Discriminatory Decisions as an Agency Problem:

Multi-dimensional Screening and Optimal Contracts

wiroy shin†

June, 2022

Abstract

A number of the largest U.S. firms have been involved in labor discrimination despite having policies in place designed to avoid that outcome. This paper diagnoses the phenomenon and proposes contractual and regulatory solutions to ameliorate the situation. Existing research (e.g., Becker (1957), Coate and Loury (1993)) studies situations in which individual persons practice discrimination. In contrast, this paper considers a hierarchical organization in which a manager (the agent) may or may not have a discriminatory taste toward his subordinates, while an owner (the principal) is unbiased, only cares about her profit, and does not observe the productivity of each subordinate. In this environment, I study a direct mechanism and characterize an optimal contract. Additionally, I compare the allocation implemented by the optimal direct mechanism to other allocations, including the first-best (full information) one, and discuss the effectiveness of current regulations (e.g., affirmative action). I find that a regulator (such as the U.S. Equal Employment Opportunity Commission) can improve compliance with non-discriminatory conduct even though that the person on whom the regulation is directly incident (i.e., the principal) is not intrinsically biased. I also show that the regulation can be counterproductive if it attempts to enforce perfect fairness (the first-best allocation) when that allocation is not incentive feasible. Finally, I review the U.S. laws regarding discrimination and analyze statutory and jurisprudential issues regarding the optimal mechanism.

Keywords: multidimensional screening, optimal incentives in hierarchies, mechanism design, theories of discrimination
JEL Classification: D02, D04, D21, D63, D82, D86

◇ I thank the late Edward J. Green for his guidance and support throughout this project. I also gratefully acknowledge the comments of the late James Jordan, Paul Grieco, Balázs Szentes, and participants of the 11th World Congress of Econometric Society, the 26th Stony Brook International Conference on Game Theory, and the 2015 Public Economic Theory Conference, as well as the advice of Andrew Wartluft on U.S. laws regarding discrimination and affirmative action. An earlier version of this paper circulated under the title "Discrimination in Organizations: Optimal Contracts and Regulation." This research was supported by a grant from the Rock Ethics Institute at The Pennsylvania State University, the Crawford Graduate Fellowship in Ethical Inquiry.

† KIET (Korea Institute for Industrial Economics and Trade); wiroypsu@gmail.com

1 Introduction

A number of the largest U.S. firms have been involved in illegal labor discrimination. In 2000, Coca-Cola paid \$192 million to its African-American employees in a racial discrimination settlement, and more recently, in 2013, Bank of America paid \$39 million in a gender discrimination settlement. FedEx and Wachovia were also accused of racial and gender discrimination, respectively, at various points and agreed to very costly settlements.

These lawsuits have two notable features. First, the discriminatory treatment was executed by managers at the lowest level. These managers, who directly oversaw line workers, were accused of giving unfair treatment in promotions and wages. Second, the aforementioned defendant firms are listed among the *Fortune* Top 100 companies. This implies that they are under sophisticated and successful management and that their senior-level management makes highly profit-oriented business decisions. Moreover, the management seems to clearly understand that discrimination toward their workers is not in their firms' best interests, as the firms have explicit non-discrimination policies. Accordingly, the fact that discrimination was practiced by the low-level managers toward their subordinates, even though it was corporate policy to scrupulously avoid discrimination, suggests that the firms did not provide effective incentives for these managers to set aside their personal preferences when making business decisions.[1]

In other words, these acts of discrimination resulted from an *agency problem* between the top management and the low-level management as an information gap was created by delegating labor supervision and related promotion decisions to the low-level managers. Under these circumstances, the following economic question can be studied: When an appointed low-level manager is considered to be likely biased, how can he be controlled by contractual arrangements? This paper answers the question by characterizing the optimal incentive contract using a mechanism design approach: If there is a fair chance that the manager is biased, the optimal mechanism for the organization is that which provides an incentive for the discriminatory manager to promote a minority worker. This mechanism not only reduces discrimination but also increases the profit of the organization as well.

The paper also analyzes the effectiveness of anti-discrimination regulations. To some extent, the existence of administrative agencies such as the Equal Employment Opportunity Commission (EEOC) is controversial.[2] This paper points out that profit-maximizing organizations with possible discriminatory managers will not achieve perfect fairness (the first-best allocation) even if the owners (the top management) are unbi-

[1]Perhaps the firms should also have had mechanisms to prevent prejudiced individuals from being assigned to low-level management positions, but prejudice may be so endemic in the workforce that such appointments cannot be avoided.

[2]In the United States, the EEOC is responsible for enforcing federal laws that make it illegal to discriminate against a job applicant or an employee.

ased. It also shows that a regulator can improve compliance with non-discriminatory conduct; however, such regulation may be counterproductive if it attempts to enforce perfect fairness when the conditions for perfect fairness are not feasible.

The following describes the model's environment in this paper.[3] The manager's personal type on his preference toward his subordinate workers (B and W)—whether he is fair or discriminatory about B—is unknown to the organization's owner who decides the manager's compensation. While the manager perfectly observes the productivity levels of two subordinates, the owner cannot see them. The manager promotes one of the subordinates, and then the owner compensates the manager based on what she observes: the promotion decision by the manager and the output of the promoted worker. By the *revelation principle* (Myerson (1981))[4], I consider a direct mechanism that abstracts the details of any particular contractual setting between the owner and the manager. In this direct mechanism, the manager (the agent) reports all available information to the owner (the principal), and the owner makes a promotion decision based on that information. Moreover, the manager's compensation is determined as a function of the report and the output of the promoted worker to maximize the organization's profit. Consequently, an optimal incentive for the biased manager in the hierarchy can be obtained by solving a direct mechanism problem in which the agent has three-dimensional private information (i.e., the productivity levels of both subordinate workers and the manager's discriminatory type) that is partially revealed (i.e., the output of the promoted worker) after promotion.

In summary, this paper addresses a hierarchical environment in which discrimination against some subordinates impairs an institution's profit, and it explains what an owner can do best to mitigate discriminatory decisions without compromising the institution's profit. As the manager may have a discriminatory preference, taste-based discrimination is assumed. In contrast to prior research, this paper considers a *contractual problem* between a principal and a *labor-related decision-maker agent* who has authority over subordinate workers' promotion (see Subsection 2.1 for details). It also addresses a regulation problem by assuming the existence of another principal (a regulator) outside of the organization.

The structure of this paper is as follows. Section 2 reviews the literature on theories of discrimination and multidimensional screening. Section 3 presents the model with a direct mechanism setting, and Section 4 investigates optimal mechanisms. Section 5 characterizes conditions under which a regulator can improve compliance with non-discriminatory conduct. In Section 6, I discuss implementation of the optimal mech-

[3]As the model investigates an agency problem in which the agent is an informed and biased decision-maker, the model applies to diverse adverse selection problems between the discriminatory decision-maker and his welfare-maximizing principal: e.g., labor hiring, resource allocations to subordinate institutions, favoritism in public procurement.

[4]Given a mechanism and an equilibrium for that mechanism, there exists a direct mechanism that gives the same outcome (e.g., who gets promoted; how much the firm produces and pays the manager).

anism and implications of the paper's model related to other research in discrimination, anti-discrimination policies (including legal issues surrounding affirmative action), and multidimensional adverse selection. Section 7 explores potential extensions of the study and concludes the paper.

2 Related Literature

2.1 Theories of Discrimination

There are two main existing branches of research on theories of discrimination: theories of taste-based discrimination and theories of statistical discrimination.[5] In the seminal paper on theories of discrimination, Becker (1957) defines a *taste for discrimination* such that "if someone has a taste for discrimination, he must act as if he were willing to forfeit income in order to avoid certain transactions." It argues that the distribution of discriminatory taste among employers determines the difference in market wages between workers of different races. Moreover, the subsequent theoretical and empirical research on taste-based discrimination (Stiglitz (1973); Black and Strahan (2001)) suggests a positive correlation between market competition and fairness of market outputs: more competitive markets have less discriminatory outputs in workers' wages.

According to theories of statistical discrimination, a decision-maker's belief about workers' outcome-relevant characteristics is a key yielding discriminatory decisions. Statistical discrimination assumes that workers' skills or productivity levels are unobservable by an employer. Instead, workers' physical attributes are used as a signal of their outcome-relevant features. Phelps (1972) introduces a model of discrimination in which statistical distributions of production-skill variables are different across groups. Arrow (1973) and Coate and Loury (1993) develop such statistical differences endogenously. In their environments, ex-ante identical groups can derive different skill investment choices in an asymmetric equilibrium.

While the existing models of discrimination are suitable for an analysis of small organizations consisting of sole proprietorships with such few workers that hierarchical management is not necessary, they are not satisfactory for guiding policy in large firms. Owners of large firms neither have face-to-face interactions with workers nor receive direct benefits from discrimination: in fact, they do not even make employment decisions on low-level workers. Large firms constitute most employment in the United States,[6]

[5]In addition to those two main branches, there is recent research that takes other approaches: in this research, discriminatory decisions are shown as an optimal outcome in various economic environments even when the decision-makers are unbiased and workers are identical. Winter (2004) suggests an equilibrium in a team project environment in which a principal wants to provide different rewards to team members for the same effort levels. Peski and Szentes (2013) provides a repeated matching environment in which employers do not want to be matched with other racial workers in an equilibrium.

[6]Firms with more than 100 workers each were responsible for 63% of U.S. employment in 2014.

and therefore, to maximize the efficiency of anti-discrimination policies and to properly analyze agency problems relating to discrimination in large organizations, it is important to consider a unique treatment for large firms in which another principal exists above the labor-related decision-makers in the hierarchy. This paper develops a model reflecting those hierarchical aspects and provides answers to contractual and policy questions.

2.2 Multidimensional Screening

This paper contributes to the literature on multidimensional screening (see Rochet and Stole (2003) for a survey), as the screening problem this paper presents has distinct features that have not been investigated before. The screening problem here involves an agent (a manager)'s three-dimensional private information. One aspect of this information (the manager's type on discrimination) plays a different role than the other two (the productivity levels of B and W) in the agent's utility: the productivity information implicitly affects the agent's utility through a payment scheme, and the manager's discriminatory taste endogenously appears on the agent's utility (when B is promoted and the agent is discriminatory). This setup is different from other multidimensional problems such as nonlinear pricing in which the consumer's marginal utility on heterogenous goods directly and exogenously appears on the agent's utility function (Armstrong (1996); Sibley and Srinagesh (1997); Armstrong and Rochet (1999)).

Some multidimensional screening problems show an aggregation of information (Armstrong (1996); Biais et al. (2000)) in which a summary statistic of the original multidimensional information is sufficient for the principal. However, the problem discussed here does not exhibit such an aggregation feature, as each piece of information is crucial to the principal (see Subsection 6.3). Nevertheless, this problem is one of the few multidimensional screening problems to which a simple and tractable mechanism can be a solution.[7]

Another distinguishable feature of the problem in this paper is that one piece of information (the productivity of the promoted worker) is revealed after the principal observes the promoted worker's output and before she decides on the payment for the agent. Such sequential information-revealing characteristics between the two decisions (promotion and payment) links this paper to research on sequential screening (Courty and Li (2000); Krähmer and Strausz (2015)). However, this paper's problem is distinguished from the existing standard sequential screening problems, as the person in this paper who receives the information update is the principal, not the agent. In this sense, this paper has similarities with existing research on mechanism design with partially verifiable information (Green and Laffont (1986); Hart et al. (2017)).

[7]Galperti (2015) also addresses a multidimensional resource allocation problem in which delegation with simple standards (a floor or a gap) is a solution.

3 Direct Mechanism

3.1 Environment

The following timeline captures a promotion procedure in which a direct mechanism (a promotion choice rule and a payment rule) contract exists between the organization's owner (the principal) and the manager (the agent).

Timeline

1. The owner specifies a contract.
2. The manager (but not the owner) observes the productivity of two workers–B and W.
3. The manager reports this productivity information, including information regarding his personal discriminatory preference type on the workers, to the owner.
4. The owner promotes one worker and observes the output (perfectly correlated with the productivity) of the promoted worker. However, she remains ignorant about the worker who was not promoted and the type of the manager.[8]
5. The owner compensates the manager in accordance with the contract.

Assumption 1 (Discrimination coefficient)**.** $\theta \in \Theta = \{0, d\}$ denotes the manager's discrimination coefficient type: he could be either discriminatory ($\theta = d$) or fair ($\theta = 0$), where $\nu : \Theta \to [0, 1]$ is the probability mass function. If the manager is discriminatory, and the identity of the promoted worker is B, the manager earns disutility equivalent to the discrimination coefficient $d > 0$. The scalar d is known to both the manager and the owner.

Note that when $\theta = d$, d is a personal cost of the manager affecting the manager's utility only. It does not have an impact on either the organization's profit or the organization's budget for the manager's compensation.

[8] Note that under the actual (non-direct) mechanism contractual setting, the manager makes the promotion decision without providing reports to the owner: 1. The owner specifies a contract. 2. The manager observes the productivity levels of the two workers . 3. The manager decides whom to promote. 4. The promoted worker produces an output, and such output is revealed to the owner. 5. The owner provides a payment according to the contract.

Assumption 2 (Productivity)**.** Let $I = \{B, W\}$ be a set of subordinates. x_B and x_W denote the productivity levels of B and W , respectively. Each x_i is i.i.d drawn from a set $X_i = [0, \bar{\zeta}] \in \mathbb{R}_+$, where $\bar{\zeta} > d$, with a continuous distribution $F_i(\cdot)$ and a density function $f_i(\cdot)$. $\forall x_i \in X_i$, $f_i(x_i) > 0$. Let $x = (x_B, x_W) \in X$, where $f(x) = f_B(x_B) \times f_W(x_W)$ and a measure $\mu : X \to [0, 1]$. The productivity x_i is perfectly correlated with the worker i's output after i's promotion: i's output level equals i's productivity.

Assumption 3 (Absence of outside options)**.** Neither the owner nor the manager has any outside options. That is, the owner cannot fire the manager[9]nor can the manager refuse to provide reports about θ and x.

3.2 Direct Mechanism $< Q, P >$

Let $t \in \Theta$ be a manager's report on his discrimination coefficient type and $z = (z_B, z_W) \in X$ be a manager's productivity report on subordinates B and W. A *direct mechanism* $< Q, P >$ consists of a choice rule $Q : \Theta \times X \to I$ that appoints one subordinate for promotion, as well as a payment rule $P : \Theta \times X \times X_i \to \mathbb{R}_+$ that is a transfer from the owner to the manager. Q is a function of the manager's report (t, z), and P is a function of (t, z) and the promoted subordinate's true productivity $x_{Q(t,z)}$. Let $\mathbb{Q}$ be a set of all choice rules. Note that the manager originally has three-dimensional private information (t, x_B, x_W), but one of the pieces of information—$x_{Q(t,z)}$—is revealed to the owner after the promoted worker's output is verified. That is, in this direct mechanism setting, partial verification of the agent's private information is enabled by the owner: after she decides whom to promote, she observes the outcome of the promoted worker. Therefore, the second decision, the compensation of the manager, is a function of the verified partial information and the manager's initial report.

Definition 1 (Owner's informational state)**.** Under an allocation rule $Q(t, z)$, *the owner's informational state* is defined as $\xi_Q(t, z; x) = (t, z_B, z_W, x_{Q(t,z)})$.

Given $< Q, P >$, *the owner's profit (or the organization's profit)* is

$$\pi(t, z; x) = x_{Q(t,z)} - P(\xi_Q(t, z; x)).$$

Assuming that the budget for the manager's compensation is limited to the organization's output $x_{Q(t,z)}$, $P(\xi_Q(t, z; x)) \in [0, x_{Q(t,z)}]$. Given $< Q, P >$, if the manager

[9]This assumption is reasonable as the owner cannot confirm the manager's discrimination coefficient type.

reports (t, z), *the manager's utility* is

$$u(t, z; \theta, x) = P(\xi_Q(t, z; x)) - d \cdot \mathbb{1}_{Q(t,z)=B} \cdot \mathbb{1}_{\theta=d}.$$

4 Optimal Mechanism

In this section, the owner's optimization problem is presented and the solutions (optimal mechanisms) are characterized. I first analyze a special case in which the manager is deterministically discriminatory ($\nu(d) = 1$). An exploration of this deterministic case helps to clarify the unobservable case in two ways: (1) It provides the necessary conditions for the optimal mechanism of the unobservable case; and (2) The optimal mechanism of the deterministic case can be a simple alternative mechanism for the owner improving the status quo of the unobservable type case. After considering the deterministic case, I investigate the original incomplete information case ($\nu(d) \in (0, 1)$) in Section 4.2.

4.1 The Manager Is Discriminatory, $\nu(d) = 1$.

Under the assumption $\nu(d) = 1$, the workers' productivity levels are the only information that is relevant to the owner's mechanism design problem. Therefore, the manager does not need to report t. Accordingly, the choice rule, the owner's informational state, and the payment rule are simplified without t: $Q : X \to I$, $\xi_Q : X \times X \to X \times X_i$, and $P : X \times X_i \to \mathbb{R}_+$. The owner's profit and the manager's utility are defined as follows.

$$\begin{aligned} \pi(z; x) &= x_{Q(z)} - P(\xi_Q(z; x)). \\ u(z; x) &= P(\xi_Q(z; x)) - d \cdot \mathbb{1}_{Q(z)=B}. \end{aligned}$$

A mechanism $< Q, P >$ is *incentive-compatible (IC)* if truthful reporting is a weakly dominant strategy for the manager, i.e.

$$P(\xi_Q(x; x)) - d \cdot \mathbb{1}_{Q(x)=B} \geq P(\xi_Q(z; x)) - d \cdot \mathbb{1}_{Q(z)=B} \quad \forall x, z \in X$$

Given an allocation rule Q, in this paper, the organization's welfare is defined as a sum of the owner's profit and the manager's compensation. Note that the discriminatory coefficient d is not included in the organization's welfare, so the welfare level is equivalent to the promoted worker's output.

$$w(z; x) = \pi(z; x) + P(\xi_Q(z; x)) = x_{Q(z)}. \tag{1}$$

Note that when two or more mechanisms are compared to each other, each specific $< Q, P >$ is indicated. (e.g., $\pi(z; x, Q, P)$)

Given $< Q, P >$ and the manager's truthful reporting, the profit of the owner is $\pi(x; x) = x_{Q(x)} - P(\xi_Q(x; x))$. The owner's problem is thus about choosing optimal Q and P to maximize the expected profit, subject to the incentive compatibility constraint. That is,

$$\begin{aligned} &\max_{Q,P} \int_{x \in X} f(x) \cdot \pi(x; x) \, dx \\ \text{s.t.} \quad & u(x; x) \geq u(z; x) \ \ \forall x, z \in X. \end{aligned} \tag{2}$$

4.1.1 Full information benchmark

Suppose that no information gap exists between the manager and the owner. In this case, the owner does not need to pay information rent to the manager, and she can promote whomever has a higher productivity level. Therefore, the first-best allocation maximizing the owner's expected profit without the incentive compatibility constraint can be achieved with a choice rule Q as a function of the true productivity vector x.[10] The following mechanism $< Q^F, P^F >$ achieves the first-best allocation.[11]

$$\begin{aligned} Q^F(x) &= B, && \text{if } x_B > x_W \\ Q^F(x) &= W, && \text{if } x_B < x_W \\ Q^F(x) &= B \text{ or } W, && \text{if } x_B = x_W. \end{aligned} \tag{3}$$

$$P^F(\xi_Q(z; x)) = 0, \qquad \forall \xi_Q(z, x) \in X \times X_i.$$

4.1.2 Optimality conditions

Let $\mathbb{M}^*$ be a set of optimal mechanisms, which are solutions to the owner's optimization problem (2), i.e., profit maximizing mechanisms subject to the incentive compatibility constraint. In this section, I characterize conditions for the optimal mechanism. First, one specific class, an unconditional mechanism, is defined as a candidate of the optimal mechanism.

[10] The first-best allocation is also ex-post efficient in terms of the organization's welfare.

[11] Define $x_{(1)} = \max\{x_B, x_W\}$. Following $< Q^F, P^F >$, the expected profit in the first-best case is $E[\pi(\cdot; Q^F, P^F)] = E(x_{(1)})$. The expected surplus of the organization is defined by a sum of the owner's expected profit and the manager's expected utility: $E(s(\cdot; Q^F, P^F)) = E[\pi(\cdot; Q^F, P^F)] + E[u(\cdot; Q^F, P^F)] = E(x_{(1)}) - d \cdot \mathrm{pr}(x_B > x_W)$. Note that under the first-best allocation, the expected profit is equivalent to the expected welfare because the payment to the manager is always zero under $< Q^F, P^F >$. Moreover, such a payment arrangement causes the expected welfare $E[w(\cdot; Q^F, P^F)]$ to exceed the expected surplus $E[s(\cdot; Q^F, P^F)]$: whenever B is promoted, it results in negative externality d to the manager.

Definition 2 (Unconditional mechanism)**.** A mechanism $< Q^\lambda, P^\lambda >$ is an *unconditional mechanism* if it randomizes B's promotion with probability $\lambda \in [0,1]$ and pays zero to the manager regardless of the manager's reports and the owner's informational state. That is,

$$\begin{aligned} \forall z \text{ and } \forall \xi_Q(z,x), \quad Q^\lambda(z) &= B \quad \text{with probability } \lambda \\ Q^\lambda(z) &= W \quad \text{with probability } 1-\lambda, \text{ and} \\ P^\lambda(\xi_Q(z;x)) &= 0. \end{aligned}$$

Let $\mathbb{Q}^\Lambda = \{Q^\lambda | \lambda \in [0,1]\}$. Note that the unconditional mechanism with $\lambda = 0$ represents *the status quo*, where the manager practices discrimination against worker B. At the status quo, the owner does not provide any incentive to the manager, and the manager always promotes W. Therefore, the situation is equivalent to the unconditional allocation s.t. $\forall z, Q(z) = W$. Since the productivity levels of B and W are ex-ante identical, the expected profit of the owner under any unconditional mechanisms is equal to $E(x_i)$, $\forall i \in \{B, W\}$:

$$E[\pi(\cdot; Q^\lambda, P^\lambda)] = \lambda \cdot E(x_B) + (1-\lambda) \cdot E(x_W) = E(x_B) = E(x_W). \tag{4}$$

In addition, given $< Q^\lambda, P^\lambda >$, the manager's utility is

$$u(z,x) = \begin{cases} -d & \text{with probability } \lambda \\ 0 & \text{with probability } 1-\lambda. \end{cases}$$

Such stochastic features of $u(z,x)$ provide the following lemma.

Lemma 1 (Unconditional mechanism)**.** *All unconditional mechanisms are incentive-compatible.*

Proof. See the appendix.

Recall that the agent's private information is partially verifiable. Next, I propose a special type of untruthful reporting by the manager that the owner can detect. Subsequently, I provide a lemma about punishment levels responding to the manager's untruthful reporting strategies: in order to characterize the optimal mechanism, it is sufficient to narrow down the compensation schemes to only those schemes that punish the special type of untruthful reporting by giving a minimum level of compensation to the manager.

Definition 3 (Detectable lie)**.** An owner's informational state $\xi_Q(z;x) = (z_B, z_W; x_{Q(z)})$

is a *detectable lie* if $x_{Q(z)} \neq z_{Q(z)}$. Let $\Xi_Q^d \subset X \times X_i$ be a set of all detectable lie reports under Q.

Example 1 (Detectable lie)**.** Suppose that the discriminatory manager reports $(z_B = 0.5, z_W = 0.7)$ when the true productivity levels are $(x_B = 0.5, x_W = 0.4)$ and that the owner promotes W . After making this promotion, the owner realizes the output $x_{Q(z)} = x_W = 0.4 \neq 0.7 = z_W$. In this case, the lie is detected.

Example 2 (Undetectable lie)**.** Suppose that the discriminatory manager reports $(z_B = 0.5, z_W = 0.7)$ when the true productivity levels are $(x_B = 0.9, x_W = 0.7)$ and that the owner promotes W . After making this promotion, the owner realizes the output $x_{Q(z)} = x_W = 0.7 = z_W$. In this case, the lie about x_B is not detected.

Lemma 2 (Maximum punishment for detectable lies)**.** *Suppose* $< Q, P > \in \mathbb{M}^*$ *and for some detectable lie* $\xi_Q(z'; x)$*,* $P(\xi_Q(z'; x)) > 0$*. Then* $\exists < Q, P^0 > \in \mathbb{M}^*$ *s.t.*

1. $P^0(\xi_Q(z'; x)) = 0$*, and*
2. $E[\pi(\cdot; Q, P^0)] = E[\pi(\cdot; Q, P)]$.

Proof. See the appendix.

The following lemma shows that if two reports (one true and one false) under a specific choice rule produce identical outcomes (the same promotion choice and the same output of the promoted worker), then the payments to the manager should be equivalent to each other in a set of incentive-compatible mechanisms.

Lemma 3 (IC)**.** *Suppose that* $< Q, P >$ *is incentive-compatible. For* x *and* z *s.t.* $x \neq z \in X$*, if* $Q(x) = Q(z)$ *and* $x_{Q(x)} = z_{Q(z)}$*, then* $P(\xi_Q(x; x)) = P(\xi_Q(z; x))$*.*

Proof. See the appendix.

Definition 4 (Conditional allocation rule)**.** Define $\mathbb{Q}^{co}$ to be a set of all conditional allocation rules, i.e, $\mathbb{Q}^{co} = \mathbb{Q} \backslash \mathbb{Q}^{\Lambda}$. Under $Q \in \mathbb{Q}^{co}$, for some $z, z' \in X$, $Q(z) = B$ and $Q(z') = W$.

The next lemma shows that if an allocation rule allows deviations leading to W's promotion, in order to select B, the owner must compensate the manager at least as much as the discrimination coefficient d.

Lemma 4 (IC)**.** *Recall that* Ξ_Q^d *denotes a set of all detectable lie reports under* Q*. Suppose that* $< Q, P >$ *is incentive-compatible and* $Q \in \mathbb{Q}^{co}$*. Assume, too, that for* $\forall \xi_Q(z;x) \in \Xi_Q^d,\ P(\xi_Q(z;x)) = 0$*. If* $Q(z) = B$ *and* $\xi_Q(z;x) \notin \Xi_Q^d$*, then* $P(\xi_Q(z;x)) \geq d$*.*

Proof. See the appendix.

In the next lemma, the optimal payment scheme for conditional allocation rules is presented. The lemma shows that the output information about the promoted worker is useful only for distinguishing detectable lies. Given $Q \in \mathbb{Q}^{co}$, the optimal payment scheme compensates the agent only when his report is not a detectable lie and when it induces B's promotion. The amount of such compensation does not vary according to the output. It is fixed with d, which is exactly equivalent to the agent's discriminatory coefficient.

Lemma 5 (Profit max)**.** *Given an arbitrary conditional allocation rule* $Q \in \mathbb{Q}^{co}$*, the following payment rule* P *uniquely maximizes the expected profit subject to the incentive compatibility constraint.*
For $\forall z, x \in X$*,*

(P1) If $\xi_Q(z;x) \in \Xi_Q^D$*, then* $P(\xi_Q(z;x)) = 0$*;*

(P2) If $\xi_Q(z;x) \notin \Xi_Q^D$ *and* $Q(z) = W$*, then* $P(\xi_Q(z;x)) = 0$*;*

(P3) If $\xi_Q(z;x) \notin \Xi_Q^D$ *and* $Q(z) = B$*, then* $P(\xi_Q(z;x)) = d$*.*

Proof. See the appendix.

4.1.3 Main results

The following arrangement uniquely achieves profit maximization subject to the owner's limited information on workers' productivity levels.

Theorem 1 (Optimal Mechanism)**.** *If the productivity gap between* B *and* W *exceeds the level of the manager's disutility associated with the discriminatory preference, then the owner will promote* B *and compensate the manager up to the disutility generated by this promotion decision. Otherwise,* W *will be promoted, and no payment will be made to the manager. That is, the following* $< Q^*, P^* >$ *is the unique optimal mechanism that maximizes the expected profit subject to the incentive compatibility constraint:*

$$
\begin{aligned}
Q^*(z) &= B && \text{if } z_B - z_W > d \\
Q^*(z) &= W && \text{if } z_B - z_W < d \\
Q^*(z) &= B \text{ } or \text{ } W && \text{if } z_B - z_W = d.
\end{aligned}
$$

$$
\begin{aligned}
P^*(\xi_Q(z;x)) &= 0 && \text{if } \xi_Q(z;x) \in \Xi_Q^D \\
P^*(\xi_Q(z;x)) &= d && \text{if } \xi_Q(z;x) \notin \Xi_Q^D \text{ and } Q(z) = B \\
P^*(\xi_Q(z;x)) &= 0 && \text{if } \xi_Q(z;x) \notin \Xi_Q^D \text{ and } Q(z) = W.
\end{aligned}
\tag{5}
$$

Proof. See the appendix.

Example 3. Suppose that each $x_i \sim$ Uniform$[0,1]$ and $d = 0.2$. Under the optimal mechanism $< Q^*, P^* >$, if $(x_B, x_W) = (0.7, 0.49)$, B is promoted and the owner pays 0.2 to the manager; if $(x_B, x_W) = (0.8, 0.61)$, W is promoted and the owner does not pay the manager. W with $x_W \in (0.8, 1]$ is always promoted regardless of x_B, and B with $x_B \in [0, 0.2)$ is never promoted. However, comparing B 's promotion probability under the optimal mechanism $< Q^*, P^* >$ with that under the status quo where B is never promoted[12], the optimal mechanism enables B with $x_B \sim [0.2, 1]$ to be promoted with probability $(x_B - 0.2)$.

The optimal mechanism $< Q^*, P^* >$ suggests that in order to maximize the organization's profit, the owner should provide an incentive to the discriminatory manager encouraging the promotion of more qualified B. As the optimal mechanism compensates d when B is promoted, the mechanism makes the discriminatory manager indifferent between promoting B and W. As a result, truthful reporting is rationalized in a direct mechanism setting. After enforcing the optimal mechanism, the expected welfare is

$$
\begin{aligned}
E[w(\cdot; Q^*, P^*)] &= \text{pr}(x_B - d > x_W) \cdot E(x_B | x_B - d > x_W) \\
&\quad + (1 - \text{pr}(x_B - d < x_W)) \cdot E(x_W | x_B - d < x_W).
\end{aligned}
$$

Out of the expected welfare, the owner takes expected profit $E[\pi(\cdot; Q^*, P^*)] = E(y_{(1)})$, where $y_{(1)} = \max\{x_W, x_B - d\}$, and pays information rent to the manager up to $d \cdot \text{pr}(x_B - d > x_W)$. Note that compared to the status quo, the owner's expected profit increases by $E(y_{(1)}) - E(x_W)$ and B's promotion probability increases

[12]Recall that the status quo allocation can be implemented by the unconditional mechanism with $\lambda = 0$, $< Q^{\lambda=0}, P^{\lambda=0} >$: $\forall z$ and $\forall \xi_Q(z;x)$, $Q(z) = W$ and $P(\xi_Q(z;x)) = 0$. Accordingly, in the status quo, the utility of the manager $u(z;x)$ is always zero, and the expected profit of the owner is $E(x_W)$.

by $(\text{pr}(x_B - d > x_W) - 0)$. In contrast, the manager's utility level remains the same regardless of the outcome.

The next corollary presents the differences between the first-best and the second-best allocations.

Corollary 1. *The differences between the full-information efficient allocation and the optimal mechanism allocation in the case of the owner's limited information are as follows.*

1. *Promotion probability of B: Compared to the first-best allocation, there is a decrease in the promotion ratio of subordinate B by* $\text{pr}(x_B > x_W) - \text{pr}(x_B - d > x_W) > 0$.

2. *Profit: Compared to the first-best allocation, the expected profit of the owner decreases by* $E(max\{x_B, x_W\}) - E(max\{x_B - d, x_W\}) > 0$.

Proof. The analysis of the full information benchmark in Section 4.1.1 and the proof of Theorem 1 imply this corollary. $\square$

4.2 Incomplete Information on Discrimination Coefficient, $0 < \nu(d) < 1$.

Now I discuss the original case in which the manager could be either *discriminatory* $\theta = d(> 0)$ or *fair* $\theta = 0$, where $\Theta = \{0, d\}$ and the probability mass function is $\nu : \Theta \to [0, 1]$. Recall that $t \in \Theta$ is the manager's report of his discrimination coefficient type. As such, the allocation rule $Q : \Theta \times X \to I$ is a function of both reports: the manager's discrimination coefficient type and the productivity levels of workers. Concurrently, the owner's informational state is $\xi_Q : \Theta \times X^4 \to \Theta \times X^3$ as $\xi_Q(t, z; x) = (t, z_B, z_W; x_{Q(z)})$, so the payment rule is $P : \Theta \times X^3 \to \mathbb{R}_+$. The owner's profit function and the manager's utility function are as follows.

$$\begin{aligned}
\pi(t, z; x) &= x_{Q(t,z)} - P(\xi_Q(t, z; x)). \\
u(t, z; \theta, x) &= P(\xi_Q(t, z; x)) - d \cdot \mathbb{1}_{Q(t,z)=B} \cdot \mathbb{1}_{\theta=d}.
\end{aligned}$$

Given $u(\cdot)$, the incentive compatibility condition is $\forall \theta, t \in \Theta$ and $\forall x, z \in X$,

$$P(\xi_Q(\theta, x; x)) - d \cdot \mathbb{1}_{Q(\theta,x)=B} \cdot \mathbb{1}_{\theta=d} \geq P(\xi_Q(t, z; x)) - d \cdot \mathbb{1}_{Q(t,z)=B} \cdot \mathbb{1}_{\theta=d}. \tag{6}$$

The following represents the owner's optimization problem subject to the incentive compatibility condition.

$$\begin{aligned} \max_{Q,P} \quad & \sum_{\theta \in \Theta} \int_{x \in X} \nu(\theta) \cdot f(x) \cdot \pi(\theta, x; x) \; dx \\ \text{s.t.} \quad & u(\theta, x; \theta, x) \geq u(t, z; \theta, x) \quad \forall \theta, t \in \Theta \text{ and } \forall x, z \in X. \end{aligned} \tag{7}$$

4.2.1 Incentive-compatible mechanisms

In this section, I introduce three kinds of mechanisms satisfying the incentive compatibility condition: unconditional mechanisms, delegation mechanisms, and projection mechanisms. Recall that $\mathbb{Q}^\lambda$ is a set of unconditional allocation rules, which promotes B with probability λ. To represent the manager's productivity reports on the workers inducing worker i's promotion under arbitrary Q and t, let $\chi_i^t(Q) = \{z \mid Q(t, z) = i\}$.

Definition 5 (Unconditional mechanism when $\nu(d) \in (0, 1)$)**.** A mechanism $< Q^\lambda, P^\lambda >$ is an *unconditional mechanism* if it promotes B with probability $\lambda \in [0, 1]$ and pays zero to the manager regardless of the manager's reports and the owner's informational state. That is,

$$\begin{aligned} \forall t, z \text{ and } \forall \xi_Q(t, z, x), \qquad Q^\lambda(t, z) &= B \quad \text{with probability } \lambda \\ Q^\lambda(t, z) &= W \quad \text{with probability } 1 - \lambda, \text{ and} \\ P^\lambda(\xi_Q(t, z; x)) &= 0. \end{aligned} \tag{8}$$

Given any λ, the owner's expected profit with unconditional mechanisms is

$$E[\pi(\cdot; Q^\lambda, P^\lambda)] = \lambda \cdot E(x_B) + (1 - \lambda) \cdot E(x_W) = E(x_B) = E(x_W). \tag{9}$$

Lemma 6 (Unconditional mechanism)**.** *All unconditional mechanisms are incentive-compatible.*

Proof. See the appendix.

Under the status quo, the manager always promotes W if he is discriminatory. If he is not discriminatory, the manager is assumed to follow the first-best allocation rule where the worker with the highest productivity is selected. The following delegation mechanism depicts the status quo.

Definition 6 (Delegation mechanism)**.** Define the following direct mechanism $< Q^0, P^0 >$ as the delegation mechanism that reflects the status quo:

$$\begin{aligned} Q^0(t,z) &= \begin{cases} B & \text{if } t = 0 \text{ and } z_B \geq z_W \\ W & \text{otherwise.} \end{cases} \\ P^0(\xi_Q(t,z;x)) &= 0, \quad \forall t \in \Theta \text{ and } \forall x, z \in X. \end{aligned} \tag{10}$$

Recall $\nu(d) = \mathrm{pr}(\theta = d)$. Then the expected profit of the delegation mechanism is

$$E[\pi(\cdot; Q^0, P^0)] = \nu(d) \cdot E(x_W) + (1 - \nu(d)) \cdot E(\max\{x_B, x_W\}). \tag{11}$$

Lemma 7 (Delegation mechanism)**.** *The delegation mechanism is incentive-compatible.*

Proof. See the appendix.

Definition 7 (Projection mechanism)**.** For an arbitrary set $X_B^d \subset X$ s.t. $X_B^d \neq \emptyset$ and $X_B^d \neq X$, let $\bar{X}_B = \mathrm{proj}_B(X_B^d) \times X_W \subset X$, the Cartesian product of X_B^d's projection on X_B and X_W. Define a projection mechanism $< Q^c(X_B^d), P^c >$ as follows:

$$\begin{aligned} Q^c(t,z;X_B^d) &= \begin{cases} B & \text{if } t = 0 \text{ and } [z \in \bar{X}_B \text{ or } \forall (z_B, z_W) \in X \backslash \bar{X}_B,\ z_B \geq z_W]; \\ & \text{or } t = d \text{ and } (z_B, z_W) \in X_B^d \\ W & \text{if } t = 0 \text{ and } \forall (z_B, z_W) \in X \backslash \bar{X}_B,\ z_B \leq z_W; \\ & \text{or } t = d \text{ and } (z_B, z_W) \notin X_B^d. \end{cases} \\ P^c(\xi_Q(t,z;x)) &= \begin{cases} d & \text{if } t = d \text{ and } Q(t,z) = B \text{ and } \xi_Q(t,z;x) \notin \Xi_Q^D; \\ & \text{or } t = 0 \text{ and } z \in \bar{X}_B \text{ and } \xi_Q(t,z;x) \notin \Xi_Q^D \\ 0 & \text{if } t = d \text{ and } Q(t,z) = W; \text{ or } t = 0 \text{ and } Q(t,z) = W; \\ & \text{or } \xi_Q(t,z;x) \in \Xi_Q^D. \end{cases} \end{aligned} \tag{12}$$

Given X_B^d, the expected profit of the projection mechanism is

$$
\begin{aligned}
& E[\pi(\cdot; Q^c(X_B^d), P^c)] \\
= \;& \nu(d) \cdot [E(x_B - d|X_B^d) \cdot \mu(X_B^d) + E(x_W|(X_B^d)^c) \cdot \mu((X_B^d)^c)] \\
& +(1-\nu(d)) \cdot [E(x_B - d|\bar{X}_B) \cdot \mu(\bar{X}_B) + E(\max\{x_B, x_W\}|(\bar{X}_B)^c) \cdot \mu(\bar{X}_B)^c)] \\
= \;& \nu(d) \cdot [E(x_B - d|X_B^d) \cdot \mu(X_B^d) + E(x_W|(X_B^d)^c) \cdot \mu((X_B^d)^c)] + (1-\nu(d)) \cdot E(y_2)
\end{aligned} \tag{13}
$$

, where y_2 is

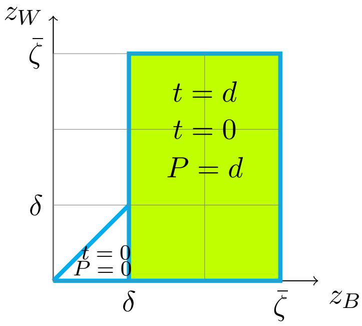
$$
y_2 = \begin{cases} x_B - d & \text{if } x \in \bar{X}_B \\ \max\{x_B, x_W\} & \text{if } x \in X \backslash \bar{X}_B. \end{cases} \tag{14}
$$

Note that in projection mechanisms, the productivity set X_B^d plays an important role: X_B^d not only determines the promotion of B when the manager reports $t = d$, but it also decides the allocation and the payment outcome when $t = 0$. Largely, X_B^d can be divided into three cases: Case 1 (*B-bar projection mechanism*). X_B^d is a correspondence of z_B only, where z_B serves as a cutoff level; Case 2 (*W-bar projection mechanism*). X_B^d is a correspondence of z_W only, where z_W serves as a cutoff level; Case 3 (*Gap projection mechanism*). X_B^d is a correspondence of both z_B and z_W, where $z_B - z_W$ serves as a cutoff level.[13] The following represent these three cases.

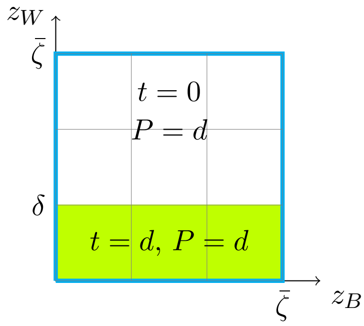
$$
X_B^d(a_B) = [a_B, \bar{\zeta}] \times [0, \bar{\zeta}]; \quad X_B^d(a_W) = [0, \bar{\zeta}] \times [0, a_W]; \quad X_B^d(a_G) = \{z_B, z_W | z_B - z_W > a_G\} \tag{15}
$$

Figure 1 presents the manager's productivity reports leading to B's promotion in the three projection mechanisms. Each green colored area denotes a set of productivity reports inducing B's promotion when the manager's discrimination coefficient type report is d (i.e. $X_B^d(\delta)$), and the bold lined area denotes a set of productivity reports inducing B's promotion when the manager's discrimination type report is 0. For example, suppose that the owner employs the gap projection mechanism with $a_G = \delta$ (see Figure 1 (c)). In this case, B is promoted if the manager reports that he is discriminatory($t = d$) and his productivity report about the workers belong to the green colored triangular area ($X_B^d(a_G = \delta)$). Following the gap projection mechanism's payment rule, a payment value P from the owner is d. When the manager reports that he is fair ($t = 0$), B is promoted if his productivity report about the workers belong to the bold lined triangular area (when $z_B \leq \delta$)or the bold lined rectangular area (when $z_B > \delta$). In this case, a payment value is 0.

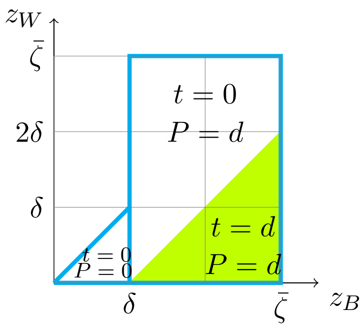
[13] When X_B^d is a correspondence of both z_B and z_W, the gap $z_B - z_W$ is considered for its tractability.



(a) B-bar Projection Mechanism



(b) W-bar Projection Mechanism



(c) Gap Projection Mechanism

Figure 1: Productivity reports leading to B's promotion in the three projection mechanisms, where $a_b = a_w = a_g = \delta$

Lemma 8 (Projection mechanism)**.** *All projection mechanisms are incentive-compatible.*

Proof. See the appendix.

Lemma 9. *Suppose that the optimal mechanism of the deterministic discriminatory coefficient case–$< Q^*, P^* >$ in Theorem 1–is proposed to the manager regardless of his discriminatory type when the owner has incomplete information on θ. Then the outcome will be equivalent to the gap projection mechanism outcome where $X_B^d = \{z | z_B - z_W > d\}$.*

Proof. See the appendix.

The following lemma shows that the optimal mechanism of the deterministic discriminatory type case is better than the status quo of the unobservable discriminatory type case in terms of profit if the probability that the manager is discriminatory is sufficiently high. Recall that $y_1 = \max\{x_B - d, x_W\}$ and $x_1 = \max\{x_W, x_W\}$, and let $\bar{\nu} = \frac{E(x_1) - E(y_2)}{(E(x_1) - E(y_2)) + (E(y_1) - E(x_W))}$.

Lemma 10 (Profit comparison)**.** *If* $\nu(d) \geq \bar{\nu}$*, the gap projection mechanism* $< Q^c(X_B^d), P^c >$ *s.t.* $X_B^d = \{z|z_B - z_W > d\}$ *provides a higher expected profit than the delegation mechanism does.*

Proof. See the appendix.

Lemma 11 (Allocation rule)**.** *Suppose that* $< Q, P >$ *is incentive-compatible. Then* $\chi_B^d(Q) \cap \chi_W^0(Q) = \emptyset$ *and* $\chi_B^d(Q) \subset \chi_B^0(Q)$*.*

Proof. See the appendix.

Lemma 12 (Payment rule)**.** *Suppose that* $< Q, P >$ *is incentive-compatible. If* $x' = (x'_B, x'_W) \in \chi_B^d(Q)$*, then* $\forall \bar{x} \in X$ *s.t.* $\bar{x}_B = x'_B$ *and* $\bar{x}_W \in X_W$*,* $P(\xi_Q(0, \bar{x}, \bar{x})) \geq P(\xi_Q(d, x', x')) = P(d, x', x'_B)$*.*

Proof. See the appendix.

Lemma 13 (Allocation rule)**.** *Let* $\bar{X}_B = \mathrm{proj}_B(\chi_B^d(Q)) \times X_W$*. Suppose that* $< Q, P >$ *is incentive-compatible. If* $\bar{x} \in \bar{X}_B$*, then* $Q(0, \bar{x}) = B$*.*

Proof. See the appendix.

4.2.2 Main results

Let $\mathbb{M}^{**}$ be a set of optimal mechanisms, i.e., the solution to the owner's profit maximization problem subject to the incentive compatibility constraint in (7). The next theorem claims that the optimal mechanism is either one of the projection mechanisms or the delegation mechanism.

Theorem 2 (Optimal mechanism)**.** *The optimal mechanism maximizing the owner's expected profit subject to the incentive compatibility constraint is either one of the projection mechanisms or the delegation mechanism. That is,*

$$\mathbb{M}^{**} \subseteq \cup_{X_B^d \subset X} \{< Q^c(X_B^d), P^c >\} \cup \{< Q^0, P^0 >\}.$$

Proof. See the appendix.

As Theorem 2 suggests, except for the delegation mechanism, the projection mechanisms are the only kind of mechanisms that satisfies the necessary conditions of the optimal mechanism. In the next example, I analyze three projection mechanisms proposed in (15) (B-bar, W-bar, gap) assuming uniform distribution on each worker's productivity. The example advises which mechanism the owner should choose depending on the probability that the manager is discriminatory.

Example 4. Suppose that $\forall i \in \{B, W\}$, $x_i \sim$ Uniform$[0, 1]$, and $d = 0.2$. I consider three different levels of probability that the manager is discriminatory: $\nu(d) \in \{0.1, 0.5, 0.9\}$. For each $\nu(d)$, to determine which mechanism maximizes the expected profit of the owner, we first need to derive an optimal level of a for each projection mechanism: that is, the optimal allocation rule that leads to the promotion of B when $t = d$. Recall that the following three X_B^d describe such rules for B-bar, W-bar, and gap projection mechanisms:

$$X_B^d = [a_B, 1] \times [0, 1]; \quad X_B^d = [0, 1] \times [0, a_W]; \quad X_B^d = \{z_B, z_W | z_B - z_W > a_G\}.$$

Table 1 presents the maximum profit and argmax a_p^*, $p \in \{B, W, G\}$ for each mechanism (see Appendix B.1 for the owner's expected profit functions of the three projection mechanisms). As Table 1 shows, when the probability that the manager is discriminatory is high ($\nu(d) = 0.9$), the gap projection mechanism dominates the other mechanisms with $a_G^* = 0.270$. That is, when the manager is discriminatory and reports $t = d$ (all projection mechanisms are incentive-compatible), the owner promotes B if $x_B - x_W > 0.270$. This optimal cut-off level $a_G^* = 0.270$ is slightly higher than the optimal cut-off level of the special case ($\nu(d) = 1$) in Section 4.1. In the special case, as Theorem 1 suggests, the owner selects $d = 0.2$ as the optimal minimum productivity gap between B and W to promote B . However, under this stochastic case, $\nu(d) = 0.9$, projection mechanisms should also promote B when $t = 0$ and $(z_B, z_W) \in \bar{X}_B = X_B^d \times [0, 1]$ and compensate d to the manager: that is, she needs to pay information rent in the case that the true productivity levels are in the projection set and the manager is a fair type. Such extra information rent increases the optimal lower bound of the productivity gap a_G^* for the gap mechanism, and such a_G^* is higher than the discrimination coefficient $d = 0.2$.

When the probability that the manager is discriminatory is intermediate ($\nu(d) = 0.5$), the B-bar mechanism is optimal for the manager. Note that the chance of the manager being a fair person is 0.5, and accordingly, the information rent for the projection set with $\theta = 0$ is much more costly than that of the gap projection mechanism. Therefore, the expected profit is maximized under the B-bar projection mechanism with a high standard for B , $X_B^d = [a_B^* = 0.904, 1] \times [0, 1]$, where the B-bar projection mechanism provides a smaller projection set compared to that of the gap mechanism.

When the probability that the manager is discriminatory is low ($\nu(d) = 0.1$), the delegation mechanism provides the highest profit. That is, providing no incentive to the manager and delegating all authority are recommended for the owner in this case. Note

$\nu(d) = 0.9$	Delegation	Gap	B-bar	W-bar
Maximum profit	0.516	0.570	0.551	0.520
Argmax		0.270	0.726	0.300
$\nu(d) = 0.5$				
Maximized profit	0.583	0.583	0.585	0.422
Argmax		1.000	0.904	0.300
$\nu(d) = 0.1$				
Maximized profit	0.650	0.650	0.650	0.3245
Argmax		1.000	1.000	0.300

Table 1: Maximum expected profit and argmax a_p^*, $p \in \{B, W, G\}$

that $a^* = 1$ for both the gap and B-bar projection mechanisms; as $a \to 1$, all projection mechanisms converge to the delegation mechanism.

Figure 2 presents the maximum expected profit levels of the four mechanisms for $\forall \nu(d) \in [0, 1]$. Additionally, in Appendix B.2, I provide graphs showing $a^*(\nu(d))$ for each mechanism.

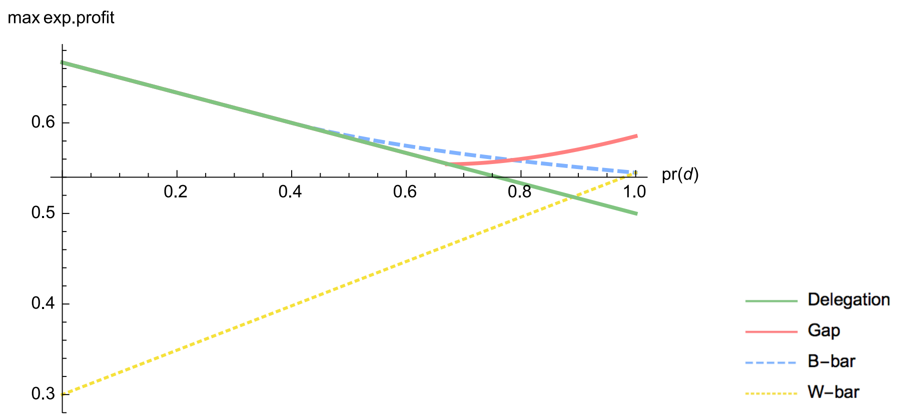


Figure 2: Maximum expected profit levels of the four mechanisms: delegation, gap, B-bar, and W-bar

5 Policy Implementation

As Corollary 1 in Subsection 4.1 and Corollary 3 in Appendix B.3 suggest, there is a gap between the first-best and the second-best allocations. Therefore, even though the owner is not personally biased, regulatory incentives are likely to improve the second-best allocation in terms of fairness. That is, it might be feasible that regulations imposed on the organization increase the probability that a more qualified worker is promoted regardless of his or her demographic category. Such regulation mechanisms do not maximize the owner's expected profit. Therefore, the following allocations would not obtain the first-best expected profit even though the allocation rule is close to the organization's first-best.

Suppose that the organization owns a nonatomic continuum of identical branches. Let m be the representative agent (the manager) of them. All assumptions about the owner, the manager, and the workers remain the same as those in Section 3.1. Suppose that a regulator can observe an aggregate promotion result of the organization: a ratio of B in the promotion. By the law of large numbers, from the allocation rule Q, the owner can perfectly forecast the ratio of B in the promotion. Suppose that the regulator wants such a ratio to be r. If the organization fails to achieve the target ratio, there is a levy τ. In this section, given (r, τ), the owner's problem deriving the optimal mechanism $< Q, P >$ is analyzed first. The regulator's optimization scheme is subsequently discussed.

5.1 Deterministic Discrimination Type Case

In this section, I discuss the special case $\nu(d)$=1, where the manager's discriminatory preference is observable by the owner. The *unobservable case ($\nu(d) \in (0,1)$)* is presented in Appendix B.3, as the results of those two (observable and unobservable) cases have identical implications.

5.1.1 The owner's problem

Given an allocation rule Q, define a set of workers' productivity levels that result in worker i's promotion, $\chi_i(Q) = \{z \mid Q(z) = i\}$. Given (r, τ), the owner's optimization problem in (2) changes as follows, combining the laissez-faire profit $\pi(x; x)$ and the regulatory penalty τ.

$$\begin{aligned} &\max_{Q,P} \int_{x \in X} [f(x) \cdot \pi(x; x)] \, dx - \tau \cdot 1_{(\mu(\chi_B(Q)) \neq r)} \\ \text{s.t. [IC]} \quad & u(x; x) \geq u(z; x) \quad \forall x, z \in X \end{aligned} \tag{16}$$

Definition 8 (Gap mechanism)**.** A mechanism $< Q^\delta, P^\delta >$ is a *gap mechanism* if it promotes B whenever the productivity difference between B and W exceeds δ and if it provides the same compensation as P^* in Theorem 1. That is,

$$
\begin{aligned}
Q^{\delta}(z) &= B \quad \text{if } z_B - z_W > \delta \\
Q^{\delta}(z) &= W \quad \text{if } z_B - z_W < \delta \\
Q^{\delta}(z) &= B \text{ or } W \quad \text{if } z_B - z_W = \delta.
\end{aligned}
$$

$$
\begin{aligned}
P^{\delta}(\xi_Q(z,x)) &= 0 \quad \text{if } Q^{\delta(r)}(z) = W \vee \xi_Q(z;x) \in \Xi_Q^D \\
P^{\delta}(\xi_Q(z,x)) &= d \quad \text{if } Q^{\delta(r)}(z) = B \wedge \xi_Q(z;x) \notin \Xi_Q^D
\end{aligned}
$$

Recall that $\mathbb{Q}^{co}$ is a set of all conditional allocation rules, i.e, $\mathbb{Q}^{co} = \mathbb{Q}\backslash\mathbb{Q}^{\lambda}$. Let $\mathbb{Q}^{co}_r = \{Q \in \mathbb{Q}^{co} \mid \mu(\chi_B(Q)) = r\}$.

Lemma 14 (Optimality of gap mechanism)**.** *Given r, define a cut-off level $\delta(r)$ s.t. $\mu(\{x|x_B - x_W > \delta(r)\}) = r$. Then, $\forall Q \in \mathbb{Q}^{co}_r$, gap mechanism $< Q^{\delta(r)}, P^{\delta(r)} >$ achieves profit maximization of the owner's problem in (16).*

Proof. See the appendix.

Expected profit of $\delta(r)$-gap mechanism is

$$E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] = E(x_W|x_W > x_B - \delta(r))\cdot(1-r) + E(x_B - d|x_W < x_B - \delta(r))\cdot r. \tag{17}$$

Suppose that the second-best allocation by $< Q^*, P^* >$ does not attain the regulatory target ratio r in worker B's promotion. To achieve the ratio r by Lemma 14, the owner considers only two mechanisms: $< Q^{\lambda=r}, P^{\lambda=r} >$ and $< Q^{\delta(r)}, P^{\delta(r)} >$. Otherwise, by Theorem 1, the owner selects the second-best allocation using $< Q^*, P^* >$ and forfeits τ.

Corollary 2 (The owner's optimization problem with policy (r, τ))**.** *Given (r, τ), the owner's optimization problem is as follows.*

$$\max_{Q,P}\{E[\pi(\cdot; Q^*, P^*)] - \tau \cdot 1_{(\mu(\chi_B(Q^*))\neq r)}, \quad E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})], \quad E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})]\} \tag{18}$$

Proof. See the appendix.

5.1.2 The regulator's problem

Theorem 3 (Range of punishment levels)**.** *A regulator can implement a specific target ratio r by setting the punishment levels $\tau(r)$ as follows.*

$$\tau \geq E[\pi(\cdot; Q^*, P^*)] - \max\{E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})], E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})]\}. \tag{19}$$

Proof. See the appendix.

Next, I define unfairness of allocation rules. Among many potential measures, I choose an ordinal measure on dichotomous events. The measure evaluates the frequency of discriminatory incidents: given an allocation rule Q, worker j is promoted even though worker i's productivity is higher than worker j's productivity.

Definition 9 (Unfairness)**.** Given an arbitrary allocation rule Q, the unfairness of the allocation rule Q is defined as follows:

$$\phi(Q) = \mu(x_W > x_B | x \in \chi_B(Q)) \cdot \mu(\chi_B(Q)) + \mu(x_B > x_W | x \in \chi_W(Q)) \cdot \mu(\chi_W(Q)).$$

Lemma 15 (Unfairness of unconditional mechanisms)**.** *Given an arbitrary $r \in (0, 1)$,*

$$\phi(Q^{\delta(r)}) < \phi(Q^{\lambda=r}).$$

Proof. See the appendix.

Suppose that the regulator is also interested in reducing unfairness. In that case, by Lemma 15, given an arbitrary regulatory ratio r, the gap mechanism is preferred to the unconditional mechanism. Note that perfect fairness—$\phi(Q) = 0$—is achieved by the first-best allocation rule described in (3). The following theorem provides conditions for achieving such perfect fairness using the regulation.

Theorem 4. *A regulator can implement the first-best allocation rule if $E(x_B - x_W | x_B > x_W) \geq d$: with (r, τ) s.t. $r = \frac{1}{2}$ and $\tau \geq E[\pi(\cdot; Q^*, P^*)] - E[\pi(\cdot; Q^{\delta(\frac{1}{2})}, P^{\delta(\frac{1}{2})})]$ where $E[\pi(\cdot; Q^{\delta(\frac{1}{2})}, P^{\delta(\frac{1}{2})})] = E(max\{x_B, x_W\}) - \frac{1}{2} \cdot d$.*

Proof. See the appendix.

5.2 Effectiveness of Anti-Discrimination Regulation

Theorem 3 (and Corollary 4 in Appendix B.3) suggests that regulators (e.g., EEOC) can enforce an organization to promote worker B as much as they want if the punishment level is high enough. However, Theorem 4 (and Example 5 in Appendix B.3) implies that such a policy decision needs attention. Without careful examination of the organization, a regulation can induce undesirable negative side effects. For example, if the regulators' goal is too ambitious (e.g., $r = \frac{1}{2}$) or if the manager is extremely discriminatory (d is too high), the owner will choose a less expensive method (i.e., unconditional mechanism, which requires no incentive for the manager) to achieve the regulatory ratio r. In that case, a high frequency of unfair events (discrimination and reverse discrimination) would occur, as shown by Lemma 15 and Lemma 17. Therefore, to succeed in quantitative equity ($r \approx \frac{1}{2}$) and to approximate qualitative fairness (minimizing ϕ), it is necessary for regulators to adjust their objectives based on the details of a specific organization.[14]

The main results presented in this section (Theorem 3 and Theorem 4) can be applied to a one-sided policy where the punishment τ is imposed to an organization with $\mu(\chi_B(Q)) < r$: the proofs are virtually same as the proofs for the two-sided policy. However, the one-sided policy allows the organization to exert reverse discrimination (by promoting B excessively) in the environments in which unconditional mechanisms are optimal for the owner. Therefore, regulators who are concerned about being criticized for reverse discrimination should not assess a one-sided policy as their best option. A one-sided policy simply directs a lower bound of the regulatory ratio, and compared to a two-sided policy, it does not produce any other unique positive effects.

6 Discussion

6.1 Relation to Becker (1957)

The model of this paper assumes that the person who has the discriminatory taste is the manager in the middle of the hierarchy, whereas in Becker (1957), it is assumed that the owner (or the whole organization) has such discriminatory taste.[15] In the sense that both models assume taste-based discrimination, the model shown in this paper can be seen as a modification of Becker's model with new features: (a) existence of another principal above the labor decision-maker in a hierarchical setup, and (b) the principal's incomplete information about the labor decision-maker's discriminatory preference and about the subordinate workers. Recall that in the special case where the manager's discriminatory type is observable by the owner (Subsection 4.1), the owner maximizes her profit by promoting B with $x_B : x_B - x_W > d$ (Q^*) and compensates d to the manager whenever B is promoted (P^*): in this case, the outcome for the owner (the promotion decision and expected profit) from the optimal direct mechanism $< Q^*, P^* >$

[14] i.e., If the conditions in Theorem 4 are not feasible, then set r with the regulatory ratio's upper bound that induces the gap mechanism to be optimal for the owner.

[15] As a result, in Becker (1957), the owner perceives the discrimination coefficient as a part of production costs.

is equivalent to the outcome of Becker's model in which the owner's utility includes the discrimination coefficient d.

However, such outcome equivalence does not hold for the general case s.t. the manager's discrimination type is unobservable. As shown in Subsection 4.2 (Lemma 9 and Example 4), for any given $\nu(d) \in (0,1)$, $< Q^*, P^* >$ is no longer an optimal mechanism. Furthermore, the owner's optimal incentive arrangement[16] is contingent on the probability that the manager is discriminatory. This implies that discrimination in large (multi-hierarchical) organizations is completely different from the discrimination in small organizations described by Becker's model[17] in terms of the treatments to be given.

6.2 Alternative to Gap Projection Mechanism as Affirmative Action

In practice, when an owner chooses an optimal mechanism to adopt, the legal aspect of the mechanism is an important factor. That is, the owner tries to minimize any possibility of criminal prosecution or lawsuits from implementing the mechanism, so as to avoid penalties and future litigation costs. In the Equal Employment Opportunity (EEO) law, such incentives mainly center on civil, not criminal, litigation.[18]

Unfortunately, if interpreted literally, the EEO law (e.g., Title VII of the Civil Rights Act of 1964) makes it virtually impossible for a firm to always make personnel decisions in a completely lawful way. In fact, the analysis in this paper shows that among incentive-compatible direct mechanisms, only the unconditional mechanism can succeed in this respect . However, the unconditional mechanism blatantly disregards the *spirit* of the EEO law, as is shown by the fact that, in equilibrium, it leads to the highest incidence of unfair promotions among incentive-compatible direct mechanisms. This mechanism also harms the firm's profit since it does not optimize the efficiency of promoted workers.

Thus, if a firm's owner justifiably believes that it is highly probable that their manager is discriminatory based on past and current imbalances in promotion results, they will choose the gap projection mechanism over the unconditional mechanism and will accept the risk of making some decisions contrary to the letter of the EEO law. In this situation, the owner will want to implement the equilibrium outcome of the gap projection mechanism in a way that minimizes the litigation risk (e.g., applying an identical promotion choice rule for both workers) and civil penalties that might arise from the gap projection mechanism's well-intentioned conduct, which increases the fairness in an effort to maximize profit. [19]

[16]It could be any one of the three–gap, B-bar, W-bar–projection mechanisms or a delegation mechanism.

[17]In small organizations (two-level hierarchies), the principal directly interacts with subordinate workers and makes labor-related decisions

[18]See Appendix A for statutes about discrimination and affirmative action.

[19]See Appendix A.3 for possible legal issues associated with the gap projection mechanism.

The following alternative indirect mechanism (the alternative gap projection mechanism) generates the same equilibrium outcome of the gap projection mechanism but is a litigation-proof way of providing the same promotion rule for B and W and the same payment rule for both types of manager.[20]

$A1$. The owner asks the manager only about the two workers' productivity levels.

$A2$. The worker with higher reported productivity is promoted (i if $z_i > z_j$). The manager receives d only if the owner observes the promoted B 's productivity (x_B) is higher than δ.

Under the alternative mechanism, the truthful reporting equilibrium outcome of the gap projection mechanism can be obtained using an untruthful reporting equilibrium of the alternative mechanism. The outcome can be checked as follows: If the manager is fair, he reports productivity information truthfully if $x_B < \delta$. If $x_B > \delta$, he always reports $[z_B > z_W$ s.t. $z_B = x_B]$ to earn the bonus d regardless of the true productivity difference $x_B - x_W$. If he is discriminatory, he reports productivity values with $[z_B > z_W$ s.t. $z_B = x_B]$ when the true productivity gap exceeds δ ($x_B - x_W > \delta$), [21] and reports $[z_W > z_B$ s.t. $z_W = x_W]$ if $x_B < \delta$.

Note that these bonus schemes in the mechanism (monetary rewards to the manager for promoting underrepresented groups) have been adopted in some private and public U.S. institutions as part of affirmative action plans.[22][23]

6.3 Identity-based Affirmative Action

To encourage diversity and fairness, affirmative action based on identity markers (race, gender, or ability, or a mixture of these features) has been employed in many cases (e.g., college admissions and employment). There have been debates on the effectiveness of such identity-based affirmative action efforts. The main argument against identity-based affirmative action is that it creates unfair competition for those who are not part of one of the categories recognized by affirmative action: that is, it leads to reverse

[20] Appendix A.3.2. discusses the alternative gap projection mechanism's fulfillment of the Weber test (legal standards for ensuring the legitimacy of affirmative action plans).

[21] The discriminatory manager is indifferent between the two promotion results if $x_B > \delta$. Therefore, any allocation is supported in this case.

[22] *"The program ties executive and senior manager compensation to a 2% net increase in representation of women and minorities at salary grades 10 and above ... All senior managers based in North America will have a portion of their incentive tied to the achievement of the Company's diversity goals."*, Alexis M. Herman et al., Ingram et al. v. The Coca-Cola Co., Third Annual Report of the Task Force, No. 1-98-CV-3679 (RWS) (N.D. Ga. Dec. 1, 2004), pp. 42.

[23] *"Special funds have been established as incentives to increase the number of minorities and women employed at the University. The president's opportunity fund serves as an incentive for the recruitment of minorities and women into areas in which they have historically been underrepresented."*, Section 8, The Pennsylvania State University Affirmative Action Plan 2015, The Pennsylvania State University.

discrimination. However, the results from Sections 4 and 5 show that such an argument is not necessarily true, especially when there is a high chance that the decision-maker is biased. In this case, the profit-maximizing mechanism does not completely mitigate the discriminatory outcome.[24] In fact, by selecting a proper quota level (a level that induces the gap projection mechanism to be optimal), the benefit goes to the competent minorities would have already been promoted if the manager were fair. Therefore, when bias exists, identity-based affirmative action helps capable workers to be chosen regardless of their identities.

Another implication of this paper on affirmative action is the outcome equivalence between quotas and preferential treatments (boosting scores for those that fall into one of the recognized affirmative action categories).[25][26] As shown in Section 5, for each equilibrium outcome with a properly chosen racial quota r that does not induce the unconditional allocation rule to be optimal for the owner, there is a unique productivity leverage preferable for B that creates an outcome equivalent to the racial quota[27]: under the optimal mechanism, the owner achieves the specific targeted quota by considering B's productivity with the extra boosting treatment. In this case, consequently, the preferential treatment policy and the quota policy are actually equivalent in terms of their effects.

6.4 Utilization of Reports on the Agent's Private Information

In the literature of multidimensional screening, some N-dimensional problems can be reduced to lower-dimensional problems (e.g., Armstrong (1996) and Biais et al. (2000)). However, the three-dimensional screening problem analyzed here (incomplete information on the manager's discrimination type and productivity levels of B and W) cannot be simplified to a lower-dimensional problem. That is, summary information is insufficient because each of the reports plays a unique role in the principal's understanding of the given screening problem and derivation of the optimal contract. The productivity levels of B and W , not the difference between them, provide a way to identify the detectable lie, a status in which the final output (the owner's updated informational state $x_{Q(z)}$) is different from what the agent reported ($z_{Q(z)}$). In addition, the payment arrangement contingent on the detectable lie helps to characterize the owner's problem (Lemma 2) and becomes part of the optimal payment scheme (Lemma 5). Therefore,

[24] See Corollary 1 in Subsection 4.1 and Corollary 3 in Appendix B.3.

[25] Bodoh-Creed and Hickman (2018) also shows such equivalence results in a contest environment.

[26] However, the U.S. courts have treated them differently— e.g., in *Grutter v. Bollinger, 539 U.S. 306 (2003)*, the U.S. Supreme Court justices upheld the University of Michigan Law School's affirmative action policy that awarded extra points to blacks, Hispanics, and Native Americans; in *Regents of the University of California v. Bakke, 438 U.S. 265 (1978)*, the U.S. Supreme Court justices ruled that specific quotas by the University of California's Davis School of Medicine were illegal. Otherwise, the quota policy for minorities has been lawfully and actively employed in many European countries.

[27] $|d - \delta(r)|$ in Lemma 14 of Subsection 5.1 and $|a_p^* - a_p(r)|$ in Lemma 16 of Appendix B.3

the exact levels of both workers' productivity are crucial to the principal. Furthermore, as shown in Subsection 4.2, the manager's discrimination type is a main element in constructing the projection mechanism that is the optimal mechanism when the probability that the manager is discriminatory is sufficiently fair (Theorem 2 and Example 4). Hence, without information on the manager's type, the optimal incentive schemes cannot be obtained.

7 Conclusion

Expanding upon the existing research, this paper provides a baseline model for a contractual relationship between a profit-maximizing owner and a discriminatory manager in a hierarchical institution. Discrimination among workers impairs the institution's profit as well as offends standards of equity in broader society. This paper reveals that the owner's optimal mechanism entails providing the discriminatory manager with an incentive to promote a competent minority worker. The optimal mechanism yields benefits in terms of both reducing discrimination and increasing the profit of the organization. This paper also shows how a regulator can improve compliance with non-discriminatory conduct.

The model studied here can be extended by adding choice problems associated with the agent's subordinates (e.g., effort levels); such a model would incorporate statistical discrimination naturally. The subordinates' decisions regarding their effort levels affect their principals' (the manager and the owner) beliefs about the subordinates' productivity: more specifically, an asymmetric equilibrium of the subordinate workers' choice problem can create different statistical distribution on each worker's productivity, and it can lead to statistical discrimination. With the extension, we can study the behavior of subordinates who are exposed to discriminatory treatment in a hierarchical organization: in particular, we can see how minority and non-minority workers strategically approach for a trade-off between their labor cost and promotion opportunity when they acknowledge that the organization's owner is trying to limit the discriminatory manager's discretion. Accordingly, this extension creates another principal-agent relationship between the manager and the subordinate workers that includes the workers' strategies. Therefore, we can formulate other competition environments for the subordinates. For example, we could consider a contest in which the manager wants to design the contest to favor his preferred candidate, and the owner tries to provide him with incentives to design a fair contest.

Finally, it would be meaningful to test the optimal mechanism proposed here by conducting experiments. The optimal payment scheme (that induces the discriminatory agent to be indifferent between his subordinates) and the assumption (such that the agent advocates for the principal's preferred selection when the agent does not have an incentive to deviate from that selection) are well justified theoretically. In reality, however, biased individuals' behavior could be conditional on additional factors such as the nature of the discriminatory preference (e.g., race, gender, or individual favoritism). Experiments that take these factors into consideration would be a good complement to the theoretical study undertaken in this paper.

Appendix A: On the Legal Status of Optimal Mechanisms

This appendix reviews the U.S. laws regarding discrimination and analyzes statutory and jurisprudential issues related to the optimal mechanisms discussed in the paper.[28]

A.1 Statutes regarding Discrimination and Affirmative Action

A.1.1 Core statutes

The Civil Rights Act of 1964 is the main U.S. law prohibiting discrimination in employment opportunities (e.g., hiring, job assignments, promotions, pay and benefits, and discharge) and educational opportunities (e.g., college admission).

1. Employment:

> Title VII of the Civil Rights Act of 1964 (Title VII) makes it unlawful to discriminate against someone on the basis of race, color, national origin, sex or religion. The Act also makes it unlawful to retaliate against a person because the person complained about discrimination, filed a charge of discrimination, or participated in an employment discrimination investigation or lawsuit. [29]

2. Education:

> Specifically, the Educational Opportunities Section is responsible for enforcing Title IV of the Civil Rights Act of 1964, which gives the Attorney General authority to address certain complaints of discrimination alleging denials of equal protection to students based on race, color, national origin, sex, and religion by public schools and institutions of higher learning[30]

[28] This appendix was written based on my understanding of the U.S. laws regarding discrimination and affirmative action, and it is especially tailored to the context of this paper. It should not be used to provide legal advice. Anyone seeking legal advice should consult with legal counsel. All errors are mine.

[29] "Laws Enforced by the Employment Litigation Section," U.S. Department of Justice, accessed April 23, 2021, https://www.justice.gov/crt/laws-enforced-employment-litigation-section.

[30] "Educational Opportunities Section," U.S. Department of Justice, accessed April 23, 2021, https://www.justice.gov/crt/educational-opportunities-section.

A.1.2 Affirmative action and the Weber standard

> Broadly defined, "affirmative action" encompasses any measure that allocates goods–such as admission into selective universities or professional schools, jobs, promotions, public contracts, business loans, and rights to buy, sell, or use land and other natural resources– through a process that takes into account individual membership in designated groups, for the purpose of increasing the proportion of those groups in the relevant labor force, entrepreneurial class, or student population, where they are currently underrepresented as a result of past oppression by state authorities and/or present societal discrimination.[31]

Affirmative action measures can be adopted in three circumstances. First, employers may voluntarily use affirmative action plans to improve a demographic balance in their firms. Second, as a consequence of lawsuits, courts sometimes order affirmative action programs as a remedy for discrimination (e.g., *Sheet Metal Workers International Association v. EEOC*[32]). Finally, contractors of the federal government are required to employ affirmative action for underrepresented minorities and women.[33]

After the seminal case of *United Steelworkers v. Weber*[34], in which the Supreme Court of the United States upheld voluntary affirmative action in private workplaces, the following three conditions from the case became the legal standards for evaluating the legitimacy of affirmative action plans:

Weber Criteria

1. There must be a manifest imbalance in the relevant workforce;
2. The plan cannot unnecessarily trammel the rights of non-beneficiaries;
3. The plan must be temporary, seeking to eradicate traditional patterns of segregation.[35]

A.2 U.S. Legal Cases and Their Implications

The U.S. Supreme Court clearly acknowledges the necessity of affirmative action in certain environments. However, some particular affirmative action plans have been rejected by the court. The court's stance on specific affirmative action plans in particular situations is as follows.

- Layoff or replacement of non-beneficiaries trammel their rights. [36]

[31] Daniel Sabbagh, "Affirmative Action," In The Oxford Handbook of Comparative Constitutional Law, edited by Michel Rosenfeld and András Sajó, Oxford University Press, 2012.

[32] Sheet Metal Workers International Association v. EEOC, 478 U.S. 421 (1986).

[33] "Executive Order 11246", Office of Federal Contract Compliance Programs, U.S. Deparment of Labor, accessed April 23, 2021, `https://www.dol.gov/agencies/ofccp/executive-order-11246/regulations`.

[34] United Steelworkers v. Weber, 443 U.S. 193 (1979).

[35] Lareau (2016).

[36] Taxman v. Piscataway Board of Education, 91 F.3d 1547 (3d Cir. 1996).

- Quotas are generally not allowed, but an exception exists in court-ordered affirmative action.[37]

- Preferential treatment can be used: different cut-off levels are not allowed, but demographic identities can be used as an additional point category. In addition, Banding (e.g., grouping test scores and considering scores in a same group equivalent) might be allowed, but point boosting (e.g., giving extra points to minority candidates) is not allowed.[38]

- Improving diversity can be part of the goals of educational institutions.[39] However, in workplaces, a justification of an operational need for diversity is limited when there is not evidence of past discrimination.[40]

A.3 Implementation of Gap Projection Mechanism

Suppose the following case. The owner of a firm believes that based on past and current imbalances in promotion results, it is highly probable that a given manager is discriminatory. In this case, the gap projection mechanism is the optimal mechanism for the owner to adopt. In this section, I review features of the gap projection mechanism and the alternative gap projection mechanism in the context of the preceding discussion.

A.3.1 Features of the gap projection mechanism

The main features of the gap projection mechanism are as follows:

1. It is a direct mechanism. The owner and the manager communicate regarding the manager's private information: the manager's type and the two workers' (B and W) productivity levels.

2. Payment and promotion rules are dependent on the manager's type.[41]

   (a) If the manager reports that he is *fair*, the owner selects B when B's reported productivity is higher than threshold δ and pays d to the manager. When B's reported productivity is less than threshold δ, a worker with higher reported productivity is promoted, and the manager does not receive any bonus.

   (b) If the manager reports that he is *discriminatory*, the owner selects B when the reported productivity gap between the two workers ($z_B - z_W$) is higher

[37] In the Regents of the University of California v. Bakke, 438 U.S. 265 (1978), the court ruled that quotas used in school admission were not lawful, but it upheld affirmative action in which race is considered one of factors in school admission decisions; in the United States v. Paradise, 480 U.S. 149 (1987), quota-based affirmative action was ordered by the court.

[38] Johnson v. Transportation Agency, 480 U.S. 616, 632 (1987), §78.05 in Lareau (2016).

[39] e.g. Grutter v. Bollinger, 539 U.S. 306 (2003).

[40] See §78.05[4] and §78.12[e] in Lareau (2016).

[41] Note that if the owner identifies a detectable lie, then the payment to the manager is always 0.

than threshold δ and pays d to the manager. Otherwise, W is promoted and the manager does not receive any bonus.

As this description suggests, the gap projection mechanism has affirmative action components and ameliorates the discriminatory outcome of the status quo: it provides a bonus to the manager for promoting B when certain conditions are met, and compared to the status quo, it increases the promotion ratio of B when the manager is discriminatory in its truthful equilibrium outcome.

However, the mechanism can also be seen as having a discriminatory component. This is because the promotion rule specifies that if the manager is discriminatory, it is insufficient for B to have a higher productivity than W to be promoted: in fact, B needs to be better than W at least as much as δ. Therefore, to implement the gap projection mechanism in practice as a form of affirmative action, it is crucial to remove this discriminatory feature.

A.3.2 Weber test

If the owner's belief about the manager's discriminatory type is a function of a past imbalance in promotion results and if the probability that the manager is discriminatory is sufficiently high, then the first Weber criterion for adopting the alternative gap projection mechanism as the optimal mechanism (evidence of discrimination) is met. The second Weber criterion (necessity of trammeling the rights of non-beneficiaries) is also satisfied, as the alternative mechanism simply selects a worker with a higher productivity report, and the mechanism's setup is about promotion, not the replacement of the non-beneficiaries.[42] Finally, the alternative gap projection mechanism meets the third Weber criterion (non-permanent policy), as the mechanism is not selected as the optimal mechanism if institutional discrimination is unobserved.[43]

[42] The original gap projection mechanism also meets the second criterion. The Supreme Court has ruled that in promotion or hiring, considering a minority candidate first if the person is qualified is a justifiable form of affirmative action. Under the gap projection mechanism, W is a non-beneficiary, as the qualified B ($z_B > \delta$) is promoted if the manager is fair. Such a promotion rule can be regarded as containing "banding" and an additional demographic identity point category.

[43] "One factor that helped to make the Kaiser plan a permissible one was the fact that the plan was temporary, to be in effect only until the percentage goal was reached. ... The Supreme Court found that it was unnecessary ... to have an explicit end date or to expressly state that it is temporary.", §78.05 [3] in Lareau (2016).

Appendix B: Proofs and Examples

B.0 Proofs

Proof of Lemma 1

For arbitrary λ, $< Q^{\lambda}, P^{\lambda} >$, the payment rule P^{λ} is independent of the manager's report. Therefore, the agent does not have a deviation incentive, i.e., $\forall \lambda \in [0,1]$, $< Q^{\lambda}, P^{\lambda} >$ is incentive-compatible. $\square$

Proof of Lemma 2

For every $\xi_Q(z;x) \neq \xi_Q(z';x)$, let $P^0(\xi_Q(z;x)) = P(\xi_Q(z;x))$. For $\xi_Q(z';x)$, let $P^0(\xi_Q(z';x)) = 0$. Then the new payment rule P^0 affects the IC condition of the report $\xi_Q(z',x)$. I show that the IC condition still holds as follows:

$$\begin{aligned} P(\xi_Q(x;x)) - d \cdot \mathbb{1}_{Q(x)=B} &\geq P(\xi_Q(z';x)) - d \cdot \mathbb{1}_{Q(z')=B} & , < Q, P > \in \mathbb{M}^* \\ &\geq 0 - d \cdot \mathbb{1}_{Q(z')=B} & , P(\cdot) \in [0, x_{Q(z')}] \\ &= P^0(\xi_Q(z';x)) - d \cdot \mathbb{1}_{Q(z')=B}. \end{aligned}$$

Therefore, $< Q, P^0 >$ is incentive-compatible. For every truthful report $\xi_Q(x,x)$, $P(\xi_Q(x;x)) = P^0(\xi_Q(x;x))$, so $E[\pi(\cdot; Q, P^0)] = E[\pi(\cdot; Q, P)]$. Thus, if $< Q, P > \in \mathbb{M}^*$, then $< Q, P^0 > \in \mathbb{M}^*$. $\square$

Proof of Lemma 3

The following inequalities prove the lemma.

$$\begin{aligned} P(\xi_Q(x;x)) &\geq P(\xi_Q(z;x)) & , \text{IC} \\ &= P(z; x_{Q(x)}) & , Q(x) = Q(z) \\ &= P(z; z_{Q(z)}) & , x_{Q(x)} = z_{Q(z)} \\ &\geq P(x; z_{Q(x)}) & , \text{IC} \\ &= P(\xi_Q(x;x)) & , z_{Q(x)} = z_{Q(z)} = x_{Q(x)}. \quad \square \end{aligned}$$

Proof of Lemma 4

By way of contradiction, assume that $P(\xi_Q(z;x)) < d$ when $Q(z) = B$. Due to the discrimination coefficient of the agent, if $z = x$, $u(z;x) < 0$. Therefore, in this case, the agent wants to deviate to z', where $u(z';x) \geq 0$ as $P(\xi_Q(z',x)) \geq 0$ and $Q(z') = W$, which does not cause the discrimination disutility d. $\square$

Proof of Lemma 5

First, I show that given $Q \in \mathbb{Q}^{co}$, P uniquely maximizes the expected profit by presenting that P provides the agent with the minimum possible payoff satisfying the necessary conditions of IC that are described in the previous lemmas. After that, I show that such $< Q, P >$ is incentive-compatible.

By Lemma 2, (P1) is feasible for $\forall \xi_Q(z;x) \in \Xi_Q^D$. Suppose $\xi_Q(z;x) \notin \Xi_Q^D$. $P(\xi_Q(z;x)) = 0$ when $Q(z) = W$ ensures the minimum transfer from the owner to the manager since the payment rule is bounded below by 0. When $Q(z) = B$, by Lemma 4, $P(\xi_Q(z;x)) = d$ is the minimum possible transfer. Therefore, given $Q \in \mathbb{Q}^{co}$, the payment rule P presented in this lemma maximizes the expected profit.

To check the IC condition, consider an arbitrary conditional allocation rule $Q \in \mathbb{Q}^{co}$. First, I consider a true productivity type vector x where $Q(x) = W$. According to the payment rule P in this lemma, $u(x,x) = 0$. Suppose that the agent reports $z \neq x$. The untruthful report z is different from the truthful productivity x in either or both x_B or x_W, and the arbitrary Q can allocate such z to B or W . By (P1), any deviation to a detectable lie yields weakly smaller utility to the agent than a truthful report. Excluding such detectable lie deviations, there are only two kinds of deviation possibilities that might be profitable to the agent: $z = (x_B, x'_W \neq x_W)$, where $Q(z) = B$ and $z = (x'_B \neq x_B, x_W)$, where $Q(z) = W$. However, under (P2) and (P3), neither of the deviations produces higher utility for the agent, resulting in $u(z,x) = 0$. Therefore, under P, $\forall x$ s.t. $Q(x) = W$, the agent reports truthfully. Another case in which $Q(x) = B$ can be proven in a similar manner. $\square$

Proof of Theorem 1

First, I show that $< Q^*, P^* >$ achieves the maximum profit among incentive-compatible conditional mechanisms (mechanisms with conditional allocation rules). By Lemma 5, given an arbitrary conditional allocation rule $Q \in \mathbb{Q}^{co}$, the payment rule P^* maximizes the expected profit subject to the incentive compatibility constraint. Following such P^*, for some x, if $Q(x) = B$, $\pi(x) = x_B - d$, and if $Q(x) = W$, $\pi(x) = x_W$. Given such P^*, if $x_B - x_W > d$, it is optimal to promote B. Otherwise, promoting W is profitable. Therefore, given P^*, Q^* is an unique optimal conditional allocation rule. Accordingly, the mechanism $< Q^*, P^* >$ is uniquely optimal among conditional mechanisms. Next, I compare this maximum profit to the unconditional mechanism's maximum profit.
Define $y_{(1)} = \max\{x_W, x_B - d\}$, and let $F_{y_{(1)}}(\cdot)$ be a cumulative distribution function of $y_{(1)}$. Expected profit under $< Q^*, P^* >$ is $E(y_{(1)})$. Under unconditional mechanisms $< Q^\lambda, P^\lambda >$, the expected profit is $E(x_i)$. Since $F_{y_{(1)}}$ first-order stochastically dominates F_i, $E(y_{(1)}) > E(x_i)$. That is, the expected profit of $< Q^*, P^* >$ exceeds the expected profit of any constant mechanism. Therefore, $< Q^*, P^* >$ is the optimal solution to the

problem (2).[44] □

Proof of Corollary 1

Section 4.1.1 and the proof of Theorem 1 imply this corollary. □

Proof of Lemma 6

A proof of this lemma is equivalent to the proof of Lemma 1. □

Given an arbitrary allocation rule Q, define a set of workers' productivity levels that results in worker i's promotion when a discrimination coefficient type t is reported: $\chi_i^t(Q) = \{z \mid Q(t,z) = i\}$. Table 2 provides the outcome of the unconditional mechanism when the manager promotes B with probability 0.

	$i = B$	$i = W$
$t = 0$	$(\emptyset, 0)$	$(X, 0)$
$t = d$	$(\emptyset, 0)$	$(X, 0)$

(a) Productivity report set and payment outcome when i is promoted under t, Q^λ, P^λ: $(\chi_i^t(Q^\lambda),\ P^\lambda(\xi_Q(t,z;x)))$

$\theta = 0$	$i = B$	$i = W$
$t = 0$		0
$t = d$		0

(b) The manager's utility $u(t,z;\theta,x)$ where $t \in \{0,d\}$ and $z = x$ when $\theta = 0$ and $x \in \chi_i^t(Q^\lambda)$

$\theta = d$	$i = B$	$i = W$
$t = 0$		0
$t = d$		0

(c) The manager's utility $u(t,z;\theta,x)$ where $t \in \{0,d\}$ and $z = x$ when $\theta = d$ and $x \in \chi_i^t(Q^\lambda)$

* The shaded cells denote unrealized cases as outcomes of the mechanism: $\chi_i^t(Q^\lambda) = \emptyset$.

Table 2: Unconditional mechanism with $\lambda = 0$

Proof of Lemma 7

As Table 3 shows, for every type $(\theta, x) \in \Theta \times X$ of the agent, there is no opportunity for obtaining higher $u(t,z;\theta,x)$ than $u(\theta,x;\theta,x)$. For example, as Table 3(c) illustrates, if the manager with $x \in \chi_W^d = X$ and $\theta = d$ deviates to $t = 0$ and $z \in \chi_B^0$, it would occur utility loss as great as $-d$. □

Proof of Lemma 8

As Table 4 shows, for every type $(\theta, x) \in \Theta \times X$ of the agent, there is no opportunity for obtaining higher $u(t,z;\theta,x)$ than $u(\theta,x;\theta,x)$. For example (see Table 4(b)), the manager with $\theta = 0$ and $x \in (\bar{X}_B)^c$ cannot earn d by deviating to $t = d$ and $x \in X_B^d$: it would be a detectable lie because $(\bar{X}_B)^c = X \backslash \bar{X}_B$ where $\bar{X}_B = \text{proj}_B(X_B^d) \times X_W \subset X$ □

[44] If d is too large and $F_B \neq F_W$, then the unconditional mechanism that selects a subordinate with a higher productivity average can be optimal. However, such an environment is not our focus: this type of environment does not support a delegation situation in a hierarchical organization since the manager plays no role in such an environment. Therefore, in this paper, we focus on the ex-ante identical case.

	$i = B$	$i = W$
$t = 0$	$(\{x : x_B > x_W\}, 0)$	$(\{x : x_B < x_W\}, 0)$
$t = d$	$(\emptyset, 0)$	$(X, 0)$

(a) Productivity report set and payment outcome when i is promoted under t, Q^λ, P^λ: $(\chi_i^t(Q^\lambda)$, $P^\lambda(\xi_Q(t, z; x)))$

$\theta = 0$	$i = B$	$i = W$
$t = 0$	0	0
$t = d$		0

(b) The manager's utility $u(t, z; \theta, x)$ where $t \in \{0, d\}$ and $z = x$ when $\theta = 0$ and $x \in \chi_i^t(Q^0)$

$\theta = d$	$i = B$	$i = W$
$t = 0$	$-d$	0
$t = d$		0

(c) The manager's utility $u(t, z; \theta, x)$ where $t \in \{0, d\}$ and $z = x$ when $\theta = d$ and $x \in \chi_i^t(Q^0)$

* The shaded cells denote unrealized cases as outcomes of the mechanism: $\chi_i^t(Q^\lambda) = \emptyset$.

Table 3: Delegation mechanism

Proof of Lemma 9

Since $< Q^*, P^* >$ is incentive-compatible when $\theta = d$, $X_B^d = \{z | z_B - z_W > d\}$. Under $< Q^*, P^* >$, the allocation is determined soley by (z_B, z_W), and the payment is decided by the identity of the allocation and existence of a detectable lie. Therefore, to earn the payment d, if the agent is not discriminatory with $\theta = 0$, he will lie if $x_B - d < x_W$ and $(x_B, x_W) \in \bar{X}_B$ since the lie is undetectable. By the assumption that the agent follows the first-best allocation rule if he is indifferent between B and W, $Q(0, x) = B$ if $x \in (\bar{X}_B)^c$ and $x_B > x_W$. $\square$

Proof of Lemma 10

The lemma is proved by taking the difference of $E[\pi(\cdot; Q^c(X_B^d), P^c)]$ with $X_B^d = \{z | z_B - d > z_W\}$ in (13) and $E[\pi(\cdot; Q^0, P^0)]$ in (11). $\square$

Proof of Lemma 11

By way of contradiction, suppose that $\exists x' \in \chi_B^d(Q) \cap \chi_W^0(Q)$. Then the following inequality should be satisfied for a type $(0, x')$ not to deviate.

$$P(d, x'; x'_B) \leq P(0, x'; x'_W). \tag{20}$$

	$i=B$		$i=W$
$t=0$	$(\bar{X}_B, d)$	$(\{(\bar{X}_B)^c : x_B > x_W\}, 0)$	$(\{(\bar{X}_B)^c : x_B < x_W\}, 0)$
$t=d$	$(X_B^d, 0)$		$((X_B^d)^c, 0)$

(a) Productivity report set and payment outcome when i is promoted under t, Q^λ, P^λ: $(\chi_i^t(Q^\lambda)$, $P^\lambda(\xi_Q(t, z; x)))$

	$i=B$		$i=W$
$\theta=0$	$\bar{X}_B$	$(\bar{X}_B)^c$	
$t=0$	d	0	0
$t=d$	d		0

(b) The manager's utility $u(t, z; \theta, x)$ where $t \in \{0, d\}$ and $z = x$ when $\theta = 0$ and $x \in \chi_i^t(Q^c(X_B^d))$

	$i=B$		$i=W$
$\theta=d$	$\bar{X}_B$	$(\bar{X}_B)^c$	
$t=0$	0	$-d$	0
$t=d$	0		0

(c) The manager's utility $u(t, z; \theta, x)$ where $t \in \{0, d\}$ and $z = x$ when $\theta = d$ and $x \in \chi_i^t(Q^c(X_B^d))$

Table 4: Projection mechanism with X_B^d

Also, there is another condition for a type (d, x'):

$$P(0, x'; x'_W) \leq P(d, x'; x'_B) - d. \tag{21}$$

Since $d \geq 0$, (20) and (21) contradict each other. Therefore, $\chi_B^d(Q) \cap \chi_W^0(Q) = \emptyset$ and $\chi_B^d(Q) \subset (\chi_W^0(Q))^c = \chi_B^0(Q)$. $\square$

Proof of Lemma 12

By way of contradiction, suppose $P(\xi_Q(0, \bar{x}, \bar{x})) < P(\xi_Q(d, x', x'))$. Then, regardless of $Q(0, \bar{x})$, the agent with a type $(\theta = 0, x = \bar{x})$ can report as $(t = d, z = x')$ without being detected and can earn the higher utility: $P(\xi_Q(d, x', x')) - \theta = P(\xi_Q(d, x', x')) - 0$. This contradicts the assumption that $< Q, P >$ is incentive-compatible. $\square$

Proof of Lemma 13

By way of contradiction, suppose $Q(0, \bar{x}) = W$. Note that $\bar{x} \in \chi_B^d(Q)$ or $\bar{x} \in \chi_W^d(Q)$. By Lemma 12 and Lemma 4,

$$P(0, \bar{x}; \bar{x}_w) \geq P(d, \bar{x}; \bar{x}_B) \geq P(d, \bar{x}; \bar{x}_W) + d.$$

Therefore, if $\bar{x} \in \chi_B^d(Q)$, the agent type $(d, \bar{x})$ can deviate to $(0, \bar{x})$ and obtain the higher utility:

$$P(0, \bar{x}; \bar{x}_w) \geq P(d, \bar{x}; \bar{x}_B) > P(d, \bar{x}; \bar{x}_B) - d.$$

If $\bar{x} \in \chi_W^d(Q)$, the agent type $(d, \bar{x})$ can similarly deviate to $(0, \bar{x})$ and obtain the higher utility:

$$P(0, \bar{x}; \bar{x}_w) \geq P(d, \bar{x}; \bar{x}_W) + d > P(d, \bar{x}; \bar{x}_W).$$

This contradicts the assumption that $< Q, P >$ is incentive-compatible. $\square$

Proof of Theorem 2

In this proof, I consider all types of allocation rules and then derive necessary conditions satisfying incentive compatibility and profit maximization. Specifically, I partition the allocation rules into three cases based on existence and nonexistence of an empty set in $\{\chi_B^t(Q) | t \in \Theta, i \in I\}$. That is, cases s.t. $Q : \chi_B^d(Q) = \emptyset$, $Q : \chi_B^0(Q) = \emptyset$, and a case s.t. $Q : \chi_B^t(Q) \neq \emptyset$, $\forall t \in \Theta$.

1. Q: $\chi_B^d(Q) = \emptyset$.
   Since $\chi_B^d(Q) \cup \chi_W^d(Q) = X$, $\chi_W^d(Q) = X$. That is, if $t = d$, $Q(d, z) = W$, $\forall z \in X$. For $t = 0$, to maximize profit, let Q be the first-best allocation rule: $Q(0, z) = B$ if $z_B > z_W$, and let $Q(0, z) = W$ otherwise. With such Q, by setting the lowest payment in all cases ($\forall t \in \Theta, \forall z, x \in X$, $P(\xi_Q(t, z, x)) = 0$) if the mechanism $< Q, P >$ is incentive-compatible, the owner can achieve profit maximization. Note that this allocation and payment rule exactly match the delegation mechanism $< Q^0, P^0 >$ described in (10), and by Lemma 7 $< Q^0, P^0 >$ is incentive-compatible. Therefore, $< Q^0, P^0 >$ is the optimal mechanism subject to the constraint of Q: $\chi_B^d(Q) = \emptyset$.

2. Q: $\chi_B^0(Q) = \emptyset$.
   Since $\chi_B^0(Q) = \emptyset$, $\chi_W^0(Q) = X$, then, by Lemma 11, $\chi_B^d = \emptyset$. For profit maximization, let $\forall t \in \Theta, \forall z, x \in X$, $P(\xi_Q(t, z, x)) = 0$. This matches the unconditional mechanism $< Q^\lambda, P^\lambda >$ with $\lambda = 0$, and by Lemma 6 $< Q^{\lambda=0}, P^{\lambda=0} >$ is incentive-compatible. Therefore, $< Q^{\lambda=0}, P^{\lambda=0} >$ is the optimal mechanism subject to the constraint of Q: $\chi_B^0(Q) = \emptyset$. However, $< Q^{\lambda=0}, P^{\lambda=0} >$ is dominated by the delegation mechanism $< Q^0, P^0 >$ in terms of expected profit: $E[\pi(\cdot; Q^0, P^0)] - E[\pi(\cdot; Q^{\lambda=0}, P^{\lambda=0})] = (1 - \nu(d)) \cdot [E(\max\{x_B, x_W\}) - E(x_W)] > 0$. Therefore, $< Q^{\lambda=0}, P^{\lambda=0} > \notin \mathbb{M}^{**}$.

3. $Q : \chi_B^t(Q) \neq \emptyset$, $\forall t \in \Theta$.
   If $\chi_B^d(Q) = X$, by Lemma 11, $\chi_B^0(Q) = X$. For profit maximization, let $P(\xi_Q(t, z; x)) = 0$, $\forall t, z, x$. Then such $< Q, P >$ is equivalent to unconditional mechanism $< Q^\lambda, P^\lambda >$ with $\lambda = 1$. However, $E[\pi(\cdot; Q^0, P^0)] - E[\pi(\cdot; Q^{\lambda=1}, P^{\lambda=1})] = (1 - \nu(d)) \cdot [E(\max\{x_B, x_W\}) - E(x_W)] > 0$. Therefore, $< Q^{\lambda=1}, P^{\lambda=1} > \notin \mathbb{M}^{**}$.
   Suppose $\chi_B^d(Q) = X_B^d \subsetneq X$. Then, by Lemma 13, $Q(0, \bar{x}) = B$. For profit maximization and following Lemma 4 and Lemma 12, let $P(\xi_Q(t, z; x)) = d$ if $[t = d$

and $z \in X_B^d$] or [$t = 0$ and $z \in \bar{X}_B$]. Otherwise, let $P(\xi_Q(t, z; x)) = 0$. Also, for profit maximization, let $Q(0, z) = B$ only if $z_B > z_W$. These conditions lead such $< Q, P >$ to be a projection mechanism $< Q^c(X_B^d), P^c >$, and by Lemma 8, $< Q^c(X_B^d), P^c >$ is incentive-compatible. $\square$

Proof of Lemma 14

By Lemma 5, given an arbitrary $Q \in \mathbb{Q}^{co}$, the payment rule $P^{\delta(r)}$ maximizes the expected profit subject to the incentive compatibility constraint. Moreover, because $d \cdot r$ is a constant, $\forall Q \in \mathbb{Q}_r^{co}$, the owner's problem (16) can be simplified as follows.

$$\max_{Q \in \mathbb{Q}_r^{co}} E(x_W \mid Q(x) = W) \cdot (1 - r) + E(x_B \mid Q(x) = B) \cdot r \tag{22}$$

By way of contradiction, suppose that there exists an incentive-compatible mechanism $< Q', P^{\delta(r)} >$ s.t. $Q' \in \mathbb{Q}_r^{co}$ and $E[\pi(\cdot; Q', P')] > E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})]$. Then $\exists\, \eta > 0$ s.t. $\eta = \mu(A_B)$, where $A_B = \{x \mid [Q'(x) = B] \wedge [x_B - x_W < \delta(r)]\}$. Additionally, since $Q' \in \mathbb{Q}_r^{co}$, $\exists$ a set A_W s.t. $A_W = \{x \mid [Q'(x) = W] \wedge [x_B - x_W > \delta(r)\}$.

Take arbitrary subsets $a_B \subset A_B$ and $a_W \subset A_W$ where $\mu(a_B) = \mu(a_W) = \frac{\eta}{2}$, and switch the allocation rule. That is, create another allocation rule Q'' s.t. [$\forall x \in a_B,\ Q''(x) = W$], [$\forall x \in a_W,\ Q''(x) = B$] and [$\forall x \notin a_B \cup a_W,\ Q''(x) = Q(x)$]. Note that in $a_B,\ x_W - x_B > -\delta(r)$, and in $a_W,\ x_B - x_W > \delta(r)$. Therefore, there is an expected profit change Δ from a_B and a_W with Q'': $\Delta_{a_B} > \frac{\eta}{2} \cdot \delta(r)$, $\quad \Delta_{a_W} > \frac{\eta}{2} \cdot (-\delta(r))$. Consequently, $\Delta = \Delta_{a_B} + \Delta_{a_W} > 0$. This contradicts the fact that $Q' \in \mathbb{Q}_r^{co}$ is optimal. Therefore, subject to $Q \in \mathbb{Q}_r^{co}$, $< Q^{\delta(r)}, P^{\delta(r)} >$ is a solution to the owner's problem. $\square$

Proof of Corollary 2

If $\chi_B(Q^*) = r$, by Theorem 1, the optimal mechanism is $< Q^*, P^* >$, and it is supported by (18). Suppose $\chi_B(Q^*) \neq r$. Then the owner should decide either to follow the regulatory rule r or to disobey the rule r and pay the levy τ. If the owner decides to ignore the rule, by Theorem 1, the best strategy for the owner is choosing $< Q^*, P^* >$. If the owner decides to follow the rule, by Lemma 6, she chooses between $< Q^{\lambda=r}, P^{\lambda=r} >$ and $< Q^{\delta(r)}, P^{\delta(r)} >$. $\square$

Proof of Theorem 3

By Corollary 1, the owner chooses to follow the regulation r if $E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})] \geq E[\pi(\cdot; Q^*, P^*)] - \tau \cdot 1_{(\mu(\chi_B(Q^*)) \neq r)}$ or $E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] \geq E[\pi(\cdot; Q^*, P^*)] - \tau \cdot 1_{(\mu(\chi_B(Q^*)) \neq r)}$. $\square$

Proof of Lemma 15

First, I claim $\forall \lambda \in [0,1],\ \phi(Q^\lambda) = \frac{1}{2}$. Recall $Q^\lambda(z) = B$ with probability λ. Therefore,

$$\begin{aligned}
\phi(Q^\lambda) &= \mu(x_W > x_B | x \in \chi_B(Q^\lambda)) \cdot \mu(\chi_B(Q^\lambda)) + \mu(x_B > x_W | x \in \chi_W(Q^\lambda)) \cdot \mu(\chi_W(Q^\lambda)) \\
&= \frac{1}{2} \cdot \lambda + \frac{1}{2} \cdot (1-\lambda) = \frac{1}{2}.
\end{aligned}$$

Now, I claim $\forall r \in (0,1),\ \phi(Q^{\delta(r)}) < \frac{1}{2}$.

1. Suppose $0 < r \leq \frac{1}{2} \Leftrightarrow \delta(r) \geq 0$.

$$\begin{aligned}
\phi(Q^{\delta(r)}) &= \mu(x_W > x_B | x \in \chi_B(Q^{\delta(r)})) \cdot \mu(\chi_B(Q^{\delta(r)}) \\
&\quad + \mu(x_B > x_W | x \in \chi_W(Q^{\delta(r)})) \cdot \mu(\chi_W(Q^{\delta(r)})) \\
&= \mu(x_W > x_B | x \in \chi_B(Q^{\delta(r)})) \cdot r + \mu(x_B > x_W | x \in \chi_W(Q^{\delta(r)})) \cdot (1-r) \\
&= \mathrm{pr}(\delta(r) < x_B - x_W < 0) \cdot r + \mathrm{pr}(0 < x_B - x_W < \delta(r)) \cdot (1-r) \\
&\leq 0 \cdot r \ + \ \frac{1}{2} \cdot (1-r) \\
&< \frac{1}{2}.
\end{aligned}$$

2. Suppose $\frac{1}{2} < r < 1 \Leftrightarrow \delta(r) < 0$.

$$\begin{aligned}
\phi(Q^{\delta(r)}) &= \mathrm{pr}(\delta(r) < x_B - x_W < 0) \cdot r + \mathrm{pr}(0 < x_B - x_W < \delta(r)) \cdot (1-r) \\
&\leq \frac{1}{2} \cdot r + 0 \cdot (1-r) \\
&< \frac{1}{2}.
\end{aligned}$$

Therefore, $\forall r \in (0,1),\ \phi(Q^{\delta(r)}) < \frac{1}{2} = \phi(Q^{\lambda=r})$. $\square$

Proof of Theorem 4

First, I derive a condition in which the owner chooses a gap mechanism over an unconditional mechanism. $E(\pi_U)$ in (4) can be rewritten as follows:

$$E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})] = E(x_W | x_B - x_W < \delta(r)) \cdot (1-r) + E(x_W | x_B - x_W > \delta(r)) \cdot r.$$

Therefore, $E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] \geq E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})] \Leftrightarrow E(x_B - x_W | x_B - x_W > \delta(r)) \geq d$. By definition of $\delta(r)$, $\delta(\frac{1}{2}) = 0$. Then $Q^{\delta=0}$ is equivalent to the first-best allocation rule Q^F defined in (3). Finally, the condition in which the owner follows the regulatory ratio r is given by (19) and (17). $\square$

B.1 Projection Mechanisms (continued)

In this section, I present the expected utility functions of the three (B-bar, W-bar, gap) projection mechanisms discussed in Example 4.

$$
\begin{aligned}
& E[\pi(\cdot; Q^c(X_B^d), P^c)] \\
= \; & \nu(d) \cdot [E(x_B - d | X_B^d) \cdot \mu(X_B^d) + E(x_W | (X_B^d)^c) \cdot \mu((X_B^d)^c)] \\
& +(1-\nu(d)) \cdot [E(x_B - d | \bar{X}_B) \cdot \mu(\bar{X}_B) + E(\max\{x_B, x_W\} | (\bar{X}_B)^c) \cdot \mu(\bar{X}_B)^c)].
\end{aligned} \tag{23}
$$

Note that $\forall i \in \{B, W\},\ x_i \sim$ Uniform$[0,1]$: $F_i(x_i) = x_i$ and $F(x_B, x_W) = F_B(x_B) \cdot F_W(x_W)$.

1. B-bar projection mechanism: For an arbitrary $a \in [0,1]$, $X_B^d = [a,1] \times [0,1]$ and $\bar{X} = [a,1] \times [0,1]$

$$
\begin{aligned}
& E[\pi(\cdot; Q^c([a,1] \times [0,1]), P^c)] \\
= \; & \nu(d) \cdot [(\int_a^1 x_B \cdot \frac{1}{1-a}\ dF_B - d) \cdot (1 - F(a,1)) + \int_0^1 x_W\ dF_W \cdot F(a,1)] \\
+ \; & (1-\nu(d)) \cdot [(\int_a^1 x_B \cdot \frac{1}{1-a}\ dF_B - d) \cdot (1 - F(a,1)) + (\int_0^1 x_1\ dF_1) \cdot F(a,1)] \\
& \text{where } F_1(x_1) = F_B(x_1 | x_B \leq a) \cdot F_W(x_1).
\end{aligned} \tag{24}
$$

2. W-bar projection mechanism: For an arbitrary $a \in [0,1]$, $X_B^d = [0,1] \times [0,a]$ and $\bar{X} = [0,1] \times [0,1]$

$$
\begin{aligned}
& E[\pi(\cdot; Q^c([0,1] \times [0,a]), P^c)] \\
= \; & \nu(d) \cdot [(\int_0^1 x_B\ dF_B - d) \cdot F(1,a) + \int_a^1 x_W \cdot \frac{1}{1-a}\ dF_W \cdot (1 - F(1,a))] \\
+ \; & (1-\nu(d)) \cdot [(\int_0^1 x_B\ dF_B - d) \cdot F(1,1)]
\end{aligned} \tag{25}
$$

3. Gap projection mechanism: For an arbitrary $a \in [0,1]$, $X_B^d = \{x | x_B - x_W > a\}$ and $\bar{X} = [a,1] \times [0,1]$

$$
\begin{aligned}
& E[\pi(\cdot; Q^c(\{x|x_B - x_W > a\}), P^c)] \\
= \ & \nu(d) \cdot [(\int_0^1 \int_{x_W+a}^1 x_B \ dF_B dF_W - d) \cdot (\int_0^1 \int_{x_W+a}^1 1 \ dF_B dF_W) \\
& \quad +(\int_0^1 \int_{x_B-a}^1 x_W \ dF_W dF_B) \cdot (\int_0^1 \int_{x_B-a}^1 1 \ dF_W dF_B)] \qquad (26) \\
+ \ & (1-\nu(d)) \cdot [(\int_a^1 x_B \cdot \frac{1}{1-a} \ dF_B - d) \cdot (1 - F(a,1)) + (\int_0^1 x_1 \ dF_1) \cdot F(a,1)] \\
& \text{where } F_1(x_1) = F_B(x_1|x_B \leq a) \cdot F_W(x_1).
\end{aligned}
$$

B.2 Figures for Example 4

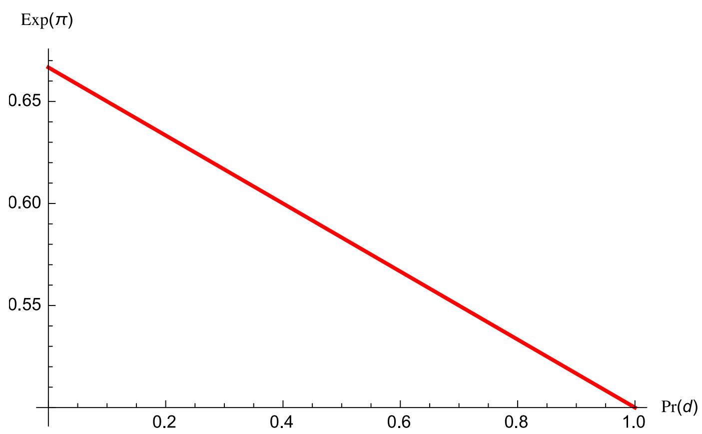
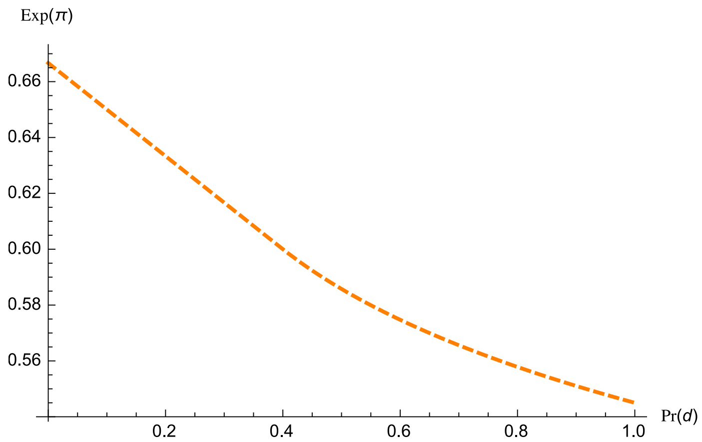
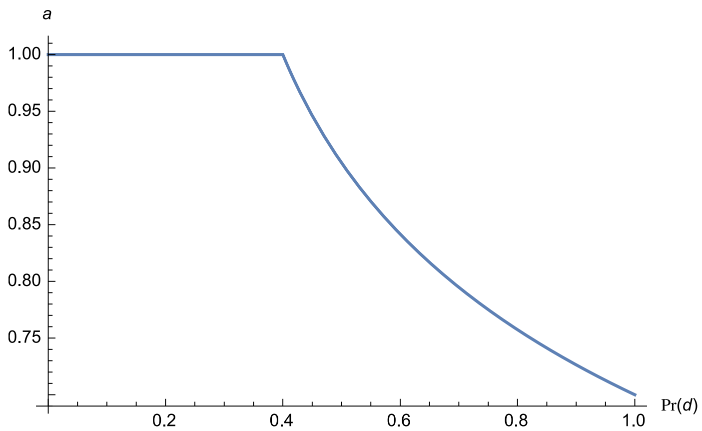


Figure 3: Expected Profit of Delegation Mechanism

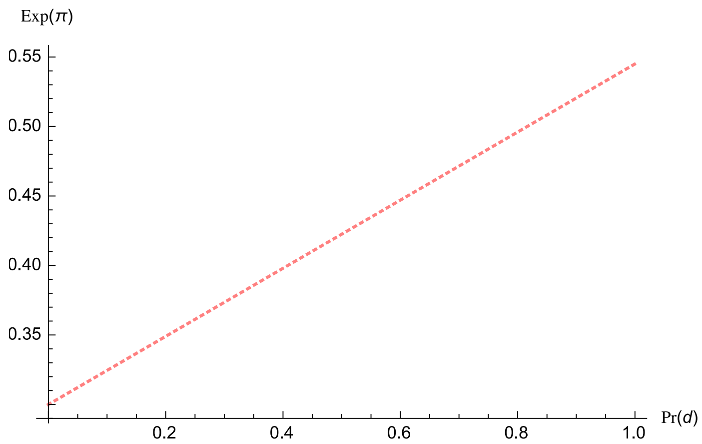


(a) Maximum expected profit

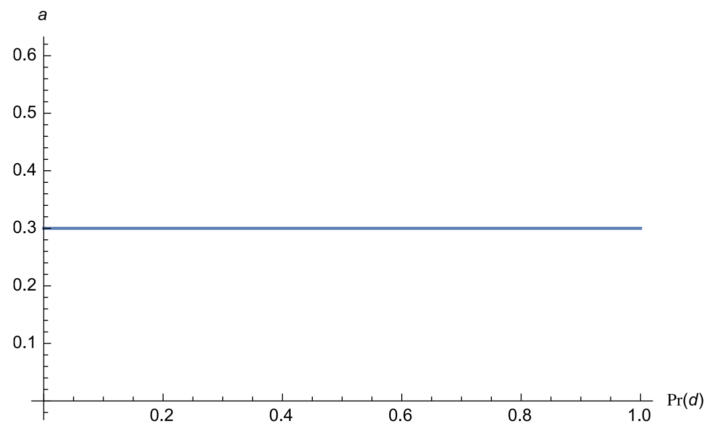


(b) Argmax a_B^*

Figure 4: B-bar projection mechanism: $X_B^d = [a, 1] \times [0, 1]$

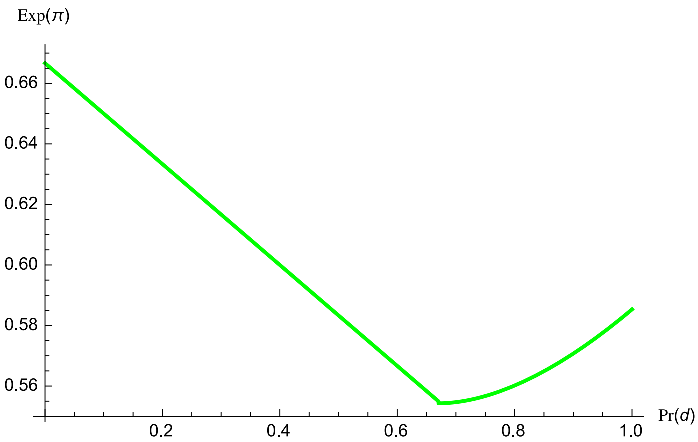


(a) Maximum expected profit

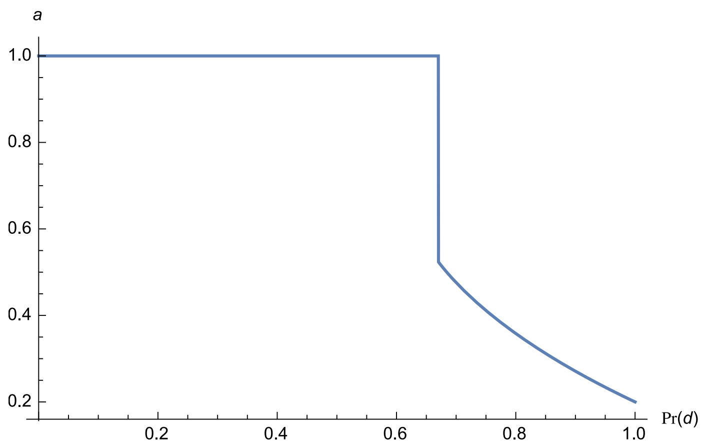


(b) Argmax a^*_W

Figure 5: W-bar projection mechanism: $X^d_B = [0, 1] \times [a, 1]$



(a) Maximum expected profit



(b) Argmax a_G^*

Figure 6: Gap projection mechanism: $X_B^d = \{z | z_B - z_W > a\}$

B.3 Policy Implementation with $\nu(d) \in (0, 1)$

Corollary 3. *If $\nu(d) \in (0, 1)$, the first-best allocation (Q^F) in (3) cannot be achieved by any incentive-compatible mechanisms.*

Proof. By way of contradiction, suppose that there exists an incentive-compatible mechanism $< Q, P >$ s.t. $\forall t \in \{0, d\}$, $\chi_B^t(Q) = \{x|x_B > x_W\}$. If $\chi_B^d(Q) = \{x|x_B > x_W\}$, by Lemma 13 and by the symmetry assumption of x_B and x_W, $\forall z \in X$, $Q(0, z) = B$. That is, $\chi_B^0(Q) = X \neq \{x|x_B > x_W\}$, and this contradicts the assumption s.t. $\forall t \in \{0, d\}$, $\chi_B^t(Q) = \{x|x_B > x_W\}$. □

Note that the promotion ratio of B with a mechanism $< Q, P >$ is $\mu(\chi_B(Q)) = \nu(d) \cdot \mu(\chi_B^d(Q)) + (1 - \nu(d)) \cdot \mu(\chi_B^0(Q))$. In addition, recall that for the projection mechanism $< Q^c(X_B^d), P^c >$, a_p ($p \in \{G, B, W\}$) specifies $\chi_B^d(Q^c(X_B^d))$ s.t. $X_B^d(a_p) = \chi_B^d(Q^c(X_B^d(a_p)))$ (see (15)).

Lemma 16. *Given $\nu(d)$, suppose that (a_G^*, a_B^*, a_W^*) maximizes the owner's expected profit for the gap, B-bar, and W-bar projection mechanisms, respectively. For each P and for a regulatory ratio $r \in (0, 1)$ s.t. $r > \mu(\chi_B(Q^c(X_B^d(a_p^*)))$, there exists $a_p(r)$ s.t. $r = \mu(\chi_B(Q^c(X_B^d)))$ where $X_B^d = X_B^d(a_p(r))$.*

Proof. Let $Q^c(X_B^d)$ be an arbitrary projection mechanism's allocation rule. Then $\mu(\chi_B(Q_B^c(X_B^d)))$ is a strictly increasing function w.r.t. $\mu(X_B^d)$ as follows.

$$\begin{aligned} & \mu(\chi_B(Q_B^c(X_B^d(a_p)))) \\ = \quad & \nu(d) \cdot \mu(X_B^d(a_p)) \cdot 1 + (1 - \nu(d)) \cdot [\mu(\bar{X}_B) + (1 - \mu(\bar{X}_B)) \cdot \mathrm{pr}(x_B > x_W|(\bar{X}_B)^c)] \end{aligned} \tag{27}$$

As F_i is a continuous cumulative distribution function, given P, (27) is continuous in each a_p. In addition,

$$\lim_{a_W \to \bar{\zeta}} \mu(\chi_B(Q_B^c(X_B^d(a_W)))) = 1;$$

$$\lim_{a_B \to 0} \mu(\chi_B(Q_B^c(X_B^d(a_B)))) = 1;$$

$$\lim_{a_G \to -\bar{\zeta}} \mu(\chi_B(Q_B^c(X_B^d(a_G)))) = 1.$$

Therefore, there exists $a(r)$ for the gap, B-bar, and W-bar projection mechanisms achieving $r = \mu(\chi_B(Q^c(X_B^d(a(r)))))$ where $r \in (\mu(\chi_B(Q^c(X_B^d(a_p^*))), 1)$. □

Corollary 4. *A regulator can implement a specific target ratio $r \in [0, 1]$.*

Proof. Let $< Q^{**}, P^{**} >= \arg\max_{Q,P}\{E[\pi(\cdot; Q^0, P^0)], \quad \max_{a_p}\{E[\pi(\cdot; Q^c(X_B^d(a_p)), P^c)]\}\}$. By Theorem 2 and Lemma 16, given regulation (r, τ), the owner's optimization problem is as follows:

$$\max_{Q,P}\{E[\pi(\cdot;Q^{**},P^{**})]-\tau\cdot 1_{(\mu(\chi_B(Q^{**}))\neq r)},\ E[\pi(\cdot;Q^{\lambda=r},P^{\lambda=r})],\ \max_{a_p(r)}\{E[\pi(\cdot;Q^c(X_B^d(a_p(r))),P^c)]\}\} \tag{28}$$

By (28), a regular can implement a specific target ratio r by setting the punishment level $\tau(r)$ as follows:

$$\tau \geq E[\pi(\cdot;Q^{**},P^{**})]-\max\{E[\pi(\cdot;Q^{\lambda=r},P^{\lambda=r})],\ \max_{a_p(r)}\{E[\pi(\cdot;Q^c(X_B^d(a_p(r))),P^c)]\}\}. \quad \square$$

Definition 10 (Unfairness)**.** Given an arbitrary allocation rule Q and an arbitrary probability that the manager is discriminatory $\nu(d) \in (0,1)$, the unfairness of the allocation rule Q is defined as follows:

$$\phi(Q) = \sum_{t\in\{0,d\}} \nu(t)\cdot[\mu(x_W > x_B|x \in \chi_B^t(Q))\cdot\mu(\chi_B^t(Q))+\mu(x_B > x_W|x \in \chi_W^t(Q))\cdot\mu(\chi_W^t(Q))].$$

Lemma 17 (Unfairness of unconditional mechanisms)**.** *Given an arbitrary $r \in (0,1)$ and an arbitrary set $X_B^d(a_p) \in X$ where $a_p \in (0,\bar{\zeta}) \times \{G,B,W\}$,*

$$\max\{\max_{a_p}\{\phi(Q^c(X_B^d(a_p)))\},\phi(Q^0)\} < \phi(Q^{\lambda=r}).$$

Proof. First, note that by Lemma 15, $\phi(Q^{\lambda=r}) = \frac{1}{2}$. I show that $\phi(Q^0) < \frac{1}{2}$ and $\forall\ a_p \in (0,\bar{\zeta}) \times \{G,B,W\}$, $\phi(Q^c(X_B^d(a_p))) < \frac{1}{2}$.

$$\phi(Q^0) = \nu(d)\cdot\frac{1}{2} + (1-\nu(d))\cdot 0.$$

$$\begin{aligned}\phi(Q^c(X_B^d)) &= \nu(d)\cdot[\mathrm{pr}(x_W > x_B|X_B^d)\cdot\mu(X_B^d) + \mathrm{pr}(x_B > x_W|(X_B^d)^c)\cdot(1-\mu(X_B^d))]\\ &\quad +(1-\nu(d))\cdot\mathrm{pr}(x_W > x_B|\bar{X}_B)\cdot\mu(\bar{X}_B).\end{aligned}$$

Since $\nu(d) < \frac{1}{2}$, $\phi(Q^0) < \frac{1}{2}$. Also, $\forall X_B^d = X_B^d(a_p)$, $\mathrm{pr}(x_W > x_B|X_B^d) < \frac{1}{2}$, $\mathrm{pr}(x_B > x_W|(X_B^d)^c) < \frac{1}{2}$, and $\mathrm{pr}(x_W > x_B|\bar{X}_B) \leq \frac{1}{2}$. Therefore, $\phi(Q^c(X_B^d)) < \frac{1}{2}$. $\square$

Example 5. There exists an environment (d,F,ν) with a regulatory target ratio r s.t.

$$E[\pi(\cdot;Q^{\lambda=r},P^{\lambda=r})] > \max_{a_p(r)}\{E[\pi(\cdot;Q^c(X_B^d(a_p(r))),P^c)\}. \tag{29}$$

Assume that $d = 0.5$, $\forall i \in \{B,W\}$, $x_i \sim \mathrm{Uniform}[0,1]$, and $\nu(d) = 0.9$. Suppose that a regulator wants to implement $r = \frac{1}{2}$. Note that $E[\pi(\cdot;Q^{\lambda=\frac{1}{2}},P^{\lambda=\frac{1}{2}})] = 0.5$, and $\mu(\chi_B(Q^0)) = 0.05$. Each projection mechanism (gap, B-bar, W-bar) can implement

$r = 0.5$ by setting $a_G(\frac{1}{2}) = 0.051$, $a_B(\frac{1}{2}) = 0.513$, and $a_W(\frac{1}{2}) = 0.444$; the corresponding expected profits are $E[\pi(\cdot; Q^c(X_B^d(a_G(\frac{1}{2}))), P^c)] = 0.401$, $E[\pi(\cdot; Q^c(X_B^d(a_B(\frac{1}{2}))), P^c)] = 0.383$, and $E[\pi(\cdot; Q^c(X_B^d(a_W(\frac{1}{2}))), P^c)] = 0.361$. $\square$

References

Armstrong, Mark (1996) "Multiproduct nonlinear pricing," *Econometrica*, Vol. 64, No. 1, pp. 51–75.

Armstrong, Mark and Jean-Charles Rochet (1999) "Multi-dimensional screening:: A user's guide," *European Economic Review*, Vol. 43, No. 4, pp. 959–979.

Arrow, Kenneth J (1973) "The theory of discrimination," *Discrimination in labor markets*, Vol. 3, No. 10, pp. 3–33.

Becker, Gary S (1957) *The Economics of Discrimination*: University of Chicago Press.

Biais, Bruno, David Martimort, and Jean-Charles Rochet (2000) "Competing mechanisms in a common value environment," *Econometrica*, Vol. 68, No. 4, pp. 799–837.

Black, Sandra E and Philip E Strahan (2001) "The division of spoils: rent-sharing and discrimination in a regulated industry," *American Economic Review*, Vol. 91, No. 4, pp. 814–831.

Bodoh-Creed, Aaron L and Brent R Hickman (2018) "College assignment as a large contest," *Journal of Economic Theory*, Vol. 175, pp. 88–126.

Coate, Stephen and Glenn C Loury (1993) "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review*, Vol. 83, No. 5, pp. 1220–1240.

Courty, Pascal and Hao Li (2000) "Sequential Screening," *Review of Economic Studies*, Vol. 67, No. 4, pp. 697–717.

Galperti, Simone (2015) "Delegating Resource Allocation: Multidimensional Information vs. Decisions," *Working Paper*.

Green, Jerry R and Jean-Jacques Laffont (1986) "Partially verifiable information and mechanism design," *Review of Economic Studies*, Vol. 53, No. 3, pp. 447–456.

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017) "Evidence games: Truth and commitment," *American Economic Review*, Vol. 107, No. 3, pp. 690–713.

Krähmer, Daniel and Roland Strausz (2015) "Ex post information rents in sequential screening," *Games and Economic Behavior*, Vol. 90, pp. 257–273.

Lareau, N. Peter et al. (2016) *Labor and Employment Law*: Matthew Bender & Company, Inc.

Myerson, Roger B (1981) "Optimal auction design," *Mathematics of Operations Research*, Vol. 6, No. 1, pp. 58–73.

Peski, Marcin and Balázs Szentes (2013) "Spontaneous discrimination," *American Economic Review*, Vol. 103, No. 6, pp. 2412–2436.

Phelps, Edmund S (1972) "The statistical theory of racism and sexism," *American Economic Review*, Vol. 62, No. 4, pp. 659–661.

Rochet, Jean-Charles and Lars A Stole (2003) "The economics of multidimensional screening," *Econometric Society Monographs*, Vol. 35, pp. 150–197.

Sibley, David S and Padmanabhan Srinagesh (1997) "Multiproduct nonlinear pricing with multiple taste characteristics," *RAND Journal of Economics*, Vol. 28, No. 4, pp. 684–707.

Stiglitz, Joseph E (1973) "Approaches to the Economics of Discrimination," *American Economic Review*, Vol. 63, No. 2, pp. 287–295.

Weber, Andrea and Christine Zulehner (2014) "Competition and gender prejudice: are discriminatory employers doomed to fail?" *Journal of the European Economic Association*, Vol. 12, No. 2, pp. 492–521.

Winter, Eyal (2004) "Incentives and discrimination," *American Economic Review*, Vol. 94, No. 3, pp. 764–773.