior belief" for implementing a Bayesian approach is difficult in practice. Since the frequentist approach is valid for *any* realization of the parameter $(\boldsymbol{\theta}_g)_{g\in G}$, it produces a more robust solution. Based on this trend, we have developed and analyzed a frequentist model.

Nevertheless, we will come to a similar conclusion even when we adopt a Bayesian setting, as long as all firms share a common prior belief. Specifically, when the common prior belief is endowed as a normal distribution, i.e.,

$$\boldsymbol{\theta}_g \sim \mathcal{N}(0, \kappa^2 \boldsymbol{I}_d),$$

then the posterior belief would also be a normal distribution $\mathcal{N}(\boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_{g(i)}(n), \boldsymbol{\Sigma}_t)$. Furthermore, $\hat{\boldsymbol{\theta}}_g(n)$ is exactly the same as the one of Eq. (2) where $\lambda = \sigma_\epsilon^2/\kappa^2$ (c.f., p.120 in Kaufmann, 2014). Therefore, under the laissez-faire mechanism, the firms would optimally apply a ridge regression even in a Bayesian model.

Regarding the UCB mechanism, there exists a Bayesian version of confidence region[23] $\mathcal{C}_g(n)^{\text{Bayes}}$ such that such that

$$\Pr^{\text{Bayes}}\left(\bigcap_n \{\boldsymbol{\theta}_g \in \mathcal{C}_g^{\text{Bayes}}(n)\}\right) \geq 1-\delta$$

by defining

$$\mathcal{C}_g^{\text{Bayes}}(n) = \left\{\bar{\boldsymbol{\theta}}_g \in \mathbb{R}^d : \left\|\bar{\boldsymbol{\theta}}_g - \hat{\boldsymbol{\theta}}_g(n)\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \sigma_\epsilon \sqrt{d + \log\left(\frac{\pi^2 N^2}{6\delta}\right) + 2\sqrt{d\log\left(\frac{\pi^2 N^2}{6\delta}\right)}}\right\}.$$

By using $\mathcal{C}_g^{\text{Bayes}}(n)$, we can obtain a Bayesian version of the UCB mechanism.[24]

10 Contribution to the Multi-Armed Bandit Literature

Thus far, we have stated all the results in the terminology of the economics and statistical discrimination literature. However, this paper also makes several technological contributions to the literature of the contextual bandit problems, which are of independent interest. In particular, we consider non-discounted reward formalization (Robbins, 1952; Lai and Robbins, 1985). Unlike other formalization such as Gittins's (1979) one (e.g., Sundaram, 2005; Bergemann and Välimäki, 2006), this formalization weights future rewards and the current reward equally. The greedy and the UCB algorithms have been intensively studied in this literature, and we made several contributions to it. For convenience of the readers, we state our technological contributions using the bandit terminology.

Perpetual Underestimation The greedy algorithm (which takes optimal decision at each round based on plug-in parameters) fails due to the randomness in finite samples. This concept originated in a "context-less" bandit, a traditional model that corresponds to the limit of $\sigma_x \to 0$. We prove that, when the context is fixed (or has very small variance), exploration is required to mitigate perpetual underestimation (Theorem 1).

[23] Here, $\Pr^{\text{Bayes}}$ denotes a probability over the Bayes posterior. The bound here is derived from Eq. (4.8) in Kaufmann (2014).

[24] Note that, to run the UCB mechanism in a model, the regulator needs to know the common prior belief of firms to calculate the confidence bound $\mathcal{C}_g^{\text{Bayes}}(n)$.

Analysis of the Greedy Algorithm in a Disproportionate Model Some previous studies (Bastani et al., 2020; Kannan et al., 2018) show that the greedy algorithm performs well if the context variation is sufficient. Our results (Theorem 3) indicate that, when multiple arms form a group (cluster) and share the coefficient parameter, the ratio of the group size is crucial for the performance of the greedy algorithm (laissez-faire). This is a novel finding in the contextual multi-armed bandit literature. When the contexts have limited variance, the greedy algorithm fails.

Development of the Hybrid Algorithm Thus far, the contextual bandit literature (e.g., Chu et al., 2011) has studied the regret with an "adversarial" setup where the contexts (characteristics) are chosen to maximize the regret, and the UCB algorithm was designed to solve such an adversarial bandit problem.

By contrast, this paper assumes that the contexts are drawn from a fixed distribution. Our hybrid algorithm, which switches from an UCB algorithm to a greedy algorithm, takes advantage of the knowledge about the context distribution (more specifically, the information about σ_x), and selects an appropriate time for switching. Consequently, we obtained the proportionality (Lemma 28), which is a crucial lemma to evaluate the performance of the hybrid algorithm. As shown theoretically (Theorem 9) and numerically (Subsection 8.6), the hybrid algorithm outperforms the UCB algorithm in terms of the total budget required to a large extent.

Analysis of the Rooney Rule To our knowledge, this is the first multi-armed bandit study on the Rooney Rule.[25] We show that, the greedy algorithm underexplores some arms even when agents are unbiased and fully rational, and the Rooney Rule can mitigate that underexploration (Theorem 11). The uncertainty in the first stage (the realization of η_i) helps to mitigate perpetual underestimation by implicitly encouraging exploration.

11 Conclusion

We studied statistical discrimination using a contextual multi-armed bandit model. Our dynamic model articulates that statistical discrimination can be caused by the failure of social learning. In our model, the insufficiency of the data about the minority group is endogenously generated. The lack of data prevents firms from estimating the candidate workers' skill accurately. Consequently, firms tend to prefer hiring a majority worker, which

[25]Prior to our work, Kleinberg and Raghavan (2018) study the Rooney Rule in the context of evaluation bias.

makes the data sufficiency persistent (perpetual underestimation). In our setting, this form of statistical discrimination is not only unfair but also inefficient. We showed that when the population ratio is imbalanced, laissez-faire tends to cause this phenomenon.

We analyzed two possible policy interventions for mitigating statistical discrimination due to the data insufficiency. One is the subsidy rules for incentivizing firms to hire minority workers. We established the UCB and hybrid mechanisms and analyzed their performance theoretically and numerically. Another policy is the Rooney Rule, which requires firms to interview at least one minority candidate. Our result indicates that the Rooney Rule with an appropriate termination would resolve statistical discrimination, while maintaining the level of social welfare. These results contrast with to some of the previous studies (e.g., Foster and Vohra, 1992; Coate and Loury, 1993; Moro and Norman, 2004) that have shown that affirmative-action policies can be counterproductive.

Our analyses of the subsidy rules and the Rooney Rule provide a consistent practical policy implication: Affirmative actions are useful for resolving statistical discrimination caused by the data insufficiency, but they should be terminated once information acquisition is completed. If we start with laissez-faire, firms may be reluctant to hire minority workers, and perpetual underestimation could occur. Conversely, if we keep using an affirmative-action policy for a long period of time, the policy may unfairly crowd out skilled majority workers. To summarize, a *temporary* affirmative action would be the best solution to resolve statistical discrimination as a failure of social learning.

Appendix

A Lemmas

This section describes the technical lemmas that are used for deriving the theorems.

The Hoeffding inequality, which is one of the most well-known versions of concentration inequality, provides a high-probability bound of the sum of bounded independent random variables.

Lemma 13 (Hoeffding inequality)**.** Let $x_1, x_2, \ldots, x_n$ be i.i.d. random variables in $[0, 1]$. Let $\bar{x} = (1/n) \sum_{t=1}^{n} x_t$. Then,

$$\Pr\left[\bar{x} - \mathbb{E}[\bar{x}] \geq k\right] \leq e^{-2nk^2}$$
$$\Pr\left[\bar{x} - \mathbb{E}[\bar{x}] \leq -k\right] \leq e^{-2nk^2}$$

and taking union bound yields

$$\Pr\left[|\bar{x} - \mathbb{E}[\bar{x}]| \geq k\right] \leq 2e^{-2nk^2}.$$

The following is a version of concentration inequality for a sum of squared normal variables.

Lemma 14 (Concentration Inequality for Chi-squared distribution)**.** Let $Z_1, Z_2, \ldots, Z_n$ be independent standard normal variables. Then,

$$\Pr\left[\left|\frac{1}{n}\sum_{k=1}^{n} Z_k^2 - 1\right| \geq t\right] \leq 2e^{-nt^2/8}$$

Lemma 15 (Normal Tail Bound, Feller, 1968)**.** Let $\phi(x) := \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ be the probability density function (pdf) of a standard normal random variable. Let $\Phi^c(x) = \int_x^\infty \phi(x')dx'$. Then,

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}} \leq \Phi^c(x) \leq \frac{1}{x}\frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

Lemma 16 (Largest Context, Theorem 1.14 in Rigollet, 2015)**.** Let

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_x, \sigma_x \boldsymbol{I}_d)$$

for each $i \in I(n)$. Let $\mu_x = ||\boldsymbol{\mu}_x||$ and

$$L = L(\delta) := \mu_x + \sigma_x\sqrt{2d(2\log(KN) + \log(1/\delta))}$$

Then, with probability at least $1 - \delta$, we have

$$\forall i \in I(n), n \in [N],\ ||\boldsymbol{x}_i||_2 \leq L(\delta).$$

The following bounds the variance of a conditioned normal variable.

Lemma 17 (Conditioned Tail Deviation)**.** Let $x \sim \mathcal{N}(a, 1)$ be a scalar normal random variable with its mean $a \in \mathbb{R}$ and unit variance. Then, for any $b \in \mathbb{R}$, the following two inequalities hold.

$$\mathrm{Var}(x|x \geq b) \geq \frac{1}{10}$$

Proof of Lemma 17. Without loss of generality, we assume $b = 0$ because otherwise we can

reparametrize $x' = x - b \sim \mathcal{N}(a-b, 1)$. If $a \leq 0$, the pdf of conditioned variable $x|x \geq 0$ is $2\psi(x)$ for $x \geq 0$. Manual evaluation of this distribution[26] reveals that $\mathrm{Var}(x) \geq 1/10$. Otherwise ($a > 0$), the pdf of $x|x \geq b$ is $p(x) \geq \psi(x-a)$ for $x \geq a$, which implies $\mathrm{Var}(x|x \geq b) \geq \mathrm{Var}(z)$, where z be a "half-normal" random variable[27] with its cumulative distribution function

$$P(z) = \begin{cases} \Phi(z) & \text{if } z > 0 \\ 1/2 & \text{if } z = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Manual evaluation of $\mathrm{Var}(z)$ also shows that $\mathrm{Var}(z) \geq 1/10$. □

The following diversity condition that simplifies the original definition of (Kannan et al., 2018) is used to lower-bound the expected minimum eigenvalue of $\bar{\boldsymbol{V}}_g$.

Lemma 18 (Diversity of Multivariate Normal Distribution)**.** The context $\boldsymbol{x}$ is λ_0-diverse for $\lambda_0 \in \mathbb{R}$ if for any $\hat{b} \in \mathbb{R}$, $\hat{\boldsymbol{\theta}} \in \mathbb{R}^d$

$$\lambda_{\min}\left(\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}'|\boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right]\right) \geq \lambda_0.$$

Let $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \sigma_x \boldsymbol{I}_d)$. Then, the context $\boldsymbol{x}$ is λ_0-diverse with $\lambda_0 = \sigma_x^2/10$.

Proof of Lemma 18.

$$\begin{aligned}\lambda_{\min}\left(\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}'|\boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right]\right) &= \min_{\boldsymbol{v}:||\boldsymbol{v}||=1} \mathbb{E}\left[(\boldsymbol{v}'\boldsymbol{x})^2|\boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right] \\ &\geq \min_{\boldsymbol{v}:||\boldsymbol{v}||=1} \mathrm{Var}\left[\boldsymbol{v}'\boldsymbol{x}|\boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right]\end{aligned}$$

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d$ be the orthogonal bases. Without loss of generality, we assume $\hat{\boldsymbol{\theta}} = \theta_1 \boldsymbol{e}_1$ for some $\theta_1 \geq 0$ and $\mu_x = u_1 \boldsymbol{e}_1 + u_2 \boldsymbol{e}_2$ for some $u_1, u_2 \in \mathbb{N}$. Let

$$\boldsymbol{x} = x_1 \boldsymbol{e}_1 + x_2 \boldsymbol{e}_2 + \cdots + x_d \boldsymbol{e}_d.$$

Due to the property of the normal distribution, each coordinate x_l for $l \in [d]$ are independent each other. We will show the variance of

$$\mathrm{Var}\left[x_l|\boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right] \geq \sigma_x^2/10, \tag{10}$$

which suffices to prove Lemma 18.

[26] This distribution is called a folded normal distribution.
[27] Half of the mass lies in $z > 0$, the other half of mass is at $z = 0$.

- For the first dimension, we have $x_1 \sim \mathcal{N}(u_1, \sigma_x^2)$ and

$$\mathrm{Var}\left[x_1 | \boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right] = \mathrm{Var}\left[x_1 | x_1 \geq \hat{b}/\theta_1\right].$$

  Applying Lemma 17 with $x = \mathrm{sgn}(\hat{b}/\theta_1)/\sigma_x$, $a = \mu_x/\sigma_x$, $b = |\hat{b}/\theta_1|$ yield $\mathrm{Var}\left[x_1 | x_1 \geq \hat{b}/\theta_1\right] \geq \sigma_x^2/10$.

- For the second dimension, we have $x_2 \sim \mathcal{N}(u_2, \sigma_x^2)$ and

$$\mathrm{Var}\left[x_2 | \boldsymbol{x}'\hat{\boldsymbol{\theta}} \geq \hat{b}\right] = \mathrm{Var}\left[x_2\right] = \sigma_x^2 > \sigma_x^2/10.$$

- $(x_3, x_4, \ldots, x_d) \sim \mathcal{N}(0, \sigma_x^2 \boldsymbol{I}_{d-2})$. In other words, these characteristics are normally distributed and thus $\mathrm{Var}(x_l) = \sigma_x^2 > \sigma_x^2/10$.

In summary, we have Eq. (10), which concludes the proof. □

Lemma 19 (Martingale Inequality on Ridge Regression, Abbasi-Yadkori et al., 2011)**.** Assume that $||\boldsymbol{\theta}_g|| \leq S$. Take $\delta > 0$ arbitrarily. With probability at least $1 - \delta$, the true parameter $\boldsymbol{\theta}_g$ is bounded as

$$\forall n,\ \left\|\hat{\boldsymbol{\theta}}_g(n) - \boldsymbol{\theta}_g\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \sigma_\epsilon \sqrt{2d \log\left(\frac{\det(\bar{\boldsymbol{V}}_g(n))^{1/2} \det(\lambda \boldsymbol{I})^{-1/2}}{\delta}\right)} + \lambda^{1/2} S. \tag{11}$$

Moreover, let $L = \max_{i,n} \|\boldsymbol{x}_i(n)\|_2$ and

$$\beta_n(L, \delta) = \sigma_\epsilon \sqrt{d \log\left(\frac{1 + nL^2/\lambda}{\delta}\right)} + \lambda^{1/2} S.$$

Then, with probability at least $1 - \delta$,

$$\forall n,\ \left\|\hat{\boldsymbol{\theta}}_g(n) - \boldsymbol{\theta}_g\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \beta_n(L, \delta). \tag{12}$$

The following lemma is used in deriving a regret bound.

Lemma 20 (Sum of Diminishing Contexts, Lemma 11 in Abbasi-Yadkori et al., 2011)**.** Let $\lambda \geq 1$ and $L = \max_{n,i} \|\boldsymbol{x}_i(n)\|_2$. Then, the following inequality holds:

$$\sum_{n:\iota(n)=g} \left\|\boldsymbol{x}_{\iota(n)}\right\|^2_{(\bar{\boldsymbol{V}}_g(n))^{-1}} \leq 2L^2 \log\left(\frac{\det(\bar{\boldsymbol{V}}_g(N))}{\det(\lambda \boldsymbol{I}_d)}\right)$$

for any group g.

The following inequality is used to bound the variation of the minimum eigenvalue of the sum of characteristics (contexts).

Lemma 21 (Matrix Azuma Inequality, Tropp, 2012)**.** Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be adaptive sequence of $d \times d$ symmetric matrices such that $\mathbb{E}_{k-1}\boldsymbol{X}_k = \boldsymbol{0}$ and $\boldsymbol{X}_k^2 \preceq \boldsymbol{A}_k^2$ almost surely, where $\boldsymbol{A} \succeq \boldsymbol{B}$ between two matrices denotes $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite. Let

$$\sigma_A^2 := \left\| \frac{1}{n} \sum_k \boldsymbol{A}_k^2 \right\|$$

where the matrix norm is defined by the largest eigenvalue. Then, for all $t \geq 0$,

$$\Pr\left[\lambda_{\min}\left(\sum_k \boldsymbol{X}_k \right) \leq t \right] \leq d \exp(-t^2/(8n\sigma_A^2)).$$

Proof. The proof directly follows from Theorem 7.1 and Remark 3.10 in Tropp (2012). □

The following lemma states that the selection bias makes its variance slightly ($O(1/\log K)$ times) smaller than the original variance.

Lemma 22 (Variance of Maximum, Theorem 1.8 in Ding, Eldan, and Zhai, 2015)**.** Let $x_1, \ldots, x_K \in \mathbb{R}$ be i.i.d. samples from $\mathcal{N}(0,1)$. Let $I_{\max} = \arg\max_{i \in [K]} x_i$. Then, there exists a distribution-independent constant $C_{\text{varmax}} > 0$ such that

$$\mathrm{Var}[I_{\max}] \geq \frac{C_{\text{varmax}}}{\log(K)}.$$

B Proofs

This section is structured as follows. Section B.1 describes the common inequalities that we assume throughout the section.[28] Proofs of individual theorems are shown in what follows.

B.1 Common Inequalities

In the proofs, we often ignore the events that happen with probability $O(1/N)$. The expected regret per round is at most $\max_i \boldsymbol{x}_i'\boldsymbol{\theta}_{g(i)} - \min_i \boldsymbol{x}_i'\boldsymbol{\theta}_{g(i)}$, which is $O(1)$ in expectation. Hence, the events that happen with probability $O(1/N)$ contributes to the regret by $O(1/N \times N) =$

[28] Except for Theorem 1 that does not pose distributional assumptions.

$O(1)$, which are insignificant in our analysis. In particular,

$$\Pr\left[\forall n \in [N], i \in I(n),\ ||\boldsymbol{x}_i(n)||_2 \leq L\left(\frac{1}{N}\right)\right] \geq 1 - \frac{1}{N}. \quad \text{(by Lemma 16)} \tag{13}$$

Moreover,

$$\Pr\left[\forall n \in [N], g \in G,\ \left\|\hat{\boldsymbol{\theta}}_g(n) - \boldsymbol{\theta}_g\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \beta_n\left(L, \frac{1}{N}\right)\right] \geq 1 - \frac{|G|}{N}. \quad \text{(by Eq. (12) in Lemma 19)} \tag{14}$$

and throughout the proof we ignore the case these events do not hold: All the contexts are bounded by $L(1/N) = O(\sqrt{\log N})$, and all the confidence bounds hold with $\beta_n\left(L(1/N), \frac{1}{N}\right) \leq \beta_N\left(L(1/N), \frac{1}{N}\right) = O(\sqrt{\log N}) = \tilde{O}(1)$, which grows very slowly as N grows large. We also denote $L = L(1/N)$ and $\beta_N = \beta_N\left(L, \frac{1}{N}\right)$.

We next discuss the upper confidence bounds.

Remark 8 (Bound for $\tilde{\boldsymbol{\theta}}_i$)**.** Let $\tilde{\boldsymbol{\theta}}_i = \arg\max_{\bar{\boldsymbol{\theta}}_{g(i)} \in \mathcal{C}_{g(i)}(n)} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}_{g(i)}$. By definition of $\tilde{\boldsymbol{\theta}}_i$, the following inequality always holds:

$$\forall n,\ \left\|\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{g(i)}(n)\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \beta_N \tag{15}$$

and Eq. (14) implies

$$\forall n,\ \boldsymbol{x}_i(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{g(i)}(n)) \geq 0. \tag{16}$$

Moreover, by using triangular inequality, we have

$$\left\|\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_g(n)\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq \left\|\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_g(n)\right\|_{\bar{\boldsymbol{V}}_g(n)} + \left\|\hat{\boldsymbol{\theta}}_g(n) - \boldsymbol{\theta}\right\|_{\bar{\boldsymbol{V}}_g(n)}$$

and thus Eq. (14) implies

$$\forall n,\ \left\|\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_g(n)\right\|_{\bar{\boldsymbol{V}}_g(n)} \leq 2\beta_N. \tag{17}$$

We use the calligraphic font to denote events. For two events $\mathcal{A}, \mathcal{B}$, let $\mathcal{A}^c$ be a complementary event and $\{\mathcal{A}, \mathcal{B}\} := \{\mathcal{A} \cap \mathcal{B}\}$. We also use prime to denote event that is close to the original event. For example, event $\mathcal{A}'$ is different from event $\mathcal{A}$ but these two events are deeply linked. Finally, we discuss the minimum eigenvalue. We denote $\boldsymbol{A} \succeq \boldsymbol{B}$ for two $d \times d$ matrices if $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite: That is, $\lambda_{\min}(\boldsymbol{A} - \boldsymbol{B}) \geq 0$. Note that $\lambda_{\min}(\boldsymbol{A} + \boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})$ and $\lambda_{\min}(\boldsymbol{A} + \boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A})$ if $\boldsymbol{B} \succeq \boldsymbol{0}$. $\boldsymbol{x}\boldsymbol{x}' \succeq \boldsymbol{0}$ for any vector $\boldsymbol{x}\mathbb{R}^d$.

B.2 Proof of Theorem 1

Consider an environment where there are two groups, $G = \{1, 2\}$, and two workers arrive in each round, $K_1 = K_2 = 1$. We assume that error terms follow a standard normal distribution, i.e., $\sigma_\epsilon^2 = 1$. We set the ridge regression parameter λ to be 1. We assume $N_1^{(0)} = 1$ and $N_2^{(0)} = 0$. Hence, $g_1 = 1$.[29] We consider "no context" setting: Namely, $d = 1$, and $x_i = 1$ for all $i \in I$. We assume that $\theta_1 = 0$ and $\theta_2 = -b$ with $b > 0$. Under this assumption, hiring a worker from group 2 incurs regret of b. Let $R_g(n)$ be the sum of the skills of workers who have been hired until round n (i.e., have arrived from round 1 to $n-1$) and belong to group g. Then, (2) implies that $\hat{\theta}_g(n) = R_g(n)/(N_g(n) + \lambda) = R_g(n)/(N_g(n) + 1)$. Since x_i is fixed to 1, firm n chooses a group whose predicted expected skill $\hat{\theta}_g(n)$ is larger.

Since $g_1 = 1$, firm 1 hires the group-1 worker: $\iota(1) = 1$. Let $b' > b$ be a constant that we specify later. Let

$$\mathcal{A} \coloneqq \{\hat{\theta}_1(2) < -b'\} = \{\epsilon_{\iota(1)} < -2b'\}$$

be the event that the skill of the worker hired in round 1 is smaller than $2b'/(1 + \lambda) = b'$. The probability that $\mathcal{A}$ occurs is $\Phi(-2b')$, where $\Phi(x)$ is the cumulative distribution of a standard normal distribution. Let

$$\mathcal{B} \coloneqq \bigcap_{n=1}^{N} \left(\hat{\theta}_2(n) \geq -b'\right)$$

be the event that $\hat{\theta}_2(n)$ never becomes smaller than b'.

We evaluate the probability that $\mathcal{A} \cap \mathcal{B}$ occurs. When such an event happens, a group-1 worker is hired in round 1, and group-2 workers are hired all the subsequent rounds (i.e., $\iota(2) = 2$ for any round $n > 2$). Accordingly, $N_2(n) = n - 2$ is the case for all $n \geq 2$.

$$\left\{\hat{\theta}_2(n) \geq -b'\right\} = \left\{\frac{R_2(n)}{N_2(n) + 1} \geq -b'\right\} \supseteq \left\{\frac{R_2(n)}{N_2(n)} \geq -b'\right\}$$

Applying Hoeffding's inequality to the empirical average $R_g(n)/N_g(n)$, we have

$$\mathbb{P}\left(\frac{R_2(n)}{N_2(n)} < -b'\right) \leq \exp\left(-2(b' - b)^2(n - 2)\right).$$

[29] Note that these the following derivation does not strongly depend on the specific value of these parameters nor the number of groups.

Accordingly,

$$\mathbb{P}(\mathcal{B}) \geq \mathbb{P}\left(\bigcup_{n=1}^{N}\left\{\hat{\theta}_2(n) \geq -b'\right\}\right) \geq 1 - \sum_{n=3}^{N} \exp\left(-2(b'-b)^2(n-2)\right).$$

Here,

$$\begin{aligned}\sum_{n=3}^{N} \exp\left(-2(b'-b)^2(n-2)\right) &\leq \sum_{n=3}^{\infty} \exp\left(-2(b'-b)^2(n-2)\right) \\ &= \frac{\exp\left(-2(b'-b)^2\right)}{1-\exp\left(-2(b'-b)^2\right)} \\ &\leq \frac{1}{2(b'-b)^2},\end{aligned}$$

and thus $\mathbb{P}(\mathcal{B})$ occurs with constant probability $1 - \frac{1}{2(b'-b)^2} > 0$ for any $b' > b + 1/\sqrt{2}$. Remember that $\mathcal{A} \cap \mathcal{B}$ implies that arm 1 is never drawn after $n > 2$, and thus $\mathrm{Reg}(N) \geq bN$. In conclusion, we have

$$\mathbb{E}[\mathrm{Reg}(N)] \geq \Phi(-2b') \cdot \left(1 - \frac{1}{2(b'-b)^2}\right) \cdot b \cdot N = \Omega(N),$$

as desired. □

B.3 Proof of Theorem 2

We first bound regret per round $\mathrm{reg}(n) := \mathrm{Reg}(n) - \mathrm{Reg}(n-1)$ in Lemma 23. Then, we prove Theorem 2.

Lemma 23 (Regret per Round)**.** Under the laissez-faire decision rule, the regret per round is bounded as:

$$\mathrm{reg}(n) \leq 2 \max_{i \in I(n)} \|\boldsymbol{x}_i\|_{\bar{\boldsymbol{V}}_g^{-1}} \left\|\boldsymbol{\theta}_{g(i)} - \hat{\boldsymbol{\theta}}_{g(i)}\right\|_{\bar{\boldsymbol{V}}_g}.$$

Proof of Lemma 23. We denote the first-best decision rule by $i^*(n) := \arg\max_{i \in I(n)} \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)}$. Then,

$$\begin{aligned}\mathrm{reg}(n) &= \boldsymbol{x}_{i^*}' \boldsymbol{\theta}_{g(i^*)} - \boldsymbol{x}_{\iota}' \boldsymbol{\theta}_{g(\iota)} \\ &\leq \boldsymbol{x}_{i^*}' \left(\hat{\boldsymbol{\theta}}_{g(i^*)} + \boldsymbol{\theta}_{g(i^*)} - \hat{\boldsymbol{\theta}}_{g(i^*)}\right) - \boldsymbol{x}_{\iota}' \left(\hat{\boldsymbol{\theta}}_{g(\iota)} + \boldsymbol{\theta}_{g(\iota)} - \hat{\boldsymbol{\theta}}_{g(\iota)}\right) \\ &\leq \boldsymbol{x}_{i^*}' \left(\boldsymbol{\theta}_{g(i^*)} - \hat{\boldsymbol{\theta}}_{g(i^*)}\right) - \boldsymbol{x}_{\iota}' \left(\boldsymbol{\theta}_{g(\iota)} - \hat{\boldsymbol{\theta}}_{g(\iota)}\right) \quad \text{(by the greedy choice of firm)} \\ &\leq \|\boldsymbol{x}_{i^*}\|_{\bar{\boldsymbol{V}}_{g(i^*)}^{-1}} \left\|\boldsymbol{\theta}_{g(i^*)} - \hat{\boldsymbol{\theta}}_{g(i^*)}\right\|_{\bar{\boldsymbol{V}}_{g(i^*)}} + \|\boldsymbol{x}_{\iota}\|_{\bar{\boldsymbol{V}}_{g(\iota)}^{-1}} \left\|\boldsymbol{\theta}_{g(\iota)} - \hat{\boldsymbol{\theta}}_{g(\iota)}\right\|_{\bar{\boldsymbol{V}}_{g(\iota)}}\end{aligned}$$

$$
\begin{aligned}
&\text{(by the Cauchy–Schwarz inequality)} \\
&\leq 2 \max_{i \in I(n)} \|\boldsymbol{x}_i\|_{\bar{\boldsymbol{V}}_{g(i)}^{-1}} \left\| \boldsymbol{\theta}_{g(i)} - \hat{\boldsymbol{\theta}}_{g(i)} \right\|_{\bar{\boldsymbol{V}}_{g(i)}} .
\end{aligned}
$$

□

The following proves Theorem 2. For the ease of discussion, we assume $N^{(0)} = 0$. That is, there is no initial sampling phase. Taking it into consideration is trivial. We first show that regardless of estimated values $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2$, the candidate of group 2 is drawn with constant probability. Let $\mu_x = ||\boldsymbol{\mu}_x||$. Let

$$
\begin{aligned}
\mathcal{M}_1(n) &= \left\{ \boldsymbol{x}_1'(n) \hat{\boldsymbol{\theta}}_1 \leq 0 \right\} \\
\mathcal{M}_2(n) &= \left\{ \boldsymbol{x}_2'(n) \hat{\boldsymbol{\theta}}_2 > 0 \right\}
\end{aligned}
$$

The sign of $\boldsymbol{x}_1' \hat{\boldsymbol{\theta}}_1(n)$ is solely determined by the component of $\boldsymbol{x}_1(n)$ that is parallel to $\hat{\boldsymbol{\theta}}_1(n)$. This component is drawn from $\mathcal{N}(\mu_{x,\|}, \sigma_x)$ where $\mu_{x,\|}$ is the component of $\boldsymbol{\mu}_x$ that is parallel to $\hat{\boldsymbol{\theta}}_1(n)$. Therefore, for any $\hat{\boldsymbol{\theta}}_1$, we have[30]

$$
\Pr[\mathcal{M}_1(n)] \geq \Phi^c(\mu_x / \sigma_x). \tag{18}
$$

Likewise, for $\hat{\boldsymbol{\theta}}_2 \neq 0$, we have[31]

$$
\Pr[\mathcal{M}_2(n)] \geq \Phi^c(\mu_x / \sigma_x) \tag{19}
$$

Let $\mathcal{X}_2(n) = \{g(\iota(n)) = g\}$ for $g \in \{1, 2\}$. By using Eq. (18) and (19),

$$
\begin{aligned}
\Pr[\mathcal{X}_2(n)] &= \Pr[x_1'(n)\hat{\theta}_1 < x_2'(n)\hat{\theta}_2] \\
&\geq \Pr[x_1'(n)\hat{\theta}_1 \leq 0 < x_2'(n)\hat{\theta}_2] \\
&= \Pr[\mathcal{M}_1(n), \mathcal{M}_2(n)] \\
&\geq \left(\Phi^c(\mu_x/\sigma_x)\right)^2 . \qquad (20) \\
&\quad \text{(by Eq. (18), (19))}
\end{aligned}
$$

Let $N_2^{(\mathcal{M})}(n) = \sum_{n'=1}^{n} \mathbf{1}[\mathcal{M}_1(n'), \mathcal{X}_2(n')] \leq N_2(n)$. Eq. (20) implies $\mathbb{E}[N_2^{(\mathcal{M})}(n)] \geq \left(\Phi^c(\mu_x/\sigma_x)\right)^2 n$. By using the Hoeffding inequality, with probability at least $1 - 2/N^2$, we have

$$
N_2^{(\mathcal{M})} \geq n \left(\left(\Phi^c(\mu_x/\sigma_x)\right)^2 - k \right) \tag{21}
$$

[30] $\Pr[\mathcal{M}(n)] = \Phi^c(\mu_x/\sigma_x)$ when $\mu_{x,\|} = \mu_x$. Namely, the direction of $\boldsymbol{\mu}_x$ is exactly the same as $\hat{\boldsymbol{\theta}}_1$.

[31] In the subsequent discussion, we do not care point mass $\hat{\boldsymbol{\theta}}_2 = 0$ of measure zero for $N_2(n) > 0$.

for

$$k = \sqrt{\frac{\log(N)}{n}}.$$

Therefore, union bound over $n = 1, 2, \ldots, N$ implies Eq. (21) holds with probability at least $1 - \sum_n 2/N^2 = 1 - 2/N$.

In the following we bound the $\lambda_{\min}(\bar{\boldsymbol{V}}_g)$. Note that a hiring of a worker i_2 under events $\mathcal{M}_1(n), \mathcal{X}_2(n)$ satisfies a diversity condition (Lemma 18) with $\hat{b} = 0$, and we have

$$\lambda_{\min}(\mathbb{E}[\boldsymbol{x}_\iota \boldsymbol{x}_\iota' | \mathcal{M}_1(n), \mathcal{X}_2(n)]) \geq \lambda_0$$

with $\lambda_0 = \sigma_x^2/10$. Using the matrix Azuma inequality (Lemma 21) for subsequence $\{\boldsymbol{x}_\iota \boldsymbol{x}_\iota' : \mathcal{M}_1(n), \mathcal{X}_2(n)\}$ with $\boldsymbol{X} = \boldsymbol{x}_\iota \boldsymbol{x}_\iota' - \mathbb{E}[\boldsymbol{x}_\iota \boldsymbol{x}_\iota']$ and $\sigma_A = 2L^2$, for $t = \sqrt{32 N_2 \sigma_A^2} \log(dN)$, with probability $1 - 1/N$

$$\lambda_{\min}\left(\sum_{n:\iota(n)=2} \boldsymbol{x}_\iota \boldsymbol{x}_\iota'\right) \geq N_2^{(\mathcal{M})} \lambda_0 - t. \tag{22}$$

In summary, with probability $1 - 4/N$, Eq. (21) and (22) hold, and then, we have

$$\begin{aligned}
\lambda_{\min}(\bar{V}_2) &\geq N_2^{(\mathcal{M})} \lambda_0 - \sqrt{32 N_2 \sigma_A^2} \log(dN) \\
&\geq (n(\Phi^c(\mu_x/\sigma_x))^2 - k)\lambda_0 - \sqrt{32 N_2 \sigma_A^2} \log(dN) \\
&= n(\Phi^c(\mu_x/\sigma_x))^2 \lambda_0 - \tilde{O}(\sqrt{n}).
\end{aligned} \tag{23}$$

By using the symmetry of the two groups, exactly the same results as Eq. (23) holds for group 1.

In the following, we bound the regret as a function of $\min_g \lambda_{\min}(\bar{V}_g)$. Eq. (23) holds with probability $1 - O(1/N)$, and we ignore events of probability $O(1/N)$ that does not affect the analysis. The regret is bounded as

$$\begin{aligned}
\mathrm{Reg}(N) &\leq 2 \sum_n \max_i \|\boldsymbol{x}_i\|_{\bar{\boldsymbol{V}}_{g(i)}^{-1}} \left\| \boldsymbol{\theta}_{g(i)} - \hat{\boldsymbol{\theta}}_{g(i)} \right\|_{\bar{\boldsymbol{V}}_{g(i)}} \quad \text{(by Lemma 23)} \\
&\leq 2 \sum_n \max_i \|\boldsymbol{x}_i\|_{\bar{\boldsymbol{V}}_{g(i)}^{-1}} \beta_N \quad \text{(by Eq. 14)} \\
&\leq 2 \sum_n \max_i \frac{||\boldsymbol{x}_i||}{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})} \beta_N \quad \text{(by definition of eigenvalues)} \\
&\leq 2 \sum_n \max_i \frac{L}{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})} \beta_N \quad \text{(by Eq. (13))}
\end{aligned} \tag{24}$$

$$
\begin{aligned}
&\leq 2L \sum_n \max_i \min\left(\frac{1}{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}, \frac{1}{\lambda}\right)\beta_N \quad (\text{by } \lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)}) \geq \lambda) \\
&\leq 2L \sum_n \min\left(\sqrt{\frac{1}{n(\Phi^c(\mu_x/\sigma_x))^2\lambda_0 - \tilde{O}(\sqrt{n})}}, \frac{1}{\lambda}\right)\beta_N \quad (\text{by Eq. (23)}) \\
&\leq 4L\sqrt{\frac{N}{(\Phi^c(\mu_x/\sigma_x))^2\lambda_0}}\beta_N + \tilde{O}(1) \\
&\quad \left(\text{by } \sum_{n=C^2+1}^{N}\left\{\frac{1}{\sqrt{n - C\sqrt{n}}}\right\} = 2\sqrt{N} + \tilde{O}(1) \text{ for } C = \tilde{O}(1)\right)
\end{aligned}
$$

which completes Proof of Theorem 2.

B.4 Proof of Theorem 3

Since we consider $d = 1$ case in this theorem, we remove bold styles in scalar variables. In this proof, we assume $\mu_x\theta > 0$ and $\theta > 0$. The proof for the case of $\mu_x\theta < 0$ or $\theta < 0$ is similar. Let $\hat{\theta}_{g,t}$ be the value of $\hat{\theta}_g$ when group g candidate was chosen t times. With a slight abuse of notation, we use $i_2 = i_2(n)$ to denote the unique candidate of group 2 in each round n. We first define the several events that characterize the perpetual underestimation. Namely,

$$
\begin{aligned}
\mathcal{P} &= \left\{\left|\hat{\theta}_{2,N_2^{(0)}}\right| < \frac{b}{2}\theta\right\} \\
\mathcal{P}'(n) &= \left\{x_{i_2(n)}\hat{\theta}_{2,N_2^{(0)}} < \frac{1}{2}\mu_x\theta\right\} \\
\mathcal{Q} &= \left\{\forall t \geq N_1^{(0)},\ \hat{\theta}_{1,t} \geq \frac{1}{2}\theta\right\} \\
\mathcal{Q}'(n) &= \left\{\exists i \text{ s.t. } g(i) = 1, x_i\hat{\theta}_{1,N_1(n)} \geq \frac{1}{2}\mu_x\theta\right\}
\end{aligned}
$$

where b is a small[32] constant that we specify later. $\mathcal{P}$ and $\mathcal{P}'$ are about the minority whereas $\mathcal{Q}$ and $\mathcal{Q}'$ are about the majority: Intuitively, Event $\mathcal{P}$ states that $\hat{\theta}_2$ is largely underestimated, and $\mathcal{P}'$ states that the minority candidate is undervalued. $\mathcal{Q}$ states that the majority parameter $\hat{\theta}_1$ is consistently lower-bounded, and $\mathcal{Q}'$ states the stability of the best candidate of the majority after n rounds. Under laissez-faire,

$$
\bigcap_{n=1}^{N}(\mathcal{P}'(n) \cap \mathcal{Q}'(n))
$$

[32] We will specify $b = O(1/(\log N))$.

implies the majority candidate is always chosen ($g(\iota) = 1$ for all n), which is exactly the perpetual underestimation of Definition 2. Therefore, proving

$$\Pr\left[\bigcap_{n=1}^{N}(\mathcal{P}'(n) \cap \mathcal{Q}'(n))\right] \geq \tilde{O}(1) \tag{25}$$

concludes the proof. We bound these events by the following lemmas and finally derives Eq. (25).

Lemma 24.

$$\Pr[\mathcal{P}] \geq C_1 b$$

for some constant C_1.

Proof of Lemma 24. We denote $x_{i_2,t}$ for representing t-th sample of group 2 during the initial sampling phase, which is an i.i.d. sample from $\mathcal{N}(\mu_x, \sigma_x^2)$. Likewise, we also denote $y_{i_2,t} = x_{i_2,t}\theta + \epsilon_t$.

$$\begin{aligned}
\Pr[\mathcal{P}] &= \Pr\left[\left|\frac{\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}(x_{i_2,t}\theta + \epsilon_t)}{\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2 + \lambda}\right| \leq \frac{b}{2}\theta\right] \\
&= \Pr\left[\left|\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}(x_{i_2,t}\theta + \epsilon_t)\right| \leq \frac{b}{2}\theta\left(\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2 + \lambda\right)\right] \\
&= \Pr\left[-g(b) \leq \sum_{t=1}^{N_2^{(0)}} x_{i_2,t}(x_{i_2,t}\theta + \epsilon_t) \leq g(b)\right]
\end{aligned}$$

where

$$g(b) = \frac{b}{2}\theta\left(\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2 + \lambda\right).$$

Let $x_{i_2,t} = \mu_x + e_t$. Define an event $\mathcal{R}$ as follows.

$$\mathcal{R} = \left\{\sum_{t=1}^{N_2^{(0)}} e_t^2 \leq 5\sigma_x^2 N_2^{(0)}\right\} \subseteq \left\{\sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2 \leq 2N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)\right\}$$

where we used $x_{i_2,t}^2 = (\mu_x + e_t)^2 \leq 2(\mu_x^2 + e_t^2)$ in the last transformation. By using Lemma 14, we have

$$\Pr[\mathcal{R}^c] \leq 1 - 2e^{-2N_2^{(0)}} \leq 1/4.$$

Moreover, let

$$\mathcal{S} = \left\{ \sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2 = \sum_{t=1}^{N_2^{(0)}} (\mu_x + e_t)^2 \geq N_2^{(0)} \mu_x^2 \right\}.$$

It is easy to confirm that $\Pr[\sum_n (\mu_x + e_t)^2 \geq N_2^{(0)} \mu_x^2] \geq 1/2$, and thus

$$\Pr[\mathcal{R} \cap \mathcal{S}] \geq 1 - 1/4 - 1/2 = 1/4. \tag{26}$$

Note that $\mathcal{S}$ implies

$$g(b) \geq \frac{b}{2} \theta N_2^{(0)} \mu_x + \lambda. \tag{27}$$

Conditioned on $x_{i_2,t}$, we have $x_{i_2,t}\epsilon_t \sim \mathcal{N}(0, x_{i_2,t}^2 \sigma_\epsilon^2)$. Moreover, by using the property on the sum of independent normal random variables,

$$\sum_t x_{i_2,t}\epsilon_t \sim \mathcal{N}(0, \sum_t x_{i_2,t}^2 \sigma_\epsilon^2) \tag{28}$$

Letting

$$L_R = \frac{-g(b) - \sum_t x_{i_2,t}^2 \theta}{\sigma_\epsilon \sqrt{\sum_t x_{i_2,t}^2}}$$
$$U_R = \frac{g(b) - \sum_t x_{i_2,t}^2 \theta}{\sigma_\epsilon \sqrt{\sum_t x_{i_2,t}^2}}$$
$$M_R = \frac{L_R + U_R}{2} = \frac{-\left(\sqrt{\sum_t x_{i_2,t}^2}\right)\theta}{\sigma_\epsilon}$$

We have

$$\begin{aligned}
&\Pr\left[-g(b) \leq \sum_{t=1} (x_{i_2,t}^2 \theta + x_{i_2,t}\epsilon_n) \leq g(b) \right] \\
&\geq \Pr\left[-g(b) \leq \sum_{t=1} (x_{i_2,t}^2 \theta + x_{i_2,t}\epsilon_t) \leq g(b), \mathcal{R}, \mathcal{S} \right] \\
&\geq \Pr\left[-g(b) - \sum_{t=1} x_{i_2,t}^2 \theta \leq \sum_{t=1} x_{i_2,t}\epsilon_t \leq g(b) - \sum_{t=1} x_{i_2,t}^2 \theta, \mathcal{R}, \mathcal{S} \right] \\
&\geq \Pr[\mathcal{R}, \mathcal{S}] \min_{\{e_n : \mathcal{R}, \mathcal{S}\}} \left[\int_{L_R}^{U_R} \phi(y) dy \right] \quad \text{(by Eq. (28))} \\
&\geq \frac{1}{4} \min_{\{e_n : \mathcal{R}, \mathcal{S}\}} \left[\int_{L_R}^{U_R} \phi(y) dy \right] \quad \text{(by Eq. (26))}
\end{aligned} \tag{29}$$

The following bounds Eq. (29). The integral's bandwidth is

$$U_R - L_R = \frac{2g(b)}{\sigma_\epsilon \sqrt{\sum_t x_{i_2,t}^2}} \geq \frac{2g(b)}{\sigma_\epsilon \sqrt{2N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)}}. \quad \text{(by event } \mathcal{R})$$

The value of $\phi(y)$ within $[M_R - 1, M_R + 1]$ is at least $\phi(M_R)/e^{1/2} \geq (1/2)\phi(M_R)$. Therefore,

$$\int_{L_R}^{U_R} \phi(y)dy \geq \min\left(2, \frac{2g(b)}{\sigma_\epsilon \sqrt{2N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)}}\right) \times \frac{\phi(M_R)}{2}. \tag{30}$$

Moreover,

$$\begin{aligned}
\phi(M_R) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(M_R)^2}{2}\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2 \sum_{t=1}^{N_2^{(0)}} x_{i_2,t}^2}{2\sigma_\epsilon^2}\right) \\
&\leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{2\theta^2 N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)}{2\sigma_\epsilon^2}\right) \quad \text{(by event } \mathcal{R})
\end{aligned} \tag{31}$$

By using these, we have

$$\begin{aligned}
\int_{L_R}^{U_R} \phi(y)dy &\geq \min\left(2, \frac{2g(b)}{\sigma_\epsilon \sqrt{2(\mu_x^2 + 5\sigma_x^2)}}\right) \frac{\phi\left(M_R\right)}{2} \quad \text{(by Eq. (30))} \\
&= \min\left(1, \frac{g(b)}{\sigma_\epsilon \sqrt{2N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)}}\right) \phi(M_R) \\
&= O\left(b\sqrt{N_2^{(0)}} \exp\left(-\frac{2\theta^2 N_2^{(0)}(\mu_x^2 + 5\sigma_x^2)}{2\sigma_\epsilon^2}\right)\right) \quad \text{(by Eq. (27), (31))}
\end{aligned}$$

The exponent does not depend on b: Given all model parameters as constant, the probability of $\mathcal{P}$ is $O(b)$, which concludes the proof. □

The following Lemma 25 on $\mathcal{Q}$ is about the stability of the mean estimator, which is widely used to prove lower bounds in multi-armed bandit problems. Namely, for any $\Delta > 0$, a wide class of mean estimators $\hat{\theta}$ of θ satisfies

$$\Pr\left[\bigcup_{n=1}^{\infty} \left(\hat{\theta}(n) \geq \theta - \Delta\right)\right] \geq C \tag{32}$$

for some constant $C = C(\theta, \Delta) > 0$. Lemma 25 is a version Eq. (32) for our ridge estimator.

Lemma 25. There exists a constant $N_1^{(0)}$ that is independent on N such that, with a warm-start of size $N_1^{(0)}$,

$$\Pr[\mathcal{Q}] \geq C_2$$

holds.

Proof of Lemma 25. In this proof, we use $t \geq 0$ to denote the estimator where the t-th sample is drawn. For example, $\bar{V}_{g,t} := \bar{V}_g(n)$ of $n : N_1(n-1) = t$. Note that we consider $d = 1$ case and $\bar{V}_{1,t} = \sum_{t'=1}^{t} x_{1,t}^2 + \lambda$. By martingale bound (Eq. (11)), with probability $1-\delta$,

$$\forall t \geq 1, \quad |\hat{\theta}_{1,t} - \theta| \sqrt{\bar{V}_{1,t}} \leq \sigma_\epsilon \sqrt{\log\left(\frac{\bar{V}_{1,t}^{1/2}\lambda^{-1/2}}{\delta}\right)} + \lambda^{1/2} S. \tag{33}$$

Let $\delta = 1/2$. It follows from $\sqrt{\log x} \leq \sqrt{x}$ for any $x > 0$ that

$$\sqrt{\log\left(2\bar{V}_{1,t}^{1/2}\lambda^{-1/2}\right)} \leq \sqrt{2\bar{V}_{1,t}^{1/2}\lambda^{-1/2}}. \tag{34}$$

Therefore,

$$\begin{aligned}
|\hat{\theta}_{1,t} - \theta| &\leq \frac{\sigma_\epsilon \sqrt{\log\left(\frac{\bar{V}_{1,t}^{1/2}\lambda^{-1/2}}{\delta}\right)} + \lambda^{1/2} S}{\sqrt{\bar{V}_{1,t}}} \quad \text{(by Eq. (33))} \\
&\leq \frac{\sigma_\epsilon \sqrt{2\bar{V}_{1,t}^{1/2}\lambda^{-1/2}} + \lambda^{1/2} S}{\sqrt{\bar{V}_{1,t}}} \quad \text{(by (34))}
\end{aligned}$$

and thus

$$\forall t \geq N_1^{(0)}, \ |\hat{\theta}_{1,t} - \theta| \leq \frac{1}{2}|\theta|$$

holds if

$$\sqrt{\bar{V}_{1,N_1^{(0)}}} \geq 2\theta \max\left(\sigma_\epsilon \sqrt{2\bar{V}_{1,N_1^{(0)}}^{1/2}\lambda^{-1/2}}, \lambda^{1/2} S\right)$$

whose sufficient condition for the initial sample size $N_1^{(0)}$ is

$$\bar{V}_{1,N_1^{(0)}} \geq \max\left[\frac{64}{\theta^4}(\sigma_\epsilon^4/\lambda^2), \frac{4}{\theta^2}\lambda S^2\right].$$

Note that $\Pr[\bar{V}_{1,N_1^{(0)}} \geq \mu_x^2 N_1^{(0)}] \geq 1/2$. Letting the observation noise σ_ϵ and regularizer λ be constants, constant size of warm-start is enough to assure this bound with probability $C_2 = 1/2 \times 1/2 = 1/4$. □

The following lemma states that, when $\hat{\theta}_2$ is very small, the estimated quality $x_{i_2}\hat{\theta}_2$ of the minority group is likely to be small.

Lemma 26. There exists a constant C_3, C_4 that is independent of N such that

$$\Pr[\mathcal{P}'(n)|\mathcal{P}] \geq 1 - C_3 \exp\left(-C_4/b\right) \tag{35}$$

holds.

Proof of Lemma 26.

$$\begin{aligned}
\Pr[\mathcal{P}'(n)|\mathcal{P}] &\geq 1 - \Pr\left[x_{i_2}(n) \geq \frac{2}{b}\right] \\
&\geq 1 - \Phi^c\left(\frac{1}{\sigma_x}\left(\frac{2}{b} - \mu_x\right)\right) \\
&\geq 1 - \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{1}{\sigma_x}\left(\frac{2}{b} - \mu_x\right)\right), \quad \text{(by Lemma 15)}
\end{aligned}$$

where we have assumed $\left(\frac{2}{b} - \mu\right)/\sigma_x \geq 1$ in the last transformation (which holds for sufficiently small b). Eq. (35) holds for $C_3 = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\mu/\sigma_x}$ and $C_4 = 2/\sigma_x$. □

Lemma 27.

$$\Pr[\mathcal{Q}'(n) \mid \mathcal{Q}] \geq 1 - (1/2)^{K_1}.$$

Event $\mathcal{Q}'(n)$ states that all the candidates' estimated quality $x_i\hat{\theta}$ is not below mean. Lemma 27 states that the probability of $\mathcal{Q}'(n)$ is exponentially small to the number of candidates. The proof of Lemma 27 directly follows from the symmetry of normal distribution and independence of each characteristic $\boldsymbol{x}_i$.

Proof of Theorem 3. By using Lemmas 24–27, we have

$$\Pr[\mathcal{P}] \geq C_1 b \tag{36}$$

$$\Pr[\mathcal{Q}] \geq C_2 \tag{37}$$

$$\Pr[\mathcal{P}'(n)|\mathcal{P}] \geq 1 - C_3 \exp\left(-C_4 b\right) \tag{38}$$

$$\Pr[\mathcal{Q}'(n)|\mathcal{Q}] \geq 1 - (1/2)^{K_1}$$

From these equations, the probability of perpetual underestimation is bounded as:

$$
\begin{aligned}
&\Pr\left[\bigcup_n \{\iota(n)=1\}\right] \\
&\geq \Pr\left[\bigcup_n \{\mathcal{P}'(n),\mathcal{Q}'(n)\},\mathcal{P},\mathcal{Q}\right] \\
&\geq \Pr\left[\mathcal{P}\right]\Pr\left[\mathcal{Q}\right]\Pr\left[\bigcup_n \{\mathcal{P}'(n),\mathcal{Q}'(n)\} \mid \mathcal{P},\mathcal{Q}\right] \quad \text{(by the independence of } \mathcal{P} \text{ and } \mathcal{Q}) \\
&\geq C_1 b \times C_2 \times (1 - NC_3 \exp\left(-C_4 b\right)) \times \left(1 - N\left(\frac{1}{2}\right)^{K_1}\right) \quad \text{(by the union bound)} \quad (39)
\end{aligned}
$$

which, by letting $b = O(1/\log(N))$ and $K_1 > \log_2(N)$, is $\tilde{O}(1)$. □

B.5 Proof of Theorem 4

Proof. Let $\mathrm{reg}(n) = \mathrm{Reg}(n) - \mathrm{Reg}(n-1)$. Notice that under the UCB decision rule,

$$\iota(n) = \max_{i \in I(n)} (\boldsymbol{x}_i' \tilde{\boldsymbol{\theta}}_i(n)). \tag{40}$$

By Lemma 19, with probability at least $1-\delta$, the true parameter of group g lies in $\mathcal{C}_g$, and thus

$$\boldsymbol{x}_i' \tilde{\boldsymbol{\theta}}_i(n) \geq \boldsymbol{x}_i' \boldsymbol{\theta}_g \tag{41}$$

for each $i \in I(n)$.

Let $i^* = i^*(n) := \arg\max_{i \in I(n)} \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)}$ be the first-best worker, and $g^* = g(i^*)$ be the group i^* belongs to. The regret in round n is bounded as

$$
\begin{aligned}
\mathrm{reg}(n) &= \boldsymbol{x}_{i^*}' \boldsymbol{\theta}_{g^*} - \boldsymbol{x}_{\iota}' \boldsymbol{\theta}_{g(\iota)} \\
&\leq \boldsymbol{x}_{i^*}' \tilde{\boldsymbol{\theta}}_{i^*} - \boldsymbol{x}_{\iota}' \boldsymbol{\theta}_{g(\iota)} \quad \text{(by Eq. (41))} \\
&\leq \boldsymbol{x}_{\iota}' \tilde{\boldsymbol{\theta}}_{\iota} - \boldsymbol{x}_{\iota}' \boldsymbol{\theta}_{g(\iota)} \quad \text{(by Eq. (40))} \\
&\leq ||\boldsymbol{x}_{\iota}'||_{\bar{\boldsymbol{V}}_{g(\iota)}^{-1}} \left\| \boldsymbol{\theta}_{g(\iota)} - \tilde{\boldsymbol{\theta}}_{\iota} \right\|_{\bar{\boldsymbol{V}}_{g(\iota)}} \quad \text{(by the Cauchy–Schwarz inequality)} \\
&\leq ||\boldsymbol{x}_{\iota}'||_{\bar{\boldsymbol{V}}_{g(\iota)}^{-1}} \beta_N. \quad \text{(by Eq. (14))} \qquad (42)
\end{aligned}
$$

The total regret is bounded as:

$$\begin{aligned}
\mathrm{Reg}(N) = \sum_n \mathrm{reg}(n) &\le \sqrt{N \sum_n \mathrm{reg}(n)^2} \quad \text{(by the Cauchy–Schwarz inequality)} \\
&\le 2\beta_N \sqrt{N \sum_n ||\boldsymbol{x}'_\iota||^2_{\bar{\boldsymbol{V}}^{-1}_{g(\iota)}}(n)} \\
&\le 2\beta_N \sqrt{2NL^2 \sum_{g \in G} \log(\det(\bar{\boldsymbol{V}}_g(N)))} \quad \text{(by Lemma 20)} \qquad (43) \\
&\le \tilde{O}(\sqrt{N|G|})
\end{aligned}$$

where we have used the fact that $\log(\det(\bar{\boldsymbol{V}}_g)) = O(\log(N)) = \tilde{O}(1)$. □

B.6 Proof of Theorem 5

Proof of the first statement Since $s_i(n) = \tilde{q}_i(n) - \hat{q}_i(n)$, we have $\hat{q}_i(n) + s_i(n) = \tilde{q}_i(n)$. Hence, firm i's incentive is aligned with the UCB index. Accordingly, firm i follows the UCB decision rule, which maximizes the UCB index.

Proof of the second statement For notational simplicity, we drop n, $\boldsymbol{X}_g$, Y_g from this proof. Define a correspondence $\mathcal{U}$ by

$$\mathcal{U}(\tilde{q}_i; s) \coloneqq \{u_i \in \mathbb{R} \mid \exists i, \exists \boldsymbol{x}_i \text{ s.t. } \hat{q}_i(\boldsymbol{x}_i) + s_i(\boldsymbol{x}_i) = u_i, \tilde{q}_i = \tilde{q}_i(\boldsymbol{x}_i)\}.$$

The set $\mathcal{U}(\tilde{q}_i)$ represents the set of firm n's all possible payoffs from a worker whose UCB index is $\tilde{q}_i$.

Clearly, the subsidy rule s implements the UCB decision rule ι if and only if for all i, $\tilde{q}'_i > \tilde{q}_i$ implies

$$\min \mathcal{U}(\tilde{q}'_i; s) > \max \mathcal{U}(\tilde{q}_i; s). \qquad (44)$$

Since $\min \mathcal{U}(\cdot; s_i)$ is an increasing function, it is continuous at all but countably many points. Equivalently, $\mathcal{U}(\tilde{q}_i; s_i)$ is a singleton for almost all values of $\tilde{q}_i$.

Now, suppose that $\mathcal{U}(\tilde{q}^*_i)$ is not a singleton for some $\tilde{q}^*_i$. Define Δ by

$$\Delta \coloneqq \max \mathcal{U}(\tilde{q}^*_i; s) - \min \mathcal{U}(\tilde{q}^*_i; s).$$

Define another subsidy rule s' by setting

$$s'_i(\boldsymbol{x}_i) = \begin{cases} s_i(\boldsymbol{x}_i) & \text{if } \tilde{q}_i(\boldsymbol{x}_i) < \tilde{q}_i^* \\ \min \mathcal{U}(\tilde{q}_i^*) - \hat{q}_i(\boldsymbol{x}_i) & \text{if } \tilde{q}_i(\boldsymbol{x}_i) = \tilde{q}_i^* \\ s_i(\boldsymbol{x}_i) - \Delta & \text{otherwise} \end{cases}$$

for all i. Then, we have

$$\mathcal{U}(\tilde{q}_i; s') = \begin{cases} \mathcal{U}(\tilde{q}_i; s) & \text{if } \tilde{q}_i < \tilde{q}_i^* \\ \{\min \mathcal{U}(\tilde{q}_i^*; s)\} & \text{if } \tilde{q}_i = \tilde{q}_i^* \\ \mathcal{U}(\tilde{q}_i; s) - \Delta & \text{otherwise,} \end{cases}$$

which implies $\mathcal{U}(\cdot; s')$ also satisfies (44), or equivalently, s' also implements the UCB rule ι. Furthermore, $s'_i(\boldsymbol{x}_i) \leq s'_i(\boldsymbol{x}_i)$ for all $\boldsymbol{x}_i$, with a strict inequality for some $\boldsymbol{x}_i$. Accordingly, s' needs a smaller budget than s.

By the argument above, whenever $\mathcal{U}(\cdot; s_i)$ does not returns singleton for some $\tilde{q}_i$, the subsidy amount can be improved by filling a gap. From now, we discuss the case in which $\mathcal{U}(\cdot; s_i)$ returns a singleton for all $\tilde{q}_i$; i.e., $\mathcal{U}$ reduces to a function. From now, we use $u(\tilde{q}_i; s)$ to represent the firm's utility when it hires a worker whose UCB index is $\tilde{q}_i$ (which was previously written as $\mathcal{U}$ because it could take multiple values). Then, we have

$$s_i(\boldsymbol{x}_i) = u(\tilde{q}_i(\boldsymbol{x}_i); s) - \hat{q}_i(\boldsymbol{x}_i)$$

for all $\boldsymbol{x}_i$. Since we require that $s_i(\boldsymbol{x}_i) \geq 0$ for all $\boldsymbol{x}_i$,

$$u(\tilde{q}_i(\boldsymbol{x}_i); s) - \hat{q}_i(\boldsymbol{x}_i) \geq 0.$$

After some history, $\hat{q}_i$ may become arbitrarily close to $\tilde{q}_i$. Therefore, the inequality is satisfied for all $\tilde{q}_i$ and $\hat{q}_i$. Accordingly, u must satisfy

$$u(q; s) \geq q \tag{45}$$

for all q. The UCB index subsidy rule satisfies (45) with equalities for all q: The UCB index subsidy rule satisfies $s_i = \tilde{q}_i - \hat{q}_i$, and therefore, $u(\tilde{q}_i; s) = \tilde{q}_i$ for all $\tilde{q}_i$. Accordingly, it needs the minimum possible budget.

Proof of the third statement We bound the amount of total subsidy $\mathrm{Sub}(N)$.

$$\begin{aligned}\mathrm{Sub}(N) &:= \sum_n \boldsymbol{x}'_{\iota(n)}(\tilde{\boldsymbol{\theta}}_\iota - \hat{\boldsymbol{\theta}}_{g(\iota)}) \\ &\leq ||\boldsymbol{x}'_\iota||_{\bar{\boldsymbol{V}}^{-1}_{g(\iota)}}\beta_N, \quad \text{(by Eq. (15))}\end{aligned}$$

which is the same as Eq. (42) and thus the same bound as regret applies. □

B.7 Proofs of Theorems 8 and 9

Proof of Theorem 8. We adopt "slot" notation for each group. Group g is allocated K_g slots and at each round n, one candidate arrives for each slot. We use index $i \in [K]$ to denote each slot: Although $\boldsymbol{x}_i$ at two different rounds n,n' $(= \boldsymbol{x}_i(n), \boldsymbol{x}_i(n'))$ represent different candidates, they are from the identical group $g = g(i)$. In summary, we use index i to represent the i-th slot and with a slight abuse of argument. We also call candidate i to represent the candidate of slot i. Note that this does not change any part of the model, and the slot notation here is for the sake of analysis.

Under the hybrid decision rule, a firm at each round hires the candidate of the largest index. Namely,

$$\iota(n) = \arg\max_{i \in I(n)} \tilde{q}^{\mathrm{H}}_i(n)$$

where $\tilde{q}^{\mathrm{H}}_i$ is defined at Eq. (6). We also denote $\tilde{\iota}(n) = \arg\max_{i \in I(n)} \boldsymbol{x}'_i\tilde{\boldsymbol{\theta}}_i$. That is, $\tilde{\iota}$ indicates the candidate who would have been hired if we have used the standard UCB decision rule (Eq. (5))

The following bounds the regret into estimation errors of $\tilde{\iota}$ and ι.

$$\begin{aligned}\mathrm{reg}(n) &= \boldsymbol{x}'_{i^*}\boldsymbol{\theta}_{g^*} - \boldsymbol{x}'_\iota\boldsymbol{\theta}_{g(\iota)} \\ &\leq \boldsymbol{x}'_{i^*}\tilde{\boldsymbol{\theta}}_{i^*} - \boldsymbol{x}'_\iota\boldsymbol{\theta}_{g(\iota)} \quad \text{(by Eq. (16))} \\ &\leq \boldsymbol{x}'_{\tilde{\iota}}\tilde{\boldsymbol{\theta}}_{\tilde{\iota}} - \boldsymbol{x}'_\iota\boldsymbol{\theta}_{g(\iota)} \quad \text{(by definition of } \tilde{\iota}) \\ &= \boldsymbol{x}'_{\tilde{\iota}}\tilde{\boldsymbol{\theta}}_{\tilde{\iota}} - \boldsymbol{x}'_\iota\tilde{\boldsymbol{\theta}}_\iota + \boldsymbol{x}'_\iota(\tilde{\boldsymbol{\theta}}_\iota - \boldsymbol{\theta}_{g(\iota)}) \\ &\leq \boldsymbol{x}'_{\tilde{\iota}}(\tilde{\boldsymbol{\theta}}_{\tilde{\iota}} - \hat{\boldsymbol{\theta}}_{g(\tilde{\iota})}) + \boldsymbol{x}'_\iota(\tilde{\boldsymbol{\theta}}_\iota - \boldsymbol{\theta}_{g(\iota)}) \quad \text{(by definition of } \iota) \qquad (46)\end{aligned}$$

Here,

$$\begin{aligned}\boldsymbol{x}'_{\tilde{\iota}}(\tilde{\boldsymbol{\theta}}_{\tilde{\iota}} - \hat{\boldsymbol{\theta}}_{g(\tilde{\iota})}) &\leq ||\boldsymbol{x}'_{\tilde{\iota}}||_{\bar{\boldsymbol{V}}^{-1}_{g(\tilde{\iota})}} \left\|\tilde{\boldsymbol{\theta}}_{\tilde{\iota}} - \hat{\boldsymbol{\theta}}_{g(\tilde{\iota})}\right\|_{\bar{\boldsymbol{V}}_{g(\tilde{\iota})}} \quad \text{(by the Cauchy–Schwarz inequality)} \\ &\leq ||\boldsymbol{x}'_{\tilde{\iota}}||_{\bar{\boldsymbol{V}}^{-1}_{g(\tilde{\iota})}}\beta_N. \quad \text{(by Eq. (15))}\end{aligned}$$

$$\leq \frac{||\boldsymbol{x}'_{\tilde{\iota}}||}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\tilde{\iota})})}}\beta_N \quad \text{(by definition of eigenvalues)}$$
$$\leq \frac{L}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\tilde{\iota})})}}\beta_N. \quad \text{(by Eq. (13))} \tag{47}$$

Moreover, the estimation error of candidate ι is bounded as

$$\begin{aligned}
\boldsymbol{x}'_{\iota}(\tilde{\boldsymbol{\theta}}_{\iota} - \boldsymbol{\theta}_{g(\iota)}) &\leq ||\boldsymbol{x}'_{\iota}||_{\bar{\boldsymbol{V}}^{-1}_{g(\iota)}} \left\|\tilde{\boldsymbol{\theta}}_{\iota} - \boldsymbol{\theta}_{g(\iota)}\right\|_{\bar{\boldsymbol{V}}_{g(\iota)}} \quad \text{(by the Cauchy–Schwarz inequality)} \\
&\leq 2||\boldsymbol{x}'_{\iota}||_{\bar{\boldsymbol{V}}^{-1}_{g(\iota)}}\beta_N. \quad \text{(by Eq. (17))} \\
&\leq \frac{2||\boldsymbol{x}'_{\iota}||}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\iota)})}}\beta_N \quad \text{(by definition of eigenvalues)} \\
&\leq \frac{2L}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\iota)})}}\beta_N. \quad \text{(by Eq. (13))}
\end{aligned} \tag{48}$$

Based on the above bounds, the regret is bounded as follows.

$$\begin{aligned}
\mathrm{Reg}(N) &= \sum_{n=1}^{N} \mathrm{reg}(n) \\
&\leq \sum_{n=1}^{N} \left(\frac{2}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\iota)})}} + \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(\tilde{\iota})})}} \right) L\beta_N \\
&\quad \text{(by Eq.(46), (47), (48))} \\
&\leq 2L\beta_N \sum_{i\in[K]} \sum_{n=1}^{N} \mathbf{1}[\iota = i] \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} \\
&\quad + L\beta_N \sum_{i\in[K]} \sum_{n=1}^{N} \mathbf{1}[\tilde{\iota} = i] \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}}
\end{aligned} \tag{49}$$

Eq. (49) consisted of two components. The first component is the estimation error of the hired candidate ι. The second component is the estimation error of $\tilde{\iota}$, who would have hired if we had posed the UCB decision rule. The Hybrid decision rule $\tilde{\iota}$ can be different from the UCB decision rule ι, which is the main challenge of deriving regret bound in the hybrid decision rule.

We first define the following events

$$\mathcal{V}_i(n) := \left\{ \boldsymbol{x}_i(n)'(\tilde{\boldsymbol{\theta}}_i(n) - \hat{\boldsymbol{\theta}}_{g(i)}(n)) \leq a\sigma_x \left\|\hat{\boldsymbol{\theta}}(n)\right\| \right\}$$

$$\mathcal{W}_i(n) := \{\tilde{\iota}(n) = i\}$$
$$\mathcal{X}_i(n) := \{\iota(n) = i\}$$
$$\mathcal{X}'_i(n) := \left\{ \boldsymbol{x}_i(n)'\hat{\boldsymbol{\theta}}(n) \geq \arg\max_{j \neq i} \tilde{q}_j^{\mathrm{H}} \right\} \subseteq \mathcal{X}_i.$$

Event $\mathcal{V}_i$ states that the candidate i is not subsidized. Event $\mathcal{W}_i$ states that i would have been hired if it was subsidized in the UCB decision rule. Event $\mathcal{X}_i$ states that i is hired and $\mathcal{X}'_i$ states that i is hired regardless of the subsidy.

The following lemma is the crux of bounding the components in Eq. (49).

Lemma 28 (Proportionality)**.** The following two inequalities hold.

$$\Pr[\mathcal{X}'_i] \geq \exp(-a^2/2) \Pr[\mathcal{W}_i] \tag{50}$$
$$\Pr[\mathcal{X}'_i] \geq \exp(-a^2/2) \Pr[\mathcal{X}_i] \tag{51}$$

Proof of Lemma 28. We first prove, for any $c \in \mathbb{R}$, $d > 0$,

$$\Pr\left[\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)} \geq c\right] \geq \exp(-d^2/2) \Pr\left[\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)} \geq c - d\left(\sigma_x \left\|\hat{\boldsymbol{\theta}}_{g(i)}\right\|\right)^2\right]. \tag{52}$$

Let $x_{\parallel} := (\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)})/||\hat{\boldsymbol{\theta}}_{g(i)}||$ be the projection of $\boldsymbol{x}_i$ into the direction of $\hat{\boldsymbol{\theta}}_{g(i)}$. Then, $\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)} = x_{\parallel}||\hat{\boldsymbol{\theta}}_{g(i)}||$. From the symmetry of a normal distribution, $x_{\parallel}||\hat{\boldsymbol{\theta}}_{g(i)}||$ is drawn from a normal distribution with its variance $(\sigma_x||\hat{\boldsymbol{\theta}}_{g(i)}||)^2$, from which Eq. (52) follows.

Eq. (50) follows by letting $c = \max_{j \neq i} \tilde{q}_j^{\mathrm{H}}$, $d = a$ because

$$\mathcal{W}_i \subseteq \left\{\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)} \geq c - d\left(\sigma_x \left\|\hat{\boldsymbol{\theta}}_{g(i)}\right\|\right)^2\right\}$$
$$\mathcal{X}'_i \supseteq \left\{\boldsymbol{x}'_i\hat{\boldsymbol{\theta}}_{g(i)} \geq c\right\}$$

Eq. (51) also follows by letting $c = \max_{j \neq i} \tilde{q}_j^{\mathrm{H}}$ and $d = a$

$$\mathcal{X}_i \subseteq \left\{\boldsymbol{x}'_i\tilde{\boldsymbol{\theta}}_i \geq c\right\}$$
$$\mathcal{X}'_i \supseteq \left\{\boldsymbol{x}'_i\tilde{\boldsymbol{\theta}}_i \geq c + d\left(\sigma_x \left\|\hat{\boldsymbol{\theta}}_{g(i)}\right\|\right)^2\right\}$$

and exactly the same discussion as Eq. (52) applies for[33]

$$\Pr\left[\boldsymbol{x}_i'\tilde{\boldsymbol{\theta}}_i \geq c + d\left(\sigma_x \left\|\hat{\boldsymbol{\theta}}_{g(i)}\right\|\right)^2\right] \geq \exp(-d^2/2)\Pr\left[\boldsymbol{x}_i'\tilde{\boldsymbol{\theta}}_i \geq c\right]. \tag{53}$$

□

Lemma 28 is intuitively understood as follows. Assume that candidate i would have been hired under the UCB rule. The candidate may not be hired under the Hybrid rule because it can cut subsidy for that candidate. However, there is constant probability such that a slightly better ("$a\sigma$-good") candidate appears on slot i and such a candidate is hired under the Hybrid rule.

The following two lemmas, which utilizes Lemma 28, bounds the two terms of Eq. (49).

Lemma 29.

$$\mathbb{E}\left[\sum_{n=1}^{N}\mathbf{1}[\iota = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}}\right] \leq \frac{2e^{a^2/4}}{\lambda_0}\sqrt{N} + O(1).$$

Lemma 30.

$$\mathbb{E}\left[\sum_{n=1}^{N}\mathbf{1}[\tilde{\iota} = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}}\right] \leq \frac{2e^{a^2/4}}{\lambda_0}\sqrt{N} + O(1).$$

With Lemmas 29 and 30, the regret is bounded as

$$\begin{aligned}
\mathrm{Reg}(N) &\leq 2L\beta_n(L, 1/N)\sum_{i\in[K]}\sum_{n=1}^{N}\mathbf{1}[\iota = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} \\
&\quad + L\beta_n(L, 1/N)\sum_{i\in[K]}\sum_{n=1}^{N}\mathbf{1}[\tilde{\iota} = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} \quad \text{(by Eq. (49))} \\
&\leq 6L\beta_n(L, 1/N)K\frac{e^{a^2/4}\sqrt{N}}{\lambda_0} + \tilde{O}(1) \quad \text{(by Lemma 29 and 30)}
\end{aligned} \tag{54}$$

which completes the proof of Theorem 8. □

Proof of Lemma 29. Let $N_i(n)$ be the number of the rounds before n such that the worker of slot i is selected. Let τ_t be the first round such that $N_i(n)$ reaches t and $N_{i,t} = \sum_{n\leq\tau_t}\mathbf{1}[\mathcal{X}_i'(n)]$. Lemma 28 implies $\mathbb{E}[N_{i,t}] \geq e^{-a^2/2}t$ and applying the Hoeffding inequality on binary random

[33]Note that $\boldsymbol{x}_i'\hat{\boldsymbol{\theta}}_{g(i)}$ in Eq. (52) is replaced by $\boldsymbol{x}_i'\tilde{\boldsymbol{\theta}}_i$ in Eq. (53), which does not change the subsequent derivations at all.

variables $(\mathbf{1}[\mathcal{X}_i'(\tau_1)], \mathbf{1}[\mathcal{X}_i'(\tau_2)], \ldots, ..., \mathbf{1}[\mathcal{X}_i'(\tau_t)])$ yields

$$\Pr\left[N_{i,t} < \left(e^{-a^2/2}t - \sqrt{(\log N)t}\right)\right] \leq \frac{2}{N^2}. \tag{55}$$

By using this, we have

$$\begin{aligned}
&\Pr\left[\bigcap_{t=1}^{N}\left\{N_{i,t} < \left(e^{-a^2/2}t - \sqrt{(\log N)t}\right)\right\}\right] \\
&\leq \sum_t \Pr\left[N_{i,t} < \left(e^{-a^2/2}t - \sqrt{(\log N)t}\right)\right] \quad \text{(by union bound)} \\
&\leq \sum_t \frac{2}{N^2} \quad \text{(by Eq. (55))} \\
&\leq \frac{2}{N}.
\end{aligned}$$

In the following, we focus on the case

$$N_{i,t} \geq e^{-a^2/2}t - \sqrt{(\log N)t}, \tag{56}$$

which occurs with probability at least $1 - 2/N$.

Let $\bar{\boldsymbol{V}}_i(n) := \sum_{n' \leq n} \mathbf{1}[\iota = i]\boldsymbol{x}_i\boldsymbol{x}_i' \preceq \bar{\boldsymbol{V}}_{g(i)}(n)$. The context $\boldsymbol{x}_i$ conditioned on event $\mathcal{X}_i'$ satisfies assumptions in Lemma 18 with $\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_i$ and $\hat{b} = \max_{j \neq i} \tilde{q}_j^{\mathrm{H}}$. We have,

$$\begin{aligned}
\sum_{n=1}^{N} \mathbf{1}[\iota = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} &\leq \sum_{n=1}^{N} \mathbf{1}[\iota = i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i)}} \quad (\text{by } \bar{\boldsymbol{V}}_{g(i)} \succeq \bar{\boldsymbol{V}}_i) \\
&\leq \sum_{n=1}^{N}\sum_{t=1}^{N} \mathbf{1}[\iota = i, N_i(n) = t]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i)}} \\
&\quad (\text{by } N_i(N) \leq N) \\
&\leq \sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))}}. \\
&\quad (\text{by } \mathbf{1}[\iota = i, N_i(n) = t] \text{ occurs at most once})
\end{aligned}$$

In other words, lower-bounding $\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))$ suffices the regret bound, which we demonstrate in the following.

We have

$$\mathbb{E}\left[\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))\right] \geq \lambda_{\min}(\sum_n \mathbb{E}[\mathbf{1}[\mathcal{X}_i'(n)]\boldsymbol{x}_i\boldsymbol{x}_i'])$$

$$\geq \lambda_0 N_{i,t}. \quad \text{(by Lemma 18)}$$

By using the matrix Azuma inequality (Lemma 21), with probability of at least $1 - 1/N$

$$\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t)) \geq \left(\lambda_0 N_{i,t} - \sqrt{32 N_{i,t} \sigma_A^2 \log(dN)}\right) \tag{57}$$

where $\sigma_A = 2L^2$. By using Eq. (56), (57), we have

$$\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t)) \geq \lambda_0 e^{-a^2/2} t - O(\sqrt{t})$$

and thus

$$\begin{aligned}\sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))}} &\leq \sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_0 e^{-a^2/2} t - O(\sqrt{t})}} \\ &\leq \frac{2e^{a^2/4}}{\lambda_0}\sqrt{N} + O(1).\end{aligned}$$

□

Proof of Lemma 30. Let $N_i^{\mathcal{W}_i}(n) = \sum_{n' \leq n} \mathbf{1}[\mathcal{W}_i]$ and let τ_t be the first round such that $N_i^{\mathcal{W}_i}(n)$ reaches t and $N_{i,t} = \sum_{n \leq \tau_t} \mathbf{1}[\mathcal{X}_i'(n)]$. The following discussions are very similar to the one of Lemma 29, which we write for the completeness. Then, we have

$$\begin{aligned}&\Pr\left[\bigcap_{t=1}^{N} \left\{N_{i,t} < \left(e^{-a^2/2} t - \sqrt{(\log N) t}\right)\right\}\right] \\ &\leq \sum_t \Pr\left[N_{i,t} < \left(e^{-a^2/2} t - \sqrt{(\log N) t}\right)\right] \quad \text{(by union bound)} \\ &\leq \sum_t \frac{2}{N^2} \quad \text{(by Lemma 28 and the Hoeffding inequality)} \\ &\leq \frac{2}{N}.\end{aligned}$$

In the following, we focus on the case

$$N_{i,t} \geq e^{-a^2/2} t - \sqrt{(\log N) t} \tag{58}$$

that occurs with probability at least $1 - 2/N$.

We have,

$$\begin{aligned}\sum_{n=1}^{N} \mathbf{1}[\tilde{\iota}=i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} &\le \sum_{n=1}^{N} \mathbf{1}[\tilde{\iota}=i]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{i})}} \quad (\text{by } \bar{\boldsymbol{V}}_{g(i)} \succeq \bar{\boldsymbol{V}}_i)\\ &\le \sum_{n=1}^{N}\sum_{t=1}^{N} \mathbf{1}[\tilde{\iota}=i, N_i^{\mathcal{W}_i}(n)=t]\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{i})}}\\ &\le \sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))}}.\\ &\qquad (\text{by } \{\tilde{\iota}=i\} \text{ increments } N_i^{\mathcal{W}_i})\end{aligned}$$

The following lower-bounds $\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))$.

We have

$$\begin{aligned}\mathbb{E}\left[\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))\right] &\ge \lambda_{\min}\left(\sum_n \mathbb{E}[\mathbf{1}\left[\mathcal{X}_i'(n)\right]\boldsymbol{x}_i\boldsymbol{x}_i']\right)\\ &\ge \lambda_0 N_{i,t}. \quad (\text{by Lemma 18})\end{aligned}$$

By using the matrix Azuma inequality (Lemma 21), at least $1-1/N$

$$\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t)) \ge \left(\lambda_0 N_{i,t} - \sqrt{32 N_{i,t}\sigma_A^2}\log(dN)\right) \tag{59}$$

where $\sigma_A = 2(L(1/N))^2$. By using Eq. (58), (59), we have

$$\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t)) \ge \lambda_0 e^{-a^2/2}t - O(\sqrt{t})$$

and thus

$$\begin{aligned}\sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_i(\tau_t))}} &\le \sum_{t=1}^{N} \frac{1}{\sqrt{\lambda_0 e^{-a^2/2}t - O(\sqrt{t})}}\\ &\le \frac{2e^{a^2/4}}{\lambda_0}\sqrt{N} + O(1).\end{aligned}$$

□

Proof of Theorem 9. We here bound the amount of the subsidy. Eq. (47), (48) imply

$$\boldsymbol{x}_i'\left(\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{g(i)}\right) \le \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} L\beta_N$$
$$\left|\boldsymbol{x}_i'\hat{\boldsymbol{\theta}}_{g(i)} - \boldsymbol{\theta}_{g(i)}\right| \le 2\frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} L\beta_N$$

and thus the subsidy $s_i^{\text{H-I}}(n) = 0$ for

$$\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)}) \ge \left(\frac{2L\beta_N}{\|\boldsymbol{\theta}\|}\right)^2 \max\left(1, \frac{1}{a^2\sigma_x^2}\right) =: C_s = \tilde{O}(1). \tag{60}$$

Hence, it follows that

$$\begin{aligned}
\mathrm{Sub}(N) &= \sum_n s_\iota^{\text{H-I}}(n) \\
&\le \sum_n \sum_i \mathbf{1}[\mathcal{X}_i] s_\iota^{\text{H-I}}(n) \\
&\le L\beta_N \sum_i \sum_n \mathbf{1}[\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)}) \le C_s] \frac{1}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_{g(i)})}} \quad \text{(by Eq. (60))} \\
&\le L\beta_N \sum_i \sum_t \mathbf{1}[\lambda_0 e^{-a^2/2}t - O(\sqrt{t}) \le C_s] \frac{1}{\sqrt{\lambda_0 e^{-a^2/2}t - O(\sqrt{t})}} \\
&\quad \text{(by the same discussion as Lemma 29)} \\
&\le L\beta_N K \sum_t \mathbf{1}[\lambda_0 e^{-a^2/2}t \le C_s] \frac{1}{\sqrt{\lambda_0 e^{-a^2/2}t}} + \tilde{O}(1) \\
&\le L\beta_N K \frac{2e^{a^4/2}}{\lambda_0} \sqrt{\frac{C_s e^{a^2/2}}{\lambda_0}} + \tilde{O}(1) = \tilde{O}(1).
\end{aligned} \tag{61}$$

□

Note that C_s diverges as $a \to +0$. The bound of Theorem 9 is meaningful for $a > 0$. If $a = 0$, the hybrid mechanism is reduced to the UCB mechanism, and thus Theorem 5 for UCB applies.

B.8 Proof of Theorem 10

We modify the proof of Theorem 3. Accordingly, we use the same notation as the proof of Theorem 3 unless we explicitly mention.

We define

$$\mathcal{Q}''(n) = \left\{ \exists i^A, i^B \text{ s.t. } g(i^A) = g(i^B) = 1, i^A \neq i^B, \text{ and } x_i \hat{\theta}_{1,N_1(n)} \geq \frac{1}{2}\mu_x \theta \text{ for } i = i^A, i^B \right\}.$$

When the event $\mathcal{Q}''(n)$ occur, there are two majority workers whose predicted skill $\hat{q}_i(n)$ is larger than its mean.

Lemma 31.

$$\Pr[\mathcal{Q}''(n)|\mathcal{Q}] \geq 1 - (K_1 + 1)\left(\frac{1}{2}\right)^{K_1}. \tag{62}$$

Event $\mathcal{Q}''(n)$ states that the second order statistics of $\{\hat{q}_i\}_{i:g(i)=1}$ is below mean. Lemma 31 states that this event is exponentially unlikely to K_1. By the symmetry of normal distribution and independence of each characteristic $\boldsymbol{x}_i$, each candidate is likely to be below mean with probability 1/2, and the proof of Lemma 31 directly follows by counting the combinations such that at most one of the worker(s) are above mean.

When we have $\mathcal{P}'(n)$ and $\mathcal{Q}''(n)$ for all n, then for every round n, the top-2 workers in terms of quality $\hat{q}_i(n)$ are from the majority. In this case, the minority worker is not hired regardless of additional signal η_i. Accordingly, this is a sufficient condition for a perpetual underestimation.

Proof of Theorem 10. By using Lemmas 24, 25, 26, and 31, we have (36), (37), (38), and (62). From these equations, the probability of perpetual underestimation is bounded as:

$$\begin{aligned}
&\Pr\left[\bigcup_n \{\iota(n) = 1\}\right] \\
&\geq \Pr\left[\bigcup_n \{\mathcal{P}'(n), \mathcal{Q}''(n)\}, \mathcal{P}, \mathcal{Q}\right] \\
&\geq \Pr\left[\mathcal{P}\right] \Pr\left[\mathcal{Q}\right] \Pr\left[\bigcup_n \{\mathcal{P}'(n), \mathcal{Q}''(n)\} \mid \mathcal{P}, \mathcal{Q}\right] \quad \text{(by the independence of } \mathcal{P} \text{ and } \mathcal{Q}) \\
&\geq C_1 b \times C_2 \times (1 - NC_3 \exp\left(-C_4 b\right)) \times \left(1 - N\left(K_1 + 1\right)\left(\frac{1}{2}\right)^{K_1}\right) \quad \text{(by the union bound)},
\end{aligned}$$

which, by letting $b = O(1/\log(N))$ and $K_1 + \log_2(K_1 + 1) \geq \log_2 N$, is $\tilde{O}\left(1\right)$. □

B.9 Proof of Theorem 11

Proof of Theorem 11. We have

$$\begin{aligned}
|\boldsymbol{x}_i'(\hat{\boldsymbol{\theta}}_g - \boldsymbol{\theta}_g)| &\leq ||\boldsymbol{x}_i||_{\bar{\boldsymbol{V}}_g^{-1}} \left\|\hat{\boldsymbol{\theta}}_g - \boldsymbol{\theta}_g\right\|_{\bar{\boldsymbol{V}}_g} \\
&\leq \frac{L}{\lambda_{\min}(\bar{\boldsymbol{V}}_g)}\beta_n \quad \textbf{(by Eq. (13) and (14))} \\
&\leq \frac{L}{\lambda}\beta_N \quad (\text{by } \bar{\boldsymbol{V}}_g \succeq \lambda \boldsymbol{I}_d) \\
&=: C_5 = \tilde{O}(1). \qquad (63)
\end{aligned}$$

Let i_1 and i_2 be the finalists chosen from group 1 and 2, respectively. Define the following event:

$$\mathcal{J}(n) = \{\eta_{i_1}(n) - \eta_{i_2}(n) > 2C_5\}.$$

Under $\mathcal{J}$, the finalist of group 1 is chosen because Eq. (63) implies that $|\boldsymbol{x}_{i_1}'\hat{\boldsymbol{\theta}}_{g_1} - \boldsymbol{x}_{i_2}'\hat{\boldsymbol{\theta}}_{g_2}| \leq 2C_5$ and thus $\boldsymbol{x}_{i_1}'\hat{\boldsymbol{\theta}}_{g_1} + \eta_{i_1} - \boldsymbol{x}_{i_2}'\hat{\boldsymbol{\theta}}_{g_2} + \eta_{i_2} > 0$. Note that $\eta_{i_1} - \eta_{i_2}$ is drawn from $\mathcal{N}(0, 2\sigma_\eta^2)$. Let $C_6 = \Phi^c(\sqrt{2}C_5/\sigma_\eta)$. Then,

$$\Pr[\mathcal{J}(n)] = C_6. \qquad (64)$$

Let $N_1^{\mathcal{J}} = \sum_{n'=1}^{n-1} \mathbf{1}[g(\iota) = 1, \mathcal{J}] \leq N_1(n)$ be the number of hiring of group 1 under event $\mathcal{J}$. By using the Hoeffding inequality, with probability $1 - 1/N^2$ we have

$$N_1^{\mathcal{J}} \geq nC_6 - \sqrt{n\log(N)}. \qquad (65)$$

By taking union bound, Eq. (65) holds for all n with probability $1 - \sum_n 1/N^2 \geq 1 - 1/N$. From now, we evaluate $\lambda_{\min}\left(\bar{\boldsymbol{V}}_1(n)\right)$. It is easy to see that

$$\begin{aligned}
\bar{\boldsymbol{V}}_1 &:= \sum_{n'=1:\iota(n')=g}^{n} \boldsymbol{x}_{i_1}\boldsymbol{x}_{i_1}' + \lambda I \\
&\geq \sum_{n'=1:\iota(n')=g}^{n} \boldsymbol{x}_{i_1}\boldsymbol{x}_{i_1}' \\
&\geq \sum_{n'=1:\mathcal{J}}^{n} \boldsymbol{x}_{i_1}\boldsymbol{x}_{i_1}'.
\end{aligned}$$

In the following, we lower-bound the quantity

$$\lambda_{\min}(\mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i'|\mathcal{J}]) \geq \min_{\boldsymbol{v}:||\boldsymbol{v}||=1} \lambda_{\min}(\mathrm{Var}[\boldsymbol{v}'\boldsymbol{x}_i|\mathcal{J}]).$$

Note that $i_1 = \arg\max_{i:g(i)=1} \boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_1$ is biased towards the direction of $\hat{\boldsymbol{\theta}}_1$, and we cannot use the diversity condition (Lemma 18). Let $\boldsymbol{v}_{\parallel}$ and $\boldsymbol{v}_{\perp}$ be the component of $\boldsymbol{v}$ that is parallel to and perpendicular to $\hat{\boldsymbol{\theta}}_1$.[34] It is easy to confirm that $\mathrm{Var}[\boldsymbol{v}_{\perp}' \boldsymbol{x}_i] = ||\boldsymbol{v}_{\perp}||^2 \sigma_x^2$ because selection of $\arg\max_i \boldsymbol{x}_i' \hat{\boldsymbol{\theta}}_g$ does not yield any bias in perpendicular direction. Regarding $\boldsymbol{v}_{\parallel}$, Lemma 22 characterize the variance, which is slightly[35] smaller than the original variance due to biased selection. Namely,

$$\min_{\boldsymbol{v}:||\boldsymbol{v}||=1} \lambda_{\min}(\mathrm{Var}[\boldsymbol{v}' \boldsymbol{x}_i | \mathcal{J}]) \geq \sigma_x \left(\frac{C_{\mathrm{varmax}}}{\log(K)} ||\boldsymbol{v}_{\parallel}||^2 + ||\boldsymbol{v}_{\perp}||^2 \right) \geq \sigma_x \frac{C_{\mathrm{varmax}}}{\log(K)}.$$

By using the matrix Azuma inequality (Lemma 21) with $\sigma_A = 2L^2$, for $t = \sqrt{32 N_1^{\mathcal{J}} \sigma_A^2} \log(dN)$, with probability $1 - 1/N$

$$\lambda_{\min}(\bar{\boldsymbol{V}}_1) \geq \sigma_x \frac{C_{\mathrm{varmax}}}{\log(K)} N_g^{\mathcal{J}} - t. \tag{66}$$

Combining Eq. (65) and (66), with probability at least $1 - 2/N$, we have

$$\lambda_{\min}(\bar{\boldsymbol{V}}_1(n)) \geq \sigma_x \frac{C_p}{\log(K)} n - \tilde{O}(\sqrt{n}) \tag{67}$$

where $C_p = C_6 C_{\mathrm{varmax}} = \tilde{O}(1)$. By symmetry, exactly the same bound as Eq. (67) holds for group 2. Finally, by using similar transformations as Eq. (24), the regret is bounded as

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}(N)] &\leq 2 \sum_{n=1}^{N} \max_{i \in [K]} \left| \boldsymbol{x}_i'(n) (\hat{\boldsymbol{\theta}}_g - \boldsymbol{\theta}_g) \right| \\
&\leq 2 \sum_{n=1}^{N} \frac{L}{\sqrt{\lambda_{\min}(\bar{\boldsymbol{V}}_g)}} \beta_N \quad \text{(by Eq. (13), (14))} \\
&\leq 2L\beta_N \sum_{n=1}^{N} \sqrt{\frac{\log(K)}{\sigma_x C_p n - \tilde{O}(\sqrt{n})}} \quad \text{(by Eq. (67))} \\
&\leq 4L\beta_N \sqrt{\frac{N \log(K)}{\sigma_x C_p}} + \tilde{O}(1) = \tilde{O}(\sqrt{N})
\end{aligned} \tag{68}$$

which concludes the proof. □

[34] $||\boldsymbol{v}_{\parallel}||^2 + ||\boldsymbol{v}_{\perp}||^2 = 1$.
[35] $O(1/\log K)$

References

Abbasi-Yadkori, Y., D. Pál, and C. Szepesvári (2011): "Improved Algorithms for Linear Stochastic Bandits," in *Advances in Neural Information Processing Systems*, 2312–2320.

Abe, N. and P. M. Long (1999): "Associative Reinforcement Learning Using Linear Probabilistic Concepts," in *Proceedings of the Sixteenth International Conference on Machine Learning*, 3–11.

Aigner, D. J. and G. G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30, 175–187.

Al-Ali, M. N. (2004): "How to Get Yourself on the Door of a Job: A Cross-Cultural Contrastive Study of Arabic and English Job Application Letters," *Journal of Multilingual and Multicultural Development*, 25, 1–23.

Altonji, J. G. and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 313–350.

Arrow, K. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Princeton University Press, 3–33.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002): "Finite-Time Analysis of the Multiarmed Bandit Problem," *Machine Learning*, 47, 235–256.

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Bardhi, A., Y. Guo, and B. Strulovici (2020): "Early-Career Discrimination: Spiraling or Self-Correcting?" Working Paper.

Bastani, H., M. Bayati, and K. Khosravi (2020): "Mostly Exploration-Free Algorithms for Contextual Bandits," *Management Science*.

Battaglini, M., J. M. Harris, and E. Patacchini (2020): "Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary," Working Paper 26726, National Bureau of Economic Research.

Bechavod, Y., K. Ligett, A. Roth, B. Waggoner, and S. Z. Wu (2019): "Equal Opportunity in Online Classification with Partial Feedback," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing*

Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada, ed. by H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, 8972–8982.

Becker, G. S. (1957): *The Economics of Discrimination*, University of Chicago press.

Bergemann, D. and J. Välimäki (2006): "Bandit Problems," Tech. rep., Cowles Foundation for Research in Economics, Yale University.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019a): "Inaccurate Statistical Discrimination," Working Paper.

Bohren, J. A., A. Imas, and M. Rosenberg (2019b): "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 109, 3395–3436.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019): "Beliefs about Gender," *American Economic Review*, 109, 739–73.

Calders, T. and S. Verwer (2010): "Three naive Bayes approaches for discrimination-free classification," *Data Min. Knowl. Discov.*, 21, 277–292.

Che, Y.-K. and J. Hörner (2018): "Recommender Systems as Mechanisms for Social Learning," *Quarterly Journal of Economics*, 133, 871–925.

Che, Y.-K., K. Kim, and W. Zhong (2019): "Statistical Discrimination in Ratings-Guided Markets," Working Paper.

Chen, Y., A. Cuellar, H. Luo, J. Modi, H. Nemlekar, and S. Nikolaidis (2020): "The Fair Contextual Multi-Armed Bandit," in *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020*, ed. by A. E. F. Seghrouchni, G. Sukthankar, B. An, and N. Yorke-Smith, International Foundation for Autonomous Agents and Multiagent Systems, 1810–1812.

Chu, W., L. Li, L. Reyzin, and R. Schapire (2011): "Contextual Bandits with Linear Payoff Functions," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 208–214.

Coate, S. and G. C. Loury (1993): "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 1220–1240.

Cornell, B. and I. Welch (1996): "Culture, Information, and Screening Discrimination," *Journal of Political Economy*, 104, 542–571.

De Paola, M., V. Scoppa, and R. Lombardo (2010): "Can Gender Quotas Break Down Negative Stereotypes? Evidence from Changes in Electoral Rules," *Journal of Public Economics*, 94, 344 – 353.

Ding, J., R. Eldan, and A. Zhai (2015): "On Multiple Peaks and Moderate Deviations for the Supremum of a Gaussian Field," *Annals of Probability*, 43, 3468–3493.

Eddo-Lodge, R. (2017): "Why I'm No Longer Talking to White People About Race," The Gurdian, https://www.theguardian.com/world/2017/may/30/why-im-no-longer-talking-to-white-people-about-race. Accessed on 08/20/2020.

Fang, H. and A. Moro (2011): "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbook of Social Economics*, Elsevier, vol. 1, 133–200.

Farber, H. S. and R. Gibbons (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111, 1007–1047.

Feller, W. (1968): *An Introduction to Probability Theory and Its Applications.*, vol. 1 of *Third edition*, New York: John Wiley & Sons Inc.

Foster, D. and R. Vohra (1992): "An Economic Argument for Affirmative Action," *Rationality and Society*, 4, 176 – 188.

Frazier, P., D. Kempe, J. Kleinberg, and R. Kleinberg (2014): "Incentivizing Exploration," in *Proceedings of the fifteenth ACM conference on Economics and computation*, 5–22.

Fryer, R. and M. O. Jackson (2008): "A Categorical Model of Cognition and Biased Decision Making," *The BE Journal of Theoretical Economics*, 8.

Gittins, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41, 148–177.

Gu, J. and P. Norman (2020): "A Search Model of Statistical Discrimination," Working Paper.

Hanna, R. N. and L. L. Linden (2012): "Discrimination in Grading," *American Economic Journal: Economic Policy*, 4, 146–68.

Hannák, A., C. Wagner, D. Garcia, A. Mislove, M. Strohmaier, and C. Wilson (2017): "Bias in Online Freelance Marketplaces: Evidence from Taskrabbit and Fiverr," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1914–1933.

Hardt, M., E. Price, and N. Srebro (2016): "Equality of Opportunity in Supervised Learning," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, ed. by D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, 3315–3323.

Hilton, J. L. and W. Von Hippel (1996): "Stereotypes," *Annual Review of Psychology*, 47, 237–271.

Immorlica, N., J. Mao, A. Slivkins, and Z. S. Wu (2020): "Incentivizing Exploration with Selective Data Disclosure," in *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, 647–648.

Joseph, M., M. Kearns, J. H. Morgenstern, and A. Roth (2016): "Fairness in Learning: Classic and Contextual Bandits," in *Advances in Neural Information Processing Systems*, 325–333.

Judd, C. M. and B. Park (1993): "Definition and Assessment of Accuracy in Social Stereotypes," *Psychological Review*, 100, 109.

Kannan, S., M. Kearns, J. Morgenstern, M. Pai, A. Roth, R. Vohra, and Z. S. Wu (2017): "Fairness Incentives for Myopic Agents," in *Proceedings of the 2017 ACM Conference on Economics and Computation*, 369–386.

Kannan, S., J. H. Morgenstern, A. Roth, B. Waggoner, and Z. S. Wu (2018): "A Smoothed Analysis of the Greedy Algorithm for the Linear Contextual Bandit Problem," in *Advances in Neural Information Processing Systems*, 2227–2236.

Kaufmann, E. (2014): "Analyse de Stratégies bayésiennes et fréquentistes pour l'allocation séquentielle de ressources," Ph.D. thesis, Institut des sciences et technologies de Paris, thèse de doctorat dirigée par Cappé, Olivier et Garivier, Aurélien Signal et images Paris, ENST 2014.

Kleinberg, J. M. and M. Raghavan (2018): "Selection Problems in the Presence of Implicit Bias," in *9th Innovations in Theoretical Computer Science*, 33:1–33:17.

Kremer, I., Y. Mansour, and M. Perry (2014): "Implementing the 'Wisdom of the Crowd'," *Journal of Political Economy*, 122, 988–1012.

Lai, T. and H. Robbins (1985): "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics*, 6, 4 – 22.

Langford, J. and T. Zhang (2008): "The Epoch-Greedy Algorithm for Contextual Multi-Armed Bandits," in *Advances in Neural Information Processing Systems*, 817–824.

Liu, L. T., S. Dean, E. Rolf, M. Simchowitz, and M. Hardt (2018): "Delayed Impact of Fair Machine Learning," in *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, ed. by J. G. Dy and A. Krause, PMLR, vol. 80 of *Proceedings of Machine Learning Research*, 3156–3164.

Lundberg, S. J. and R. Startz (1983): "Private Discrimination and Social Intervention in Competitive Labor Market," *American Economic Review*, 73, 340–347.

MacNell, L., A. Driscoll, and A. N. Hunt (2015): "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching," *Innovative Higher Education*, 40, 291–303.

Mailath, G. J., L. Samuelson, and A. Shaked (2000): "Endogenous Inequality in Integrated Labor Markets with Two-Sided Search," *American Economic Review*, 90, 46–72.

Mansour, Y., A. Slivkins, and V. Syrgkanis (2020): "Bayesian Incentive-Compatible Bandit Exploration," *Operations Research*, 68, 1132–1161.

Mitchell, K. M. and J. Martin (2018): "Gender Bias in Student Evaluations," *PS: Political Science and Politics*, 51, 648–652.

Monachou, F. G. and I. Ashlagi (2019): "Discrimination in Online Markets: Effects of Social Bias on Learning from Reviews and Policy Design," in *Advances in Neural Information Processing Systems*, 2145–2155.

Moro, A. (2009): *Statistical Discrimination*, London: Palgrave Macmillan UK, 1–5.

Moro, A. and P. Norman (2003): "Affirmative action in a Competitive Economy," *Journal of Public Economics*, 87, 567–594.

——— (2004): "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 114, 1–30.

O'Brien, S. A. (2018): "Facebook Commits to Seeking More Minority Directors," *CNN*, https://money.cnn.com/2018/05/31/technology/facebook-board-diversity/index.html. Accessed on 09/09/2020.

Papanastasiou, Y., K. Bimpikis, and N. Savva (2018): "Crowdsourcing Exploration," *Management Science*, 64, 1727–1746.

Pedreschi, D., S. Ruggieri, and F. Turini (2008): "Discrimination-aware data mining," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, ed. by Y. Li, B. Liu, and S. Sarawagi, ACM, 560–568.

Peña, V. H., T. L. Lai, and Q.-M. Shao (2008): *Self-normalized processes: Limit theory and Statistical Applications*, Springer Science & Business Media.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

Precht, K. (1998): "A Cross-Cultural Comparison of Letters of Recommendation," *English for Specific Purposes*, 17, 241–265.

Raghavan, M., A. Slivkins, J. W. Vaughan, and Z. S. Wu (2018): "The Externalities of Exploration and How Data Diversity Helps Exploitation," in *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, ed. by S. Bubeck, V. Perchet, and P. Rigollet, PMLR, vol. 75 of *Proceedings of Machine Learning Research*, 1724–1738.

Rigollet, P. (2015): "High Dimensional Statistics," MIT OpenCourseWare, https://ocw.mit.edu/courses/mathematics/18-s997-high-dimensional-statistics-spring-2015/lecture-notes/. Accessed on 08/29/2020.

Robbins, H. (1952): "Some Aspects of the Sequential Design of Experiments," *Bulletin of the American Mathematical Society*, 58, 527–535.

Rusmevichientong, P. and J. N. Tsitsiklis (2010): "Linearly Parameterized Bandits," *Mathematics of Operations Research*, 35, 395–411.

Schumann, C., S. N. Counts, J. S. Foster, and J. P. Dickerson (2019a): "The Diverse Cohort Selection Problem," in *Proceedings of the 18th International Conference*

on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019, ed. by E. Elkind, M. Veloso, N. Agmon, and M. E. Taylor, International Foundation for Autonomous Agents and Multiagent Systems, 601–609.

Schumann, C., Z. Lang, J. S. Foster, and J. P. Dickerson (2019b): "Making the Cut: A Bandit-based Approach to Tiered Interviewing," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, ed. by H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, 4641–4651.

Schwartzstein, J. (2014): "Selective Attention and Learning," *Journal of the European Economic Association*, 12, 1423–1452.

Smith, L. and P. Sørensen (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398.

Sundaram, R. K. (2005): "Generalized Bandit Problems," in *Social Choice and Strategic Decisions*, Springer, 131–162.

Thompson, W. R. (1933): "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples," *Biometrika*, 25, 285–294.

Trix, F. and C. Psenka (2003): "Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty," *Discourse and Society*, 14, 191–220.

Tropp, J. A. (2012): "User-Friendly Tail Bounds for Sums of Random Matrices," *Foundations of Computational Mathematics*, 12, 389–434.

Wang, J., Y. Zhang, C. Posse, and A. Bhasin (2013): "Is It Time for a Career Switch?" in *Proceedings of the 22nd International Conference on World Wide Web*, New York, NY, USA: Association for Computing Machinery, WWW '13, 1377–1388.

Williams, W. M. and S. J. Ceci (2015): "National Hiring Experiments Reveal 2: 1 Faculty Preference for Women on STEM Tenure Track," *Proceedings of the National Academy of Sciences*, 112, 5360–5365.

Xu, L., J. Honda, and M. Sugiyama (2018): "A Fully Adaptive Algorithm for Pure Exploration in Linear Bandits," in *International Conference on Artificial Intelligence and Statistics*, 843–851.