Discrimination in Organizations: Optimal Contracts and Regulation

Wiroy Shin
KIET

July 19, 2016

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 a situation in which an individual person practices 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, whereas an owner (the principal) is unbiased and only cares about profit. The manager perfectly observes productivity levels of the subordinates and decides whom to promote. The owner only sees results of the manager’s decision: the promoted worker’s identity and that worker’s performance. 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 the first-best (full information) allocation and discuss the effectiveness of current regulations (e.g., affirmative action, taxation on the minority promotion ratio): I find that a regulator (such as the U.S. Equal Employment Opportunity Commission) can improve compliance with non-discriminatory conduct, despite the fact that the person on whom the regulation is directly incident—the principal—is not intrinsically biased. I also show that the regulation can be counter-productive if it attempts to enforce perfect fairness (the first-best allocation) when that allocation is not incentive feasible. Finally, I review the U.S. law of discrimination and analyze statutory and jurisprudential issues regarding the optimal mechanism.

Keywords: Theories of Discrimination, Mechanism Design, Multidimensional Screening, Optimal Incentives in Hierarchies

JEL Classification: D02, D04, D21, D63, D82, D86

‡ I thank to Edward J. Green for his guidance and support throughout this project. I also gratefully
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 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, and agreed to very costly settlements.

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

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: if the appointed low-level manager may be biased, how can he be controlled by contractual arrangements? This paper answers to 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 provides an incentive for

\[ \text{[footnote:Acknowledgment: The authors thank James Jordan, Paul Grieco, Balázs Szentes, and participants at the 11th World Congress of Econometric Society, the 26th Stony Brook International Conference on Game Theory, the 2015 Public Economic Theory Conference, the 2015 GCER Conference at Georgetown U., and the Fall 2014 Midwest Economic Theory Conference. This research is supported by a grant from the Rock Ethics Institute at Penn State, the Crawford Graduate Fellowship in Ethical Inquiry.} \]

\[ \text{[footnote:1Perhaps, they should also have had mechanisms to prevent prejudiced people from being assigned to the low-level management position, but prejudice may have been so endemic in the societal workforce that such appointments could not be avoided.} \]
the discriminatory manager to promote a minority worker, which reduces discrimination and increases 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 EEOC is controversial. 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 not biased, and it shows that a regulator can improve compliance with non-discriminatory conduct; but, the paper also shows that the regulation can be counter-productive if it attempts to enforce perfect fairness when conditions for perfect fairness are not feasible.

The following describes the model’s environment in this paper. 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 productivity levels of the two subordinates, the owner cannot see them. The manager promotes one of the subordinates, and then the owner will compensate the manager based on what she observes: the promotion decision by the manager and the output of the promoted worker. By the Revelation Principle, I consider a direct mechanism that abstracts the details of any particular contractual setting between the owner and the manager. In the direct mechanism, the manager (the agent) reports all available information to the owner (the principal), and the owner determines the promotion choice based on the report. Also, the manager’s compensation is arranged 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 where the agent has three-dimensional private information (productivity levels of both subordinate workers and the manager’s discriminatory type).

2 In the U.S., the Equal Employment Opportunity Commission (EEOC) is responsible for enforcing federal laws that make it illegal to discriminate against a job applicant or an employee.

3 As the model investigates an agency problem where 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 (who gets promoted; how much the firm produces and pays to the manager).
that is partially revealed (the output of the promoted worker) after promotion.

In summary, this paper deals with a hierarchical environment where discrimination against some subordinates impairs institution’s profit, and it explains what the owner can do best to reduce the discriminatory decisions without compromising the organization’s profit. As the manager can have a discriminatory preference, taste-based discrimination is assumed here. In contrast to prior research, this paper considers a contractual problem between a principal and a labor-related decision maker agent as the agent has authority over the subordinate workers’ promotion (see subsection 2.1 for details). Also, it deals with a regulation problem assuming existence of another principal (a regulator) outside of the organization.

The structure of this paper is as follows. Section 2 reviews literature in 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 that a regulator can improve compliance with non-discriminatory conduct. In section 6, I discuss implications of the model relating to other research in discrimination, anti-discrimination policy (including legal issues on affirmative action), and multidimensional adverse selection. Section 7 explores extensions of the study and concludes the paper.

2 Related Literature

2.1 Theories of Discrimination

There are two main existing research branches in theories of discrimination: theories of taste-based discrimination and theories of statistical discrimination. In the seminal paper in 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.”, and it argues that the distribution of

\footnote{In addition to those two main branches, there is recent research that takes other approaches: they show discriminatory decisions as an optimal outcome in various economic environments even when 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 requiring same effort levels. Koski and Szentes (2010) provides a repeated matching environment, where employers do not want to be matched with the other racial workers in an equilibrium.}

Winter (2004) suggests an equilibrium in a team project environment, in which a principal wants to provide different rewards to team members for requiring same effort levels. Koski and Szentes (2010) provides a repeated matching environment, where employers do not want to be matched with the other racial workers in an equilibrium.
the discriminatory taste among employers determines a difference in market wages between workers with different races. Moreover, the subsequent theoretical and empirical research on the taste-based discrimination (Stiglitz (1973); Black and Strahan (2001)) suggests a positive correlation between market competition and fairness of the market output: 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 to their outcome-relevant features. Phelps (1972) introduces a model of discrimination where statistical distributions of production-skill variables are different in 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 good for an analysis of small organizations consisting of a sole proprietorship with few enough workers where hierarchical management is not necessary, they are not satisfactory to guide policy for large firms. Owners of large firms neither have face-to-face interactions with workers nor get direct benefits from discrimination: they do not even make employment decisions on bottom level workers. The large firms are in charge of most employment in the U.S., and therefore, to maximize efficiency of anti-discrimination policy and to properly analyze agency problems relating to discrimination in large organizations, it is important to have own treatment on large firms where another principal exists above the labor-related decision makers in the hierarchy. This paper develops a model reflecting those hierarchical aspects and 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 in this paper presents distinct fea-

\footnote{Firms with more than 100 workers consist of 63% of U.S. employment in 2014.}
tures that have not been investigated before. The screening problem here deals with the agent (the manager)’s three-dimensional private information where one (the manager’s type on discrimination) of them plays a different role than the other two (productivity levels of B and W) in the agent’s utility: the productivity information implicitly affects to 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 where the consumer’s marginal utility on heterogenous goods directly and exogenously appears on the agent’s utility function ([Armstrong (1990); Sibley and Srinagesh (1997); Armstrong and Rochet (1999)].

Some multidimensional screening problems show aggregation of information ([Armstrong (1990); Biais et al. (2000)]) where a summary statistic of the original multidimensional information is sufficient for the principal. However, the problem discussed here does not exhibit such aggregation feature as each 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.

Another distinguishable feature of the problem in this paper is that one of the information (productivity of the promoted worker) is revealed at a time after the principal observes the promoted worker’s output and before she decides the payment for the agent. Such sequential information revealing characteristics between the two decisions (promotion and payment) relates this paper to research on sequential screening ([Courty and Li (2000); Krahmer and Strausz (2015)]); but this paper is distinguished from those standard sequential screening problems as the person who receives the information update is the principal, not the agent. In that sense, this paper has similarities with existing research on mechanism design with partially verifiable information ([Green and Laffont (1980); Hart et al. (2013)].

Galperti (2015) also addresses a multidimensional resource allocation problem where delegation with simple standards (a floor or a gap) is a solution.
3 Direct Mechanism

3.1 Environment

The following timeline provides a description of the organization’s promotion procedure where a direct mechanism (a promotion rule and a payment rule) contract exists between the 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 to the owner including information regarding his personal discriminatory preference on the workers.

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 following the contract.

Assumption 1 (Discrimination coefficient). If the manager is discriminatory, and the identity of the promoted labor is \( B \), the manager earns disutility equivalent to \( d > 0 \).

\( \theta \in \Theta = \{0, d\} \) denotes the manager’s discrimination coefficient type: he could be either discriminatory \((\theta = d)\) or fair \((\theta = 0)\) where \( \nu : \Theta \rightarrow [0, 1] \) is the probability mass function.

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 either on the organization’s profit or on 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 the reports: 1. The owner specifies a contract. 2. The manager observes 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 productivity of $B$ and $W$, respectively. Each $x_i$ is i.i.d drawn from a set $X_i = [0, \tilde{c}] \in \mathbb{R}_+$, 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 $\mu : X \rightarrow [0, 1]$. The productivity $x_i$ is perfectly correlated with the worker $i$’s output after $i$’s promotion.

Assumption 3 (Absence of outside options). Both the owner and the manager do not have any outside options. That is, the owner cannot fire the manager\footnote{This assumption is reasonable as the owner cannot prove the manager’s discriminatory preference.} nor can the manager reject to provide the reports about $\theta$ and $x$.

3.2 Direct Mechanism $< Q, P >$

Let $t \in \Theta$ be the manager’s report on his discrimination coefficient type, and $z = (z_B, z_W) \in X$ be a productivity report on subordinates $B$ and $W$ by the manager. A direct mechanism $< Q, P >$ consists of an allocation rule $Q : \Theta \times X \rightarrow I$ that appoints one subordinate for promotion and a payment rule $P : \Theta \times X \times X_i \rightarrow \mathbb{R}_+$ 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 $x_{Q(z)}$.

Note that the manager originally has three-dimensional private information, but one of them—$x_{Q(z)}$—is revealed later. That is, in this direct mechanism setting, partial verification of the agent’s private information is enabled by the owner: after she decides the promotion, 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 $\xi_Q(t, z; x) = (z_B, z_W, x_{Q(z)})$. 

---

9This assumption is reasonable as the owner cannot prove the manager’s discriminatory preference.
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 where the manager is deterministically discriminatory ($\nu(d) = 1$). After that, an analysis on the unobservable discrimination type case ($\nu(d) \in (0, 1)$) is investigated.\footnote{For the deterministic case where $\nu(d) = 1$, the manager’s discrimination coefficient $d$ is common knowledge. For the unobservable case, the manager’s discrimination coefficient type ($\theta \in \{0, d\}$) is known only to the manager; the owner knows only the distribution of $\theta$, $\nu(\theta)$.} The study on the deterministic case helps to understand the unobservable case in two ways: 1. It provides necessary conditions for the optimal mechanism of the unobservable case; 2. The optimal mechanism of the deterministic case can be a simple alternative improving the status quo of the unobservable case.

4.1 The manager is discriminatory, $\nu(d) = 1$.

Under the assumption, $\nu(d) = 1$, information regarding the workers’ productivity levels are only relevant to the owner’s mechanism design problem. Therefore, the manager doesn’t need to report $t$. Accordingly, the allocation rule, the owner’s informational state, and the payment rule are simplified without $t$: $Q : X \rightarrow I$, $\xi_Q : X \times X \rightarrow X \times X$, and $P : X \times X_i \rightarrow \mathbb{R}_+$.

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

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

Since a budget for the manager’s compensation is limited to the organization’s output $x_{Q(z)}$, $P(\xi_Q(z; x)) \in [0, x_{Q(z)}]$. Given $< Q, P >$, if the manager reports $z$, the manager’s utility is

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

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

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

Given an allocation rule $Q$, organization’s welfare is defined as a sum of the owner’s
profit and the manager’s compensation, which is equivalent to the promoted worker’s output. Note that the discriminatory coefficient \( d \) is not included in the organization’s welfare.

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

When two or more mechanisms are compared to each other, the specific \(< Q, P >\) is notated in each function (e.g., \( \pi(z; x, Q, P) \)).

With \(< Q, P >\) and the manager’s truthful reporting, profit of the owner is \( \pi(x; x) = x_{Q(x)} - P(\xi_Q(x; x)) \). Then, the owner’s problem is about choosing optimal \( Q \) and \( P \) to maximize the expected profit, subject to the incentive compatibility constraint.\(^{11}\) That is,

\[
\max_{Q, P} \int_{x \in X} f(x) \cdot \pi(x; x) \, dx \quad \text{s.t.} \quad u(x; x) \geq u(z; x) \quad \forall x, z \in X.
\] (2)

### 4.1.1 Full Information benchmark

Suppose that no information gap exists between the manager and the owner. Then, the owner doesn’t need to pay information rent to the manager, and she can promote whomever has a higher productivity. Therefore, an efficient allocation (the first-best allocation) that maximizes organization’s welfare \( w(z, x) \) in (1) can be achieved with an allocation rule \( Q \) as a function of the true productivity value \( x \). The following mechanism \(< Q^F, P^F >\) achieves the first-best allocation and expected profit maximization as well.

\[
\begin{align*}
Q^F(x) &= B, & \text{if } x_B > x_W \\
Q^F(x) &= W, & \text{if } x_B < x_W \\
Q^F(x) &= B \text{ or } W, & \text{if } x_B = x_W.
\end{align*}
\] (3)

\[
P^F(\xi_Q(z; x)) = 0, \quad \forall \xi_Q(z, x) \in X \times X_i.
\]

\(^{11}\) The manager’s participation constraint is not included in the owner’s optimization problem as the manager doesn’t have an outside option.
Define $x_{(1)} = \max\{x_B, x_W\}$. Following $< Q^F, P^F >$, expected profit in the first-best case is $E[\pi(\cdot; Q^F, P^F)] = E(x_{(1)})$. Expected surplus of the organization is defined by a sum of the owner’s expected profit and the manager’s expected utility: $E(s(\cdot; Q^F, P^F)) = E[\pi(\cdot; Q^F, P^F)] + E[u(\cdot; Q^F, P^F)] = E(x_{(1)}) - d \cdot \Pr(x_B > x_W)$. Note that under the first-best allocation, expected profit is equivalent to expected welfare, because the payment to the manager is always zero. Such payment arrangement causes that the expected welfare $E[w(\cdot; Q^F, P^F)]$ exceeds the expected surplus $E[s(\cdot; Q^F, P^F)]$: whenever $B$ is promoted, it creates negative externality $d$ to the manager.

4.1.2 Optimality conditions

Let $\mathcal{M}^*$ 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.

**Definition 2** (Unconditional mechanism). 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 information state. That is,

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

Let $\mathcal{Q}^\lambda = \{Q^\lambda| \lambda \in [0, 1]\}$. 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, and the manager always promotes $W$. Therefore, the situation is equivalent to the unconditional allocation s.t. $\forall z, Q(z) = W$. Under unconditional mechanisms, since the productivity levels of $B$ and $W$ are ex-ante identical, expected profit of the owner is equal to $E(x_i), \forall i \in \{B, W\}$:

$$
E[\pi(\cdot; Q^\lambda, P^\lambda)] = \lambda \cdot E(x_B) + (1 - \lambda) \cdot E(x_W) = E(x_B) = E(x_W). \quad (4)
$$
Note that given \(< Q^\lambda, P^\lambda >\), the manager’s utility is

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

**Lemma 1** (Unconditional mechanism). *Any unconditional mechanism is incentive compatible.*

**Proof.** For arbitrary \(\lambda, < Q^\lambda, P^\lambda >\), the payment rule \(P^\lambda\) 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. \(\square\)

Recall that the agent’s private information is partially verifiable. Next, I propose a type of untruthful reports by the manager, which the owner can detect. Subsequently, I provide a lemma about a punishment level for the manager’s untruthful strategy.

**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 the owner promotes \(W\). After 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 the owner promotes \(W\). After promotion, the owner realizes the output \(x_Q(z) = x_W = 0.7 = z_W\). In this case, the lie is not detected.

Next, I present a lemma s.t. in order to characterize the optimal mechanism, we can concentrate on the compensation schemes that punish detectible lies by giving a minimum level of compensation to the manager.
Lemma 2 (Maximum punishment on the detectable lie). 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. 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:

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

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 \( < Q, P^0 > \in \mathbb{M}^* \). \( \square \)

The following lemma shows that if two reports (one true and one false) produce the same outcome, then a payment scheme should treat them equally in a set of incentive compatible mechanisms.

Lemma 3 (IC). Suppose that \( < Q, P > \) is IC. 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. The following inequalities prove the lemma.

\[
P(\xi_Q(x; x)) \geq P(\xi_Q(z; x)) \quad , \text{IC} \\
= P(z; x_{Q(x)}) \quad , Q(x) = Q(z) \\
= P(z; z_{Q(z)}) \quad , x_{Q(x)} = z_{Q(z)} \\
\geq P(x; z_{Q(x)}) \quad , \text{IC} \\
= P(\xi_Q(x; x)) \quad , z_{Q(z)} = x_{Q(z)} = x_{Q(x)} \quad . \quad \square
\]

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

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.

**Lemma 4 (IC).** Suppose that \( < Q, P > \) is IC, and \( Q \in Q_c^2 \). Also, assume 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.** 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 \( d \). \( \square \)

In Lemma 5, the optimal payment scheme for conditional allocation rules are presented. It shows that the output of the promoted worker is useful only for distinguishing detectable lies. The optimal payment scheme with given \( Q \in Q_c^2 \) compensates the agent only when his report is not a detectable lie and induces \( B \)'s promotion. The amount of such compensation does not vary in 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 \( Q \in Q_c^2 \), the following payment rule \( P \) formulates \( < Q, P > \) to maximize 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.** First, I show that given \( Q \in Q_c^2 \), \( P \) maximizes expected profit by presenting that \( P \) provides a minimum payoff to the agent among possible payoffs satisfying necessary conditions of IC that described in 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 a 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 possible minimum transfer. Therefore, given $Q \in Q^{co}$, the payment rule $P$ presented in this lemma maximizes expected profit.

To check the IC condition, choose an arbitrary conditional allocation rule $Q \in Q^{co}$. First, I consider a true productivity type vector $x$, where $Q(x) = W$. By 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 $x$ either or both of in $x_B$ or in $x_W$, and the arbitrary $Q$ can allocate such $z$ to $B$ or $W$. By (P1), any deviations to a detectible lie yield weakly smaller utility to the agent than the truthful report. Excluding such detectible lie deviations, only two kinds of deviation possibilities 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), both deviations do not produce higher utility for the agent resulting $u(z, x) = 0$. Therefore, under $P$, $\forall x$ s.t. $Q(x) = W$, the agent report truthfully. Another case, where $Q(x) = B$ is proved in a similar manner. □

4.1.3 The Main Result

The following arrangement uniquely achieves profit maximization subject to the owner’s limited information and to the manager’s incentive compatibility constraint.

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 as much as 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 expected profit subject to the incentive compatibility constraint:
\[ Q^*(z) = B \quad \text{if} \quad z_B - d > z_W \]
\[ Q^*(z) = W \quad \text{if} \quad z_B - d < z_W \]
\[ Q^*(z) = B \text{ or } W \quad \text{if} \quad z_B - d = z_W. \]

(5)

\[ P^*(\xi_Q(z; x)) = 0 \quad \text{if} \quad \xi_Q(z; x) \in \Xi_Q^B \]
\[ P^*(\xi_Q(z; x)) = d \quad \text{if} \quad \xi_Q(z; x) \notin \Xi_Q^B \text{ and } Q(z) = B \]
\[ P^*(\xi_Q(z; x)) = 0 \quad \text{if} \quad \xi_Q(z; x) \notin \Xi_Q^B \text{ and } Q(z) = W. \]

**Proof.** First, I show that the optimal allocation rule \( Q^* \) achieves the maximum profit among incentive compatible non-constant mechanisms. By Lemma 5, given an arbitrary conditional allocation rule \( Q \in Q^{co} \), the payment rule \( P^* \) maximizes 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 \( x_B - d > x_W \), it is optimal to promote \( B \). Otherwise, promoting \( W \) is profitable. Therefore, \( Q^* \) is an unique allocation rule. Accordingly, the mechanism \(< Q^*, P^* >\) is incentive compatible and achieves a maximum profit among non-constant mechanisms.

Next, I compare this maximum profit to the constant 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 an optimal solution to the problem (2). \( \square \)

**Example 3.** Suppose that each \( x_i \sim \text{Uniform}[0, 1] \) 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; whereas if \((x_B, x_W) = (0.8, 0.61)\), \( W \) is promoted and the owner does...
not pay to 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 the one of the status quo where \( B \) is not promoted in any cases, the optimal mechanism enables \( B \) with \( x_B \sim [0.2, 1] \) to have a chance to be promoted with probability \((x_B - 0.2)\).

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

The optimal mechanism \(< Q^*, P^* >\) suggests that in order to maximize the organization’s profit, the owner should provide an incentive to the discriminatory management encouraging to promote more qualified \( B \). After enforcing the optimal mechanism, the expected welfare (which are equivalent to the promoted worker’s output) is

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

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 as much as \( d \cdot \text{pr}(x_B - d > x_W) \). Note that comparing to the status quo, the owner’s expected profit increases by \( E(y_{(1)}) - E(x_W) \), and \( B \)'s promotion probability increases by \( (\text{pr}(x_B - d > x_W) - 0) \). In contrast, the manager’s utility level remains the same regardless of the outcome.

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. Note that in an equilibrium path, as Lemma 5 suggests, the output of the promoted worker does not matter to the owner in deciding the manager’s compensation. Such information is only used to recognize detectable lies. In other words, it is optimal that compensation of the manager is determined only by the identity of the promoted worker.

\(^{13}\)For those of whom consider the optimal mechanism \(< Q^*, P^* >\) politically improper (though
For example, suppose that the owner employs a linear compensation scheme that she pays a bonus proportionally to the outcome of the promoted worker. That is, the manager receives $\alpha x_i$, where $\alpha \in (0, 1)$ and $i$ is the identity of the promoted worker. In this case, the manager only chooses $B$ if $x_B - x_W > d/\alpha$. Since $d/\alpha > d$, the threshold of the productivity difference promoting $B$ is higher than the optimal mechanism case. As a result, more qualified $B$ is not promoted, and it causes such a contract to be sub-optimal.

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

**Corollary 1.** 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 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, expected profit of the owner decreases by $E(\max\{x_B, x_W\}) - E(\max\{x_B - d, x_W\}) > 0$.

**Proof.** Section 4.1.1 and the proof of Theorem 1 implies this corollary. □

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

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

it actually ameliorates discrimination situation comparing with the status quo), it is worth to devise another indirect contract that exactly implements the optimal direct mechanism. However, such a contract will not improve any economic agents’ utilities relating to the problem, but it might be helpful alleviating opposition to executing the optimal mechanism.
\[
\pi(t, z; x) = x_{Q(t, z)} - P(\xi_Q(t, z; x)).
\]
\[
u(t, z; \theta, x) = P(\xi_Q(t, z; x)) - d \cdot 1_{Q(t, z) = B} \cdot 1_{\theta = d}.
\]

Given \( u(\cdot) \), the incentive compatibility condition is \( \forall \theta, t \in \Theta \) and \( \forall x, z \in X \),
\[
P(\xi_Q(\theta, x; x)) - d \cdot 1_{Q(\theta, x) = B} \cdot 1_{\theta = d} \geq P(\xi_Q(t, z; x)) - d \cdot 1_{Q(t, z) = B} \cdot 1_{\theta = d}.
\] (6)

Then, the following represents the owner’s optimization problem subject to the incentive compatibility condition.
\[
\max_{Q, P} \sum_{\theta \in \Theta} \int_{x \in X} \nu(\theta) \cdot f(x) \cdot \pi(\theta, x; x) \, dx
\]
\text{s.t. } u(\theta, x; \theta, x) \geq u(t, z; \theta, x) \quad \forall \theta, t \in \Theta \text{ and } \forall x, z \in X.
\] (7)

4.2.1 Incentive compatible mechanisms

Given an allocation rule \( Q \), define a set of workers’ productivity levels that result in worker \( i \)’s promotion when a discriminatory type \( t \) is reported: \( \chi^i(Q) = \{z \mid Q(t, z) = i\} \).

Recall that \( Q^\lambda \) is a set of unconditional allocation rules, which promotes \( B \) with probability \( \lambda \).

**Definition 5** (Unconditional mechanism). 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 information state. That is,
\[
\forall t, z \text{ and } \forall \xi_Q(t, z, x), \quad Q^\lambda(t, z) = B \text{ with probability } \lambda
\]
\[
Q^\lambda(t, z) = W \text{ with probability } 1 - \lambda, \text{ and } \quad P^\lambda(\xi_Q(t, z; x)) = 0.
\] (8)
Given any λ, the owner’s expected profit with the unconditional mechanism is

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

**Lemma 6 (Unconditional mechanism).** Any unconditional mechanism is incentive compatible.

**Proof.** A proof of this lemma is equivalent to the proof of Lemma \(\Box\). \hspace{1cm} \Box

Under the status quo, the manager always promotes \(W\) if he is discriminatory. If he is not discriminatory, it is assumed that the manager follows the first-best allocation rule where any worker with higher productivity is selected. The following Delegation mechanism depicts the status quo.

**Definition 6 (Delegation mechanism).** Define the following direct mechanism \(\langle Q^0, P^0 \rangle\) as Delegation mechanism that reflects the status quo:

\[ Q^0(t, z) = \begin{cases} B & \text{if } t = 0 \text{ and } z_B \geq z_W \\ W & \text{otherwise.} \end{cases} \]

\[ P^0(\xi_Q(t, z; x)) = 0, \quad \forall t \in \Theta \text{ and } \forall x, z \in X. \]

Recall \(\nu(d) = \Pr(\theta = d)\). Then, expected profit of Delegation mechanism is

\[ E[\pi(\cdot; Q^0, P^0)] = \nu(d) \cdot E(x_W) + (1 - \nu(d)) \cdot E(\max\{x_B, x_W\}). \] \hspace{1cm} (11)
\[ i = B \]
\[ \{ x : x_B > x_W \}, 0 \]
\[ t = 0 \]
\[ (0, 0) \]
\[ t = d \]
\[ (X, 0) \]
\( (a) \) Allocation and Payment outcome: \( (\chi^0_i(Q^0), P^0(\xi_Q(t, z; x))) \)

\[ i = W \]
\[ \{ x : x_B < x_W \}, 0 \]
\[ t = 0 \]
\[ (0, 0) \]
\[ t = d \]
\[ (X, 0) \]

\[ \theta = 0 \]
\[ i = B \]
\[ i = W \]
\[ t = 0 \]
\[ 0 \]
\[ 0 \]
\[ t = d \]
\[ 0 \]

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

\[ \theta = d \]
\[ i = B \]
\[ i = W \]
\[ t = 0 \]
\[ -d \]
\[ 0 \]
\[ t = d \]
\[ 0 \]

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

Table 2: Delegation mechanism

**Lemma 7** (Delegation mechanism). *Delegation mechanism is incentive compatible.*

**Proof.** As Table 2 shows, for every type \( (\theta, x) \in \Theta \times X \) of the agent, there is no opportunity obtaining higher \( u(t, z; \theta, x) \) than \( u(\theta, x; \theta, x) \). For example, as Table 2(c) describes, if the manager with \( x \in \chi^d_W = X \) and \( \theta = d \) deviates to \( t = 0 \) and \( z = x \) when \( \theta = 0 \), it would occur utility loss as much as \(-d\). □

**Definition 7** (Projection Mechanism). For an arbitrary \( X_B^d \subset X \) s.t. \( X_B^d \neq \emptyset \) and \( X_B^d \neq X \), let \( \bar{X}_B = \text{proj}_B(X_B^d) \times X_W \subset X \). Define a projection mechanism \( < Q^c(X_B^d), P^c > \) as follows:
\[ Q^c(t, z; X^d_B) = \begin{cases} 
B & \text{if } t = 0 \text{ and } [z \in X_B \text{ or } \forall (z_B, z_W) \in X \setminus X_B, \ z_B \geq z_W]; \\
\text{or } t = d \text{ and } (z_B, z_W) \in X^d_B \\
W & \text{if } t = 0 \text{ and } \forall (z_B, z_W) \in X \setminus X_B, \ z_B \leq z_W; \\
\text{or } t = d \text{ and } (z_B, z_W) \notin X^d_B. 
\end{cases} \]

\[ P^c(\xi_Q(t, z; x)) = \begin{cases} 
\nu(\cdot) \cdot [E(x_B - d|X^d_B) \cdot \mu(X^d_B) + E(x_W|(X^d_B)^c) \cdot \mu((X^d_B)^c)] \\
+ (1 - \nu(\cdot)) \cdot [E(x_B - d|\tilde{X}_B) \cdot \mu(\tilde{X}_B) + E(\max\{x_B, x_W\}|(\tilde{X}_B)^c) \cdot \mu(\tilde{X}_B)^c)] 
\end{cases} \]

\[ = \nu(\cdot) \cdot [E(x_B - d|X^d_B) \cdot \mu(X^d_B) + E(x_W|(X^d_B)^c) \cdot \mu((X^d_B)^c)] + (1 - \nu(\cdot)) \cdot E(y_2) \]

\[ y_2 = \begin{cases} 
x_B - d & \text{if } x \in \tilde{X}_B \\
\max\{x_B, x_W\} & \text{if } x \in X \setminus \tilde{X}_B. 
\end{cases} \]

Given \( X^d_B \), expected profit of the projection mechanism is

\[ E[\pi(\cdot; Q^c(X^d_B), P^c)] \]

Note that in projection mechanisms, the productivity set \( X^d_B \) plays an important role: \( X^d_B \) not only determines promotion of \( B \) when the manager reports \( t = d \), but also it decides the allocation and the payment outcome when \( t = 0 \). Largely, \( X^d_B \) can be divided into three cases: Case 1. \( X^d_B \) is a correspondence of \( z_B \) only; Case 2. \( X^d_B \) is a correspondence of \( z_W \) only; Case 3. \( X^d_B \) is a correspondence of both \( z_B \) and \( z_W \).

Let a projection mechanism with each of \( X^d_B \) in (15) be called as follows: \( B \)-bar projection mechanism, \( W \)-bar projection mechanism, and Gap projection mechanism.
\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\multicolumn{2}{|c|}{i = B} & \multicolumn{2}{|c|}{i = W} \\
\hline
\hline
\(t = 0\) & \((\tilde{X}_B, d)\) & \((\{ (\tilde{X}_B)^c : x_B > x_W \} , 0)\) & \\
\hline
\(t = d\) & \((X_B^d, 0)\) & \(( (X_B^d)^c, 0)\) & \\
\hline
\end{tabular}
\caption{Allocation and Payment outcome: \((\chi^*_i(Q^*(X_B^d)), P^c(\xi_Q(t, z; x)))\)}
\end{table}

(a) Allocation and Payment outcome: \((\chi^*_i(Q^*(X_B^d)), P^c(\xi_Q(t, z; x)))\)

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\multicolumn{2}{|c|}{\(i = B\)} & \multicolumn{2}{|c|}{\(i = W\)} & \multicolumn{2}{|c|}{} \\
\hline
\hline
\(\theta = 0\) & \(\tilde{X}_B\) & \((\tilde{X}_B)^c\) & \(\tilde{X}_B\) & \((\tilde{X}_B)^c\) & \\
\hline
\(t = 0\) & \(d\) & 0 & 0 & 0 & 0 \\
\hline
\(t = d\) & \(d\) & 0 & 0 & 0 & 0 \\
\hline
\end{tabular}
\caption{Manager’s utility \(u(t, z; \theta, x)\) where \(t \in \{0, d\}\) and \(z = x\) when \(\theta = 0\) and \(x \in \chi^*_i(Q^*(X_B^d))\)}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\multicolumn{2}{|c|}{\(i = B\)} & \multicolumn{2}{|c|}{\(i = W\)} & \multicolumn{2}{|c|}{} \\
\hline
\hline
\(\theta = d\) & \(\tilde{X}_B\) & \((\tilde{X}_B)^c\) & \(\tilde{X}_B\) & \((\tilde{X}_B)^c\) & \\
\hline
\(t = 0\) & 0 & \(-d\) & 0 & 0 & 0 \\
\hline
\(t = d\) & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{tabular}
\caption{Manager’s utility \(u(t, z; \theta, x)\) where \(t \in \{0, d\}\) and \(z = x\) when \(\theta = d\) and \(x \in \chi^*_i(Q^*(X_B^d))\)}
\end{table}

(b) The manager’s utility \(u(t, z; \theta, x)\) where \(t \in \{0, d\}\) and \(z = x\) when \(\theta = 0\) and \(x \in \chi^*_i(Q^*(X_B^d))\)

(c) The manager’s utility \(u(t, z; \theta, x)\) where \(t \in \{0, d\}\) and \(z = x\) when \(\theta = d\) and \(x \in \chi^*_i(Q^*(X_B^d))\)

Table 3: Projection mechanism with \(X_B^d\)

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

Lemma 8 (Projection mechanism). Any projection mechanism is incentive compatible.

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

\[\square\]

\begin{lemma}
Suppose that the optimal mechanism of the deterministic discriminatory coefficient case \(< Q^*, P^* >\) in Theorem \ref{thm:discriminant} is proposed to the manager regardless of his discriminatory type when the owner has incomplete information on \(\theta\). Then, the outcome will be equivalent to Gap projection mechanism outcome where \(X_B^d = \{ z \mid z_B - z_W > d \}\).
\end{lemma}

\[14\text{When } X_B^d \text{ is a correspondence of only } z_B \text{ (or } z_W), B \text{ (or } W)-\text{bar projection is the optimal correspondence to consider for the owner’s optimization problem. When } X_B^d \text{ is a correspondence of both } z_B \text{ and } z_W, \text{ I consider the gap } z_B - z_W \text{ for its tractability.}\]
Figure 1: Productivity Region for B’s promotion, where $a_B = a_W = a_G = \delta$

Proof. 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 by $(z_B, z_W)$ solely, and the payment is decided by the identity of the allocation and existence of a detectible lie. Therefore, to earn the payment $d$, if the agent is not discriminatory with $\theta = 0$, he would lie if $x_B - d < x_W$ and $(x_B, x_W) \in \tilde{X}_B$, since the lie is not detectible. 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 $x \in (\tilde{X}_B)^c$ and $x_B > x_W$. □

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 probability of being discriminatory is high enough for the manager. Recall that $y_1 = \max\{x_B - d, x_W\}$ and $x_1 = \max\{x_W, x_W\}$, and let

$$\tilde{\nu} = \frac{E(x_1) - E(y_2)}{(E(x_1) - E(y_2)) + (E(y_1) - E(x_W))}.$$ 

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

**Proof.** The lemma is proved by taking a 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 (14). □

**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.** 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.

$$P(d, x'; x_B') \leq P(0, x'; x_W') \quad (16)$$

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

$$P(0, x'; x_W') \leq P(d, x'; x_B') - d. \quad (17)$$

Since $d \geq 0$, (16) and (17) contradicts to each other. Therefore, $\chi_B^d(Q) \cap \chi_W^0(Q) = \emptyset$ and $\chi_B^d(Q) \subset (\chi_W^0(Q))^c = \chi_B^0(Q)$. □

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

**Proof.** 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 $(t = 0, x = \bar{x})$ can report as $(t = d, z = x')$ without being detected, and can earn the higher utility: $P(\xi_Q(d, x', x')) - \theta = P(\xi_Q(d, x', x')) - \theta$. This contradicts to the assumption that $< Q, P >$ is incentive compatible. □

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

**Proof.** By way of contradiction, suppose $Q(0, \bar{x}) = W$. Note that $\bar{x} \in \chi_B^d(Q)$ or $\bar{x} \in \chi_B^0(Q)$. □
\( \chi^d_W(Q) \). By Lemma 12 and Lemma 3,

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

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

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

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

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

This contradicts to the assumption that \(< Q, P >\) is incentive compatible. □

4.2.2 The Main Result

Let \( \mathcal{M}^\ast \) be a set of optimal mechanisms, i.e., the solution to the owner’s profit maximization problem subject to the incentive compatibility constraint in (1). Next theorem claims that the optimal mechanism is either one of the projection mechanisms or Delegation mechanism.

**Theorem 2** (Optimal mechanism). \( \mathcal{M}^\ast \subseteq \bigcup_{\chi^d_B \subset X} \{ < Q^c(X^d_B), P^c > \} \cup \{ < Q^0, P^0 > \} \).

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

1. \( Q: \chi^d_B(Q) = \emptyset \).

   Since \( \chi^d_B(Q) \cup \chi^d_W(Q) = X \), \( \chi^d_W(Q) = X \). 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 to Delegation mechanism \( < Q^0, P^0 > \) described in (11), and by Lemma 4 \( < Q^0, P^0 > \) is incentive compatible.
Therefore, \(<Q^0, P^0>\) is the optimal mechanism subject to the constraint of \(Q\): 
\(\chi^d_B(Q) = \emptyset\).

2. \(Q: \chi^0_B(Q) = \emptyset\).

Since \(\chi^0_B(Q) = \emptyset, \chi^0_W(Q) = X\). Then, by Lemma 1, \(\chi^d_B = \emptyset\). For profit maximization, let \(\forall t \in \Theta, \forall z, x \in X, P(\xi_Q(t, z, x)) = 0\). This matches to the unconditional mechanism \(<Q^\lambda, P^\lambda>\) with \(\lambda = 0\), and by Lemma \(\emptyset <Q^\lambda=0, P^\lambda=0>\) is incentive compatible. Therefore, \(<Q^\lambda=0, P^\lambda=0>\) is the optimal mechanism subject to the constraint of \(Q: \chi^0_B(Q) = \emptyset\). However, \(<Q^\lambda=0, P^\lambda=0>\) is dominated by Delegation mechanism \(<Q^0, P^0>\) in terms of expected profit: 
\(E[\pi(\cdot; Q^0, P^0)] - E[\pi(\cdot; Q^\lambda=0, P^\lambda=0)] = (1 - \nu(d)) \cdot [E(\max\{xB, xW\}) - E(xW)] > 0\). Therefore, \(<Q^\lambda=0, P^\lambda=0>\notin M^*\).

3. \(Q: \chi^d_B(Q) \neq \emptyset, \forall t \in \Theta\).

If \(\chi^d_B(Q) = X, \) by Lemma 2, \(\chi^d_B(Q) = X\). 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, \(E[\pi(\cdot; Q^0, P^0)] - E[\pi(\cdot; Q^\lambda=1, P^\lambda=1)] = (1 - \nu(d)) \cdot [E(\max\{xB, xW\}) - E(xW)] > 0\). Therefore, \(<Q^\lambda=1, P^\lambda=1>\notin M^*\).

Suppose \(\chi^d_B(Q) = X^d_B \subset X\). Then, by Lemma 3, \(Q(0, \bar{x}) = B\). For profit maximization and following Lemma 2 and Lemma 3, let \(P(\xi_Q(t, z; x)) = d\) if \(t = d\) and \(z \in X^d_B\) or \(t = 0\) and \(z \in X^d_B\). 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 \(<Q^c(X^d_B), P^c>\), and by Lemma 3, \(<Q^c(X^d_B), P^c>\) is incentive compatible. \(\square\)

As Theorem 2 suggests, other than Delegation mechanism, projection mechanisms only satisfy necessary conditions of the optimal mechanism. In the next example, I analyze the three projection mechanisms proposed in (E1) (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 of the manager being discriminatory.

**Example 4.** Suppose that \(\forall i \in \{B, W\}, x_i \sim \text{Uniform}[0, 1],\) and \(d = 0.2\). I consider three different levels for the probability that the manager is discriminatory: \(\nu(d) \in\)
{0.1, 0.5, 0.9}. For each \( \nu(d) \), to figure out which mechanism maximizes 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 promotes \( B \) when \( t = d \). Recall that the following three \( X^d_B \) describe such rules for B-bar, W-bar, and Gap projection mechanisms:

\[
X^d_B = \left[ a_B, 1 \right] \times [0, 1]; \quad X^d_B = [0, 1] \times [0, a_W]; \quad X^d_B = \{ z_B, z_W | z_B - z_W > a_G \}.
\]

Table 4 presents the maximum profit and \( \text{argmax } a^*_P \), \( P \in \{B, W, G\} \) for each mechanism (see appendix B.1 for the owner’s expected profit functions of the three projection mechanisms).

<table>
<thead>
<tr>
<th>( \nu(d) )</th>
<th>Delegation</th>
<th>Gap</th>
<th>B-bar</th>
<th>W-bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum profit</td>
<td>0.516</td>
<td>0.570</td>
<td>0.551</td>
<td>0.520</td>
</tr>
<tr>
<td>Argmax</td>
<td></td>
<td>0.270</td>
<td>0.726</td>
<td>0.300</td>
</tr>
<tr>
<td>( \nu(d) = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximized profit</td>
<td>0.583</td>
<td>0.583</td>
<td>0.585</td>
<td>0.422</td>
</tr>
<tr>
<td>Argmax</td>
<td></td>
<td>1.000</td>
<td>0.904</td>
<td>0.300</td>
</tr>
<tr>
<td>( \nu(d) = 0.1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximized profit</td>
<td>0.650</td>
<td>0.650</td>
<td>0.650</td>
<td>0.3245</td>
</tr>
<tr>
<td>Argmax</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.300</td>
</tr>
</tbody>
</table>

Table 4: Maximum expected profit and \( \text{Argmax } a^*_P \)

As Table 4 shows, when the probability of the manager being discriminatory is high (\( \nu(d) = 0.9 \)), 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 4 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 \( (z_B, z_W) \in \bar{X}_B = X^d_B \times [0, 1] \) and compensate \( d \) to the manager: that is, she needs to pay information rent for the case that the true productivity levels are in the projection set and the manager is a fair type. Therefore,
such extra information rent increases the optimal lower bound of the productivity gap $a_G^*$ for Gap mechanism, and such $a_G^*$ is higher than the discrimination coefficient $d = 0.2$.

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

When the probability of the manager being discriminatory is low ($\nu(d) = 0.1$), Delegation mechanism provides the highest profit. That is, providing no incentive for the manager and delegating all authorities are recommended for the owner in that case. Also, note that $a^* = 1$ for both Gap and B-bar projection mechanisms; as $a \rightarrow 1$, all projection mechanisms converge to Delegation mechanism. In other words, the argmax $a^*$ does not exist in $[0, 1)$, and therefore, Gap and B-bar projection mechanisms are inferior to Delegation mechanism when $\nu(d) = 0.1$.

Figure 2 presents maximum expected profits 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 profits of the four mechanisms: Delegation, Gap, B-bar, and W-bar](image-url)
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 a personally biased, regulatory incentives are likely to improve the second-best allocation in terms of fairness. That is, it is feasible that regulations on the organization increase a probability that more qualified worker to be promoted regardless of his or her demographic category. Such regulation-mechanisms do not maximize the owner’s expected profit, so the following allocations would not obtain the first-best expected profit though the allocation rule is close to the organization’s first-best. Current research on regulations for discrimination problems is based on the principal’s discriminatory taste. Applying those regulations to the setting in which the agent has discriminatory preference, we can ask how the effects of such regulations change.

Suppose that the organization owns a nonatomic continuum of identical branches, and let \( m \) be the representative agent (the manager) of them. All assumptions about the owner, the manager, and the workers remain the same as section 2.1. 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 ratio to be \( r \). If the organization fails to achieve the target ratio, there is a levy \( \tau \). In this section, given \( (r, \tau) \), the owner’s problem deriving the optimal mechanism \( < Q, P > \) is first analyzed. After that, the regulator’s optimization scheme is discussed.

5.1 Deterministic discrimination type case

I first discuss the special deterministic case where \( \nu(d)=1 \) in this section, and then I present the unobservable case \( (\nu(d) \in (0,1)) \) in appendix B.3 as results of those two 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 \( (r, \tau) \), the owner’s optimization problem in (2) changes as follows combining the laissez-faire profit \( \pi(x; x) \) and the
regulatory penalty $\tau$.

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

(18)

**Definition 8** (Gap mechanism). A mechanism $\langle Q^\delta, P^\delta \rangle$ is *Gap mechanism*, if it promotes $B$ whenever the productivity difference between $B$ and $W$ exceeds $\delta$ and if it provides the same compensation as $P^*$. That is,

\[
Q^\delta(z) = B \quad \text{if } z_B - z_W > \delta \\
Q^\delta(z) = W \quad \text{if } z_B - z_W < \delta \\
Q^\delta(z) = B \text{ or } W \quad \text{if } z_B - z_W = \delta.
\]

\[
P^\delta(\xi_Q(z, x)) = 0 \quad \text{if } Q^\delta(r)(z) = W \lor \xi_Q(z; x) \in \Xi_Q^P \\
P^\delta(\xi_Q(z, x)) = d \quad \text{if } Q^\delta(r)(z) = B \land \xi_Q(z; x) \notin \Xi_Q^P
\]

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

**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, $\forall Q \in Q^{co}_r$, Gap mechanism $\langle Q^\delta(r), P^\delta(r) \rangle$ achieves profit maximization of the owner’s problem in (IS).

**Proof.** By Lemma 5, given an arbitrary $Q \in Q^{co}$, the payment rule $P^\delta(r)$ maximizes expected profit subject to the incentive compatibility constraint. Also, by the fact that $d \cdot r$ is a constant, $\forall Q \in Q^{co}_r$, the owner’s problem (IS) is simplified as below.

\[
\max_{Q \in Q^{co}} E(x_W \mid Q(x) = W) \cdot (1 - r) + E(x_B \mid Q(x) = B) \cdot r
\]  

(19)

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

Take arbitrary subsets \( a_B \subset A_B \) and \( a_W \subset A_W \) where \( \mu(a_B) = \mu(a_W) = \frac{3}{2} \), 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 \( \Delta \) from \( a_B \) and \( a_W \) with \( Q''_r: \Delta_a > \frac{9}{2} \cdot \delta(r), \quad \Delta_a < -\frac{9}{2} \cdot (-\delta(r)). \) Consequently, \( \Delta = \Delta_a + \Delta_a > 0 \). This contradicts to the fact that \( Q' \in Q_{r}^{co} \) is optimal. Therefore, subject to \( Q \in Q_{r}^{co}, < Q'^{\delta(r)}, P^{\delta(r)} > \) is a solution to the owner's problem. \( \square \)

Expected profit of \( \delta(r) \)-Gap mechanism is
\[
E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] = E(x_W | x_W > x_B - \delta(r)) \cdot (1 - r) + E(x_B - d | x_W < x_B - \delta(r)) \cdot r. \tag{20}
\]

Suppose that the second-best allocation by \( < Q^*, P^* > \) does not result in the regulatory target ratio \( r \) in worker \( B \)'s promotion. To achieve the ratio \( r \), by Lemma 13, 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 \( \tau \).

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

\[
\max_{Q, P} \{ E[\pi(\cdot; Q^*, P^*)] - \tau \cdot 1_{(\mu(A_B(Q^*)) \neq r)}, \quad E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})], \quad E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] \} \tag{21}
\]

**Proof.** If \( \chi_B(Q^*) = r \), by Theorem 1, the optimal mechanism is \( < Q^*, P^* > \), and it is supported by (21). 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 \( \tau \). In case of ignoring the rule, by Theorem 1, the best strategy for the owner is choosing \( < Q^*, P^* > \). When the owner decides to follow the rule, by Lemma 6, she chooses between \( < Q^{\lambda=r}, P^{\lambda=r} > \) and \( < Q^{\delta(r)}, P^{\delta(r)} > \). \( \square \)

32
5.1.2 The regulator’s problem

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

\[
\tau \geq E[\pi(\cdot; Q^*, P^*)] - \max\{E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})], E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})]\}. \tag{22}
\]

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

Next, I define unfairness of allocation rules. Among many potential measures, I choose an ordinal measure on dichotomous events. The measure evaluates 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 \( Q \), unfairness of the allocation rule \( Q \) is defined as follows:

\[
\phi(Q) = \mu(x_W > x_B \mid x \in \chi_B(Q)) \cdot \mu(\chi_B(Q)) + \mu(x_B > x_W \mid x \in \chi_W(Q)) \cdot \mu(\chi_W(Q)).
\]

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

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

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

\[
\phi(Q^\lambda) = \mu(x_W > x_B \mid x \in \chi_B(Q^\lambda)) \cdot \mu(\chi_B(Q^\lambda)) + \mu(x_B > x_W \mid x \in \chi_W(Q^\lambda)) \cdot \mu(\chi_W(Q^\lambda))
\]

\[
= \frac{1}{2} \cdot \lambda + \frac{1}{2} \cdot (1 - \lambda) = \frac{1}{2}.
\]

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

$$\phi(Q^{\delta(r)}) = \mu(x_W > x_B \mid x \in \chi_B(Q^{\delta(r)})) \cdot \mu(\chi_B(Q^{\delta(r)}))$$

$$+ \mu(x_B > x_W \mid x \in \chi_W(Q^{\delta(r)})) \cdot \mu(\chi_W(Q^{\delta(r)}))$$

$$= \mu(x_W > x_B \mid x \in \chi_B(Q^{\delta(r)})) \cdot r + \mu(x_B > x_W \mid x \in \chi_W(Q^{\delta(r)})) \cdot (1 - r)$$

$$= \text{pr}(\delta(r) < x_B - x_W < 0) \cdot r + \text{pr}(0 < x_B - x_W < \delta(r)) \cdot (1 - r)$$

$$\leq 0 \cdot r + \frac{1}{2} \cdot (1 - r)$$

$$< \frac{1}{2}.$$

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

$$\phi(Q^{\delta(r)}) = \text{pr}(\delta(r) < x_B - x_W < 0) \cdot r + \text{pr}(0 < x_B - x_W < \delta(r)) \cdot (1 - r)$$

$$\leq \frac{1}{2} \cdot r + 0 \cdot (1 - r)$$

$$< \frac{1}{2}.$$

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

Suppose that the regulator is also interested in minimizing unfairness. In that case, by Lemma 15, given an arbitrary regulatory ratio $r,$ Gap mechanism is preferred to unconditional mechanisms. 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 $E(x_B - x_W \mid x_B > x_W) \geq d$: with $(r, \tau)$ s.t. $r = \frac{1}{2}$ and $\tau \geq E[\pi(\cdot; Q^*, P^*)] = E[\pi(\cdot; Q^{\frac{1}{2}}, P^{\frac{1}{2}})]$ where $E[\pi(\cdot; Q^{\frac{1}{2}}, P^{\frac{1}{2}})] = E(\max\{x_B, x_W\}) - \frac{1}{2} \cdot d.$

**Proof.** First, I derive a condition that the owner chooses Gap mechanism over an unconditional mechanism. $E(\pi_U)$ in (3) can be rewritten as follows:

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

Therefore, $E[\pi(\cdot; Q^{\delta(r)}, P^{\delta(r)})] \geq E[\pi(\cdot; Q^{\lambda=r}, P^{\lambda=r})] \Leftrightarrow E(x_B - x_W \mid x_B - x_W > \delta(r)) \geq$
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 (4). Finally, a condition that the owner follows the regulatory ratio $r$ is given by (22) and (20). □

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. However, Theorem 4 (and Example 5 in appendix B.3) implies that such a policy decision needs attention. Without careful examinations on 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 (Unconditional mechanism, which requires no incentive for the manager) to achieve the regulatory ratio $r$. In that case, high frequency of unfair events (discrimination and reverse discrimination) would occur as Lemma 15 and Lemma 17 show. Therefore, to succeed in quantitative equity ($r \approx \frac{1}{2}$) and to approximate qualitative fairness (minimizing $\phi$), it is necessary for regulators to adjust their objective based on details of a specific organization. □

Furthermore, main results presented in this section (Theorem 3 and Theorem 4) can be applied to a one-sided policy where the punishment $\tau$ 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 where unconditional mechanisms are optimal for the owner. Therefore, regulators who concern a criticism of reverse discrimination should not consider a one-sided policy for their best option. It just directs a lower bound of the regulatory ratio, and compared to the two-sided policy, it does not produce any other different positive effects.

\[15\text{I.e., If conditions in Theorem 4 are not feasible, then set } r \text{ with the regulatory ratio’s upper bound that induces Gap mechanism to be optimal for the owner.}\]
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 at the middle of the hierarchy, whereas in Becker (1957), it is assumed that the owner (or the whole organization) has such discriminatory taste. In the sense that both models assume taste-based discrimination, the model of this paper can be seen as modification of Becker’s model with new features: (a) existence of another principal above the labor-decision maker in a hierarchical setup, (b) the principal’s incomplete information on the labor-decision maker’s discriminatory preference and on 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 $< Q^*, P^* >$ is equivalent to the outcome of Becker’s model where 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 4 and Example 4), for any given $\nu(d) \in (0, 1)$, $< Q^*, P^* >$ is not an optimal mechanism any more. Furthermore, the owner’s optimal incentive arrangement is contingent on the probability of the manager being discriminatory. This implies that discrimination in large (multiple hierarchical) organizations is completely different from discrimination in small organizations described in Becker’s model in terms of the treatments to be given.

6.2 Identity-based affirmative action

To encourage diversity and fairness, affirmative action based on identities (race, gender, the disabled or mixture of those features) has been employed in many cases, e.g., college

---

16 As a result, in Becker (1957), the owner perceives the discrimination coefficient as a part of production costs.
17 It could be one of the three—Gap, B-bar, W-bar—projection mechanisms or Delegation mechanism.
18 In small organizations (the two-level hierarchy), the principal directly interacts with subordinate workers and makes labor related decisions.
admissions or employment in the workplace. Concurrently, there have been debates on the effectiveness of the identity-based affirmative action. The main argument opposing identity-based affirmative action is that it creates another unfair competition for those who are not included in the category of the affirmative action: that is, existence of reverse discrimination. However, results from section 4 and section 5 show that such argument is not necessarily true especially when there is an enough chance that the decision maker is biased. The profit maximizing mechanism in that case does not completely mitigate the discriminatory outcome.\textsuperscript{19} Rather, by selecting a proper quota level (the level that induces Gap projection mechanism to be optimal), the benefit goes to the competent minorities, who should have been promoted if the manager were fair. Therefore, under the circumstances where bias exists, the identity-based affirmative action helps the capable workers to be chosen regardless of their identities.

Another implication of this paper on affirmative action is the equivalence between quotas and preferential treatments (boosting scores for those in affirmative action category)\textsuperscript{20}, although the U.S. courts have treated them differently: e.g., in \textit{Grutter v. Bollinger}, 539 U.S. 306 (2003), the U.S. Supreme Court justices upheld the University of Michigan’s law school affirmative action policy that awarded extra points for blacks, Hispanics and Native Americans; in \textit{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, Davis School of Medicine were illegal.\textsuperscript{21} 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 unique productivity leverage preferable for $B$ that creates the equivalent outcome to the racial quota\textsuperscript{22}: under the optimal mechanism, the owner achieves the specific targeted quota by considering $B$’s productivity with the extra boosting treatment. Consequently, the preferential treatment policy and the quota policy are actually equivalent in terms of the effects. Therefore, it is reasonable that both methods in affirmative action do not receive the opposite legal judgements.

\textsuperscript{19}See Corollary 1 in subsection 4.1 and Corollary 3 in appendix B.3.
\textsuperscript{20}Fryer (2009) and Bodoh-Creed and Hickman (2015) also show such equivalence results in environments of the two-level hierarchy and a contest, respectively.
\textsuperscript{21}Otherwise, the quota policy for minorities has been lawfully and actively employed in many European countries.
\textsuperscript{22}$|d - \delta(r)|$ in Lemma 13 of subsection 5.1 and $|a^*_p - a_P(r)|$ in Lemma 16 of appendix B.3
6.3 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 narrowed down to a lower-dimensional problem. That is, summary information is not enough, because each of the reports plays its own role for the principal to understand the given screening problem and derive the optimal contract. The productivity levels of $B$ and $W$, not the difference between them, provide a way to find out the detectable lie, a status that the final output (the owner’s updated informational state $x_{Q(z)}$) is different from what the agent reported ($z_{Q(z)}$). Also, 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, the exact numbers of both worker’s productivity are crucial information to the principal. Furthermore, as shown in subsection 4.2, the manager’s discrimination type is a main element for constructing the projection mechanism that is the optimal mechanism when the probability of the manager being discriminatory is fair enough (Theorem 4 and Example 4). Hence, without the information on the manager’s type, the optimal incentive scheme cannot be obtained.

7 Conclusion

In contrast to the pre-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 the standard of equity in the broad society. The optimal mechanism for the owner to adopt provides an incentive to the discriminatory manager for promoting a minority worker. It yields benefits in terms of both reducing discrimination and increasing profit of the organization. The paper also shows how a regulator can improve compliance with non-discriminatory conduct.
The model studied here can be extended by adding choice problems of the agent’s subordinates (e.g., effort levels); such a model will incorporate statistical discrimination naturally. The subordinates’ decisions on their effort levels affect their principals (the manager and the owner)’ beliefs on their productivity: 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: we can see how the minority and non-minority workers strategically behave for a trade-off between their labor cost and the promotion opportunity when they acknowledge the fact that the organization’s owner tries to restrict the discriminatory manager’s discretion. Accordingly, this extension creates another principal-agent relationship between the manager and the subordinate workers, which includes the workers’ strategies. Therefore, we can formulate other competition environments for the subordinates. For example, we can consider a contest in which the manager wants to design the contest favorably to whom he prefers, and the owner tries to provide incentives to him for designing 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 (s.t. the agent advocates 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 people’s behavior could be conditional on additional factors such as bases of the discriminatory preference (e.g., race, gender, and individual favoritism). Hence, experiments testing those features will be a good complement to the theoretical results in this paper.

Appendix A: On the Legal Status of Optimal Mechanisms

This appendix reviews the U.S. law of discrimination and analyzes statutory and jurisprudential issues regarding the optimal mechanisms discussed in the paper. 

This appendix has been written based on my understanding of the U.S. law regarding discrimination and affirmative action, and it is especially tailored for the environment of this paper. It should not be
A.1 Statutes about discrimination and affirmative action

A.1.1 The core statutes

The Civil Rights Act of 1964 is the main 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.

2. Education:

   ... Title IV of the Civil Rights Act of 1964 ... prohibits discrimination on the basis of race, color, national origin, sex, and religion in public schools and institutions of higher learning.

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.

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)

used to provide legal advice. Anyone seeking legal rights should consult with legal counsel. All errors are mine.

25 https://www.justice.gov/crt/educational-opportunities-section
Finally, contractors of the federal government are required to employ affirmative action for under-represented minorities and women.  

Since the seminal case of United Steelworkers v. Weber, where the Supreme Court of the United States upheld voluntary affirmative action in private workplaces, the following three conditions from the case became legal standards for 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.  

A.2 Legal cases in the U.S. and the implications

The U.S. Supreme Court clearly acknowledges necessity of affirmative action in certain environments. However, some particular affirmative action plans were rejected by the court, and the following provides the court’s stance on specific affirmative action plans.

- Layoff or replacement of non-beneficiaries trammel their rights.

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

- Preferential treatment can be used: different cutoff levels are not allowed, but demographic identities can be used as an additional point category. Banding (e.g. test scores are categorized by ranges and compared by the category) might be allowed, but point boosting (e.g. giving extra test scores to minority candidates) is not allowed.

---

28 Presidential Executive Order 11246.
30 Lareau (2016).
32 In Regents of the University of California v. Bakke, 438 U.S. 265 (1978), the Court ruled that quotas used in the school admission was not lawful, but it upheld affirmative action where race is considered as one of factors in school admission decisions; in United States v. Paradise, 480 U.S. 149 (1987), quota-based affirmative action was ordered by the Court.
Improving diversity can be part of the goals of educational institutions. However, in workplaces, a justification of an operational need for diversity is limited without evidence of past discrimination.

A.3 Implementation of the optimal mechanism

Suppose the following case. The owner of a firm believes that based on the past and current imbalance on promotion results, it is highly probable that the 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 in the context of the preceding discussion and study how to lawfully implement the mechanism’s outcome.

A.3.1 Features of the Gap Projection Mechanism

Main features in the specification of the Gap Projection Mechanism are as follows:

1. It is a direct mechanism. Communication between the owner and the manager takes place 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.

   (a) If the manager reports that he is fair, the owner selects B when B’s reported productivity is higher than threshold $\delta$, and pays $d$ to the manager. When B’s reported productivity is less than threshold $\delta$, a worker with higher reported productivity is promoted, and the manager does not get 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 than threshold $\delta$ and pays $d$ to the manager. Otherwise, W is promoted and the manager does not get any bonus.

The Gap Projection Mechanism described above has affirmative action components and ameliorates the discriminatory outcome of the status quo: it provides a bonus to

---

35 See §78.05[4] and §78.12[e] in [Lareau (2016)].
36 Note that if the owner finds out a detectable lie, then the payment to the manager is always 0.
the manager for promoting $B$ when conditions are met, and compared to the status quo, it increases a promotion ratio of $B$ when the manager is discriminatory in its truthful equilibrium outcome.

However, it can be also claimed as discrimination because the promotion rule specifies that it is not sufficient for $B$ having higher productivity than $W$ to be promoted when the manager is discriminatory: $B$ needs to be better than $W$ at least as much as $\delta$. Therefore, to implement the Gap Projection Mechanism in practice as affirmative action, it is crucial to remove such a discriminatory feature.

A.3.2 Alternative 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 possibilities of criminal prosecution or lawsuits from implementing the mechanism to avoid penalties and future litigation costs. In the EEO (Equal Employment Opportunity) law, such incentives mainly have to do with civil, not criminal litigation.

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 always to make personnel decisions in a completely lawful way. In fact, from the analysis in this paper, only the unconditional mechanism can succeed in that respect among incentive compatible direct mechanisms. However, the unconditional mechanism blatantly disregards the spirit of the EEO law, as is shown by the fact that, in equilibrium, it has the highest incidence of unfair promotions among incentive compatible direct mechanisms. Correspondingly, that mechanism is unprofitable, since it does not optimize the efficiency of promoted workers. Thus, justifiably, firms’ owners will choose the Gap Projection Mechanism over the unconditional mechanism, and will take on 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 costs and civil penalties that might arise from this well-intentioned conduct, which elevates the level of fair promotions compared to the status quo in an effort to maximize profit.

The following alternative indirect mechanism generates the equilibrium outcome of the Gap Projection Mechanism but is a litigation proof way of providing a same
promotion rule for $B$ and $W$ and a same payment rule for both types of the manager.\footnote{Fulfillment of the Weber test is also discussed in the next section.}

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

A2. A 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 $\delta$.

Under the alternative mechanism, the truthful reporting equilibrium outcome of the Gap Projection Mechanism can be obtained with an untruthful reporting equilibrium of the alternative mechanism. It can be checked as follows:

If the manager is fair, he reports productivity information truthfully only if $x_B < \delta$. If $x_B > \delta$, he always reports $[z_B > z_W \text{ s.t. } z_B = x_B]$ to earn the bonus $d$ regardless of the true productivity difference $x_B - x_W$. If he is discriminatory, he reports productivity values with $[z_B > z_W \text{ s.t. } z_B = x_B]$ when the true productivity gap exceeds $\delta$ ($x_B - x_W > \delta$), \footnote{The discriminatory manager is indifferent between the two promotion results if $x_B > \delta$. Therefore, any allocation is supported in this case.} and reports $[z_W > z_B \text{ s.t. } z_W = x_W]$ if $x_B < \delta$.

Additionally, note that the bonus scheme in the mechanism (monetary rewards to the manager for promoting underrepresented groups) has been adopted in some private and public institutions in the U.S. as part of affirmative action plans.\footnote{“... 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.”, Coca-Cola Task Force Report 2004, pp.42.}

A.3.3 Weber test

If the owner’s belief on the manager’s discriminatory type is a function of a past imbalance in promotion results and if the probability that he is discriminatory is high enough, then adopting the alternative Gap Projection Mechanism as the optimal mechanism passes the Weber criteria 1 (evidence of discrimination). The Weber criteria 2 (necessity of trammeling the rights of non-beneficiaries) is also satisfied as the mechanism’s allocation rule is non-discriminatory to both workers, and the mechanism’s setup

\footnote{“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.}
is about promotion, not replacement of the non-beneficiaries.\footnote{The original Gap Projection Mechanism also passes the criteria 2. In promotion or hiring, considering a minority candidate first if the person is qualified has been justified as affirmative action in the Court. Under the Gap Projection Mechanism, $W$ is non-beneficiary as 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.} Finally, it passes the Weber criteria 3 (non-permanent policy) as the mechanism is not employed as the optimal mechanism if discrimination does not exist in the institution.\footnote{"... 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 \cite{Lareau}.}

**Appendix B: Proofs and Examples**

**B.1 Projection Mechanisms (continued)**

In this section, I present expected utility functions of the three (B-bar, W-bar, Gap) projection mechanisms discussed in Example \footnote{... 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 \cite{Lareau}.}

\[
E[\pi(\cdot; Q^c(X_B^d), P^c)] = \nu(d) \cdot [E(x_B - d|X_B^d) \cdot \mu(X_B^d) + E(x_W|\mu(X_B^d)) \cdot \mu((X_B^d)^c)] + (1 - \nu(d)) \cdot [E(x_B - d|\mu(X_B)) \cdot \mu(\mu(X_B)) + E(\mu(x_B, x_W)|\mu(X_B)^c) \cdot \mu(\mu(X_B)^c)].
\]

Note that $\forall i \in \{B, W\}$, $x_i \sim \text{Uniform}[0, 1]$: $F_i(x_i) = x_i$ 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 $\bar{X} = [a, 1] \times [0, 1]$

\[
E[\pi(\cdot; Q^c([a, 1] \times [0, 1]), P^c)] = \nu(d) \cdot \left(\int_a^1 x_B \cdot \frac{1}{1-a} dF_B - d\right) \cdot (1 - F(a, 1)) + \int_0^1 x_W dF_W \cdot F(a, 1) + (1 - \nu(d)) \cdot \left(\int_a^1 x_B \cdot \frac{1}{1-a} dF_B - d\right) \cdot (1 - F(a, 1)) + (\int_0^1 x_1 dF_1) \cdot F(a, 1)
\]

where $F_1(x_1) = F_B(x_1|x_B \leq a) \cdot F_W(x_1)$.

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

\[
E[\pi(\cdot; Q^e([0, 1] \times [0, a]), P^e)] \\
= \nu(d) \cdot \left((\int_0^1 x_B \ dF_B - d) \cdot F(1, a) + \int_a^1 x_W \cdot \frac{1}{1-a} \ dF_W \cdot (1 - F(1, a))\right) \tag{25} \\
+ \ (1 - \nu(d)) \cdot \left((\int_0^1 x_B \ dF_B - d) \cdot F(1,1)\right)
\]

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

\[
E[\pi(\cdot; Q^e(\{x|x_B - x_W > a\}), P^e)] \\
= \nu(d) \cdot \left((\int_0^1 \int_{x_B + a}^1 x_B \ dF_B dF_W - d) \cdot (\int_0^1 \int_{x_W + a}^1 dF_B dF_W) \right) \\
+ \left((\int_0^1 \int_{x_B - a}^1 x_W \ dF_W dF_B) \cdot (\int_0^1 \int_{x_B - a}^1 dF_W dF_B)\right) \tag{26} \\
+ \ (1 - \nu(d)) \cdot \left((\int_a^1 x_B \cdot \frac{1}{1-a} \ dF_B - d) \cdot (1 - F(a,1)) + (\int_0^1 x_1 \ dF_1) \cdot F(a,1)\right)
\]

where \( F_1(x_1) = F_B(x_1|x_B \leq a) \cdot F_W(x_1) \).

B.2 Figures for Example 3

![Figure 3: Expected Profit of Delegation mechanism](image)
Figure 4: B-bar projection mechanism: $X_B^d = [a, 1] \times [0, 1]$
Figure 5: W-bar projection mechanism: $X_B^\dagger = [0, 1] \times [a, 1]$
Figure 6: Gap mechanism: $X^d_B = \{z | z_B - z_W > a\}$
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 (8) 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 to the assumption s.t. $\forall t \in \{0, d\}$, $\chi_B^t(Q) = \{x|x_B > x_W\}$. \(\square\)

Note that the promotion ratio of $B$ with a mechanism $< Q, P >$ is $\mu(\chi_B(Q)) = \nu(d) \cdot \mu(\chi_B^d(Q)) + (1- \nu(d)) \cdot \mu(\chi_B^0(Q))$. Also, 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 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. $r = \mu(\chi_B(Q^c(X_B^d(a_P))))$ 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, $\mu(\chi_B(Q^c_b(X_B^d))))$ is a strictly increasing function w.r.t. $\mu(X_B^d)$ as follows.

\[
\mu(\chi_B(Q^c_b(X_B^d(a_P)))) = \nu(d) \cdot \mu(X_B^d(a_P)) \cdot 1 + (1- \nu(d)) \cdot [\mu(X_B) + (1- \mu(X_B)) \cdot \text{pr}(x_B > x_W | (X_B)^c)]
\]

As $F_i$ is a continuous cumulative distribution function, given $P$, (27) is continuous in each $a_P$. Also,

\[
\lim_{a_W \to 0} \mu(\chi_B(Q^c_b(X_B^d(a_W)))) = 1;
\]

\[
\lim_{a_B \to 0} \mu(\chi_B(Q^c_b(X_B^d(a_B)))) = 1;
\]

\[
\lim_{a_G \to 0} \mu(\chi_B(Q^c_b(X_B^d(a_G)))) = 1.
\]

Therefore, there exists $a(r)$ for Gap, B-bar, and W-bar projection mechanisms achieving $r = \mu(\chi_B(Q^c(X_B^d(a(r))))$ where $r \in (\mu(\chi_B(Q^c(X_B^d(a_P^*))), 1)$. \(\square\)
Corollary 4. A regulator can implement a specific target ratio $r \in [0, 1]$.

Proof. Let $< Q^{**}, P^{**} > = \arg \max_{Q, P} \{ E[\pi(\cdot; Q^0, P^0)] \}$, $\max_{a_p} \{ E[\pi(\cdot; Q^c(X_B^d(a_p)), P^c)] \}$.

By Theorem 2 and Lemma 14, given regulation $(r, \tau)$, the owner’s optimization problem is as follows:

$$\max_{Q, P} \left\{ E[\pi(\cdot; Q^{**}, P^{**})] - \tau \cdot 1_{(\mu(x_B(Q^{**})) \neq r)} \right\}, \max_{a_p(\tau)} \left\{ E[\pi(\cdot; Q^{c}(X_B^d(a_P(r))), P^c)] \right\}$$

(28)

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

$$\tau \geq E[\pi(\cdot; Q^{**}, P^{**})] - \max_{a_p(\tau)} \left\{ E[\pi(\cdot; Q^{c}(X_B^d(a_P(r))), P^c)] \right\}. \ \Box$$

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

$$\phi(Q) = \sum_{t \in \{0, \bar{d}\}} \nu(t) \cdot [\mu(x_W > x_B | x \in \chi^t_B(Q)) \cdot \mu(\chi^t_B(Q)) + \mu(x_B > x_W | x \in \chi^t_W(Q)) \cdot \mu(\chi^t_W(Q))]$$

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 $a_p \in (0, \bar{c}) \times \{G, B, W\}$,

$$\max_{a_p} \left\{ \max_{d} \{ \phi(Q^c(X_B^d(a_p))) \} \right\} < \phi(Q^{\lambda=r}).$$

Proof. First, note that by Lemma 13, $\phi(Q^{\lambda=r}) = \frac{1}{2}$. I show that $\phi(Q^0) < \frac{1}{2}$ and $\forall a_p \in (0, \bar{c}) \times \{G, B, W\}$, $\phi(Q^c(X_B^d(a_p))) < \frac{1}{2}$.

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

Since $\nu(d) < \frac{1}{2}$, $\phi(Q^0) < \frac{1}{2}$. Also, $\forall X_B^d = X_B^d(a_p)$, $\nu(x_W > x_B | X_B^d) < \frac{1}{2}$, $\nu(x_B > x_W | X_B^d) < \frac{1}{2}$, $\nu(x_H > x_B | X_B^d) < \frac{1}{2}$, $\nu(x_W > x_H | X_B^d) < \frac{1}{2}$, $\nu(x_H > x_W | X_B^d) < \frac{1}{2}$.
\[ x_W | (X_B^d)^c) < \frac{1}{2}, \text{ and } \Pr(x_W > x_B | X_B) \leq \frac{1}{2}. \text{ Therefore, } \phi(Q^c(X_B^d)) < \frac{1}{2}. \] □

**Example 5.** There exists an environment \((d, F, \nu)\) 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 \(d = 0.5, \forall i \in \{B, W\}, x_i \sim \text{Uniform}[0, 1], \text{ and } \nu(d) = 0.9.\) Suppose that a regulator wants to implement \(r = \frac{1}{2}\). Note that \(E[\pi(\cdot; Q^\lambda=\frac{1}{2}, P^\lambda=\frac{1}{2})] = 0.5, \text{ and } \mu(\chi_B(Q^0)) = 0.05.\) Each projection mechanisms (Gap, B-bar, W-bar) can implement \(r = 0.5\) by setting \(a_G(\frac{1}{2}) = 0.051, a_B(\frac{1}{2}) = 0.513, \text{ and } a_W(\frac{1}{2}) = 0.444; \) and corresponding expected profits are \(E[\pi(\cdot; Q^c(X_B^d(a_G(\frac{1}{2}))), P^c)] = 0.401, \ E[\pi(\cdot; Q^c(X_B^d(a_B(\frac{1}{2}))), P^c)] = 0.383, \text{ and } E[\pi(\cdot; Q^c(X_B^d(a_W(\frac{1}{2}))), P^c)] = 0.361. \) □
References


