An economic theory of differential treatment

Emil Temnyalov*

October 26, 2019.

Abstract

When does efficiency justify affirmative action and gender equity policies in education and labor markets? Or, more generally, when does efficiency require differential treatment based on observable and surplus-irrelevant characteristics, such as race, gender, or socioeconomic status?

This paper proposes an assignment model of differential treatment, where a policymaker assigns agents to different treatments or positions in order to maximize total surplus, based on the agents' characteristics and on noisy information about their types (i.e. abilities or productivities). I provide necessary and sufficient conditions on the agents' signaling structures which characterize whether surplus maximization requires differential treatment or not, in a general non-parametric information economics framework. I show that under certain reasonable conditions the optimal assignment policy is characterized by an index which measures the agents' expected marginal benefits from different treatments, and also examine further conditions on the bias and informativeness of signaling structures that determine the efficiency implications of differential treatment. The model also provides novel questions and predictions for empirical research on the economics of discrimination.

Keywords: differential treatment, affirmative action, gender equity, optimal assignment, signalling, efficiency, education, labor markets. **JEL codes:** D47, D61, D8, J7, I24.

*University of Technology Sydney, Department of Economics. emil.temnyalov@uts.edu.au.
For helpful comments I thank Dan Bernhard, Juan Carlos Carbajal, Federico Echenique, Jeffrey Ely, Young-Chul Kim, Bettina Klaus, Hao Li, George Mailath, Thomas Mariotti, Shiko Maruyama, Alessandro Pavan, Bruno Strulovici, Kentaro Tomoeda, Steven Williams, and seminar participants at Caltech, Collegio Carlo Alberto, Simon Fraser, Toulouse School of Economics, the Universities of British Columbia, Illinois Urbana-Champaign and New South Wales, and conference participants at AETW 2019, NASM ES 2019, PET 2019, and the Organisational Economics Workshop 2019.

1 Introduction

Affirmative action policies are often controversial in debates about public policy, education, hiring and promotions within organizations, and in other examples of differential treatment, whereby economic agents are treated differently depending on some observable characteristics. For example, in an education context, whether a student is admitted to a university or not may depend on their socio-economic status, race, or some other characteristic of the student or of their environment—an issue currently under scrutiny in the "Students for Fair Admissions v. Harvard" lawsuit, for instance. Similarly, in a labor market context, an organization's promotion or hiring decisions may depend not just on the performance of an employee, but also on characteristics such as the opportunities that the employee has had, their gender, or some other characteristic of the employee or of their environment—a topic of discussion in the ongoing global debate on the gender gap in employment. Debates about such differential treatment policies are often based on concerns about fairness, and one can make ethical arguments both in favor and against differential treatment. This paper offers a different approach and considers differential treatment from the perspective of efficiency, rather than fairness. Fairness is of course an important concern, but understanding the positive implications of affirmative action and other differential treatment policies, before debating their merits on normative grounds, may provide more constructive grounds for debate. It may also inform public policy and provide new conceptual questions for empirical research on the economics of discrimination.

I study an assignment problem where a policy-maker must assign agents to different treatments or positions, based on some observable signals and characteristics of the agents. The policy-maker's objective is to design an assignment policy which maximizes total expected surplus, subject to an ex post feasibility constraint—i.e. the number of agents assigned to each position must not exceed the capacity of that position. I highlight two leading examples from education and labor contexts: a university's admissions office must decide which students to admit or not; and an organization must decide which job candidates to hire, or how to promote employees to different ranks within the organizational hierarchy.

Each agent generates some surplus or value (e.g. student achievement or employee output), which depends on their unobservable type (e.g. the ability of a student or the productivity of an employee) and on the treatment or position they are assigned to. The policy-maker observes a noisy signal of the agent's type (e.g. a student's standardized test score or an employee's past performance), and also some characteristics of the agent (e.g. a student's socio-economic status, race, gender, etc., or an employee's demographics, gender, etc.)

This paper characterizes the surplus-maximizing policy in a general setting, under mild non-parametric assumptions on values and signal distributions. First, I show that the optimal assignment policy will generically feature differential treatment across the observable characteristics of the agents, even if these characteristics are not directly payoff relevant, as long as the distributions of signals vary across these observable characteristics. More precisely, the optimal assignment policy is assortative with respect to a *benefit index*, which measures the expected incremental benefit from being assigned to a higher position or treatment. But the optimal policy need not be monotonic with respect to signals, because signals need not monotonically translate into expected incremental benefits, if the conditional distribution of types, given signals, differ across characteristics. The optimal policy is deterministic (except for tie-breaking among payoff-equivalent agents) and can be implemented with category-specific thresholds, i.e. signal cutoffs which generally vary across agents' characteristics. In fact, I show that the optimal policy is only assortative with respect to signals, for any capacities, if and only if the conditional signal distributions are identical across observable characteristics in a particular sense. This is a very strong condition, which suggests that in many realistic settings the surplus-maximizing policy will feature some differential treatment, which favors agents whose signal distributions are "worse"—i.e. a form of affirmative action.

Second, I show that the optimal policy may also feature differential treatment even in the case where signals are unbiased predictors of agents' types, i.e. an agent's expected type, conditional on a signal, does not vary across other observable characteristics. This is because the policy-maker's objective function may feature an intrinsic preference for or against dispersion in types, and the conditional type distributions can be noisier or less noisy across different characteristics, even if expected types are equal across characteristics. That is, some categories of agents may have more or less informative signals, defined in terms of mean-preserving spreads in conditional type distributions, and as a result may optimally be treated differently, depending on the shape of the surplus function. Hence the optimal policy is monotonic with respect to the benefit index, but need not be monotonic with respect to the agents' expected types, and may favor agents whose type distributions are more or less dispersed—i.e. another form of affirmative action.

To derive my main results, outlined above, I provide a definition of *comparability* across categories, which requires that the expected incremental benefits from any two positions, as a function of an agent's signal, can be translated across characteristics. This notion allows signals to be compared across different characteristics, in an environment with any signal distributions. Comparability implies an ordering over type distributions, which is more general than the usual stochastic orderings used in economics, since it only imposes restrictions on the expected values of the conditional type distributions, rather than on the shape of the distributions. Moreover, I show that comparability is satisfied in a variety of environments. I provide 3 different sufficient conditions for the notion of comparability to hold: either (i) treatments are binary; or (ii) the surplus is multiplicatively separable in type and position; or (iii) the family of signal distributions are similar, in a particular sense. Each of these conditions is enough to ensure comparability, making it a widely applicable concept.

My main results can be interpreted for example in the context of university admissions policies. One may think that the distribution of students' university exam scores, conditional on their true abilities, varies across socio-economic status or some other observable characteristic (e.g. because parents invest different time and resources in their children's education, depending on their status). A university admissions office, seeking to maximize surplus, would therefore use socio-economic status information in forming beliefs about the students' true abilities. My first main result shows that the optimal admissions policy may admit students non-monotonically with respect to their exam scores—a lower-score student from a socio-economic group whose score distributions are worse may be preferred over a higher-score student from a socio-economic group with higher score distributions. This highlights a *nominal* motive for differential treatment, which results from the fact that surplus is a function of the agents' types, not of their signals per se. Moreover, the optimal policy may favor some group even if expected abilities are the same across groups, conditional on scores, for example if the variance of abilities is different across characteristics and if there are increasing or decreasing marginal values from higher types (e.g. if the value of "superstar" students is disproportionally high). This highlights a *substantive* motive for differential treatment, which results from the fact that the informativeness of signal distributions can vary across characteristics and thus affect the dispersion of types across characteristics. Such nominal and substantive motives may arise in the context of hiring and promotions decisions within organizations, for example in thinking about differential treatment based on gender.

The paper also highlights several important observations for public policy debates and empirical research on the topic of affirmative action and gender equity in education and labor market settings. First, differential treatment may arise out of concerns for economic efficiency, rather than fairness, which is often presumed in such debates. Second, comparing signals (e.g. university exam scores or past employee performance) may be asking a misguided question, as it focuses on the wrong metric. In the context of maximizing surplus, it is expected incremental benefits that matter, not signals per se. Therefore a more relevant empirical question would be to estimate the returns to education across different characteristics for the marginal admitted students at a university, or the productivity across different characteristics for the marginal employees at an organization. If these are consistent across characteristics, this would point to an efficiency rationale for differential treatment. Third, comparing expected abilities or productivities across characteristics, conditional on exam scores or past performance, may also be asking the wrong questions, as the optimal policy can depend on the dispersion of types, not just on means, across characteristics. So a more relevant empirical question would be to estimate whether there is inherently a disproportional benefit or loss due to dispersion in the abilities of students or in the productivity of employees. If so, this would also point to an efficiency rationale for differential treatment.

In addition to my main results, I also discuss more specific applications of the model. First, I consider an environment with biased signals, where the distributions of types, conditional on any signal, can be ordered across characteristics in terms of first-order stochastic dominance. In this application signals over-estimate types for higher categories and under-estimate them for lower ones. Such a model may accurately describe the relationship between student abilities, standardized test scores, and socio-economic status (with higher socio-economic status corresponding to higher categories). The optimal policy here favors agents from categories with worse signals, and the optimal cutoffs for each position are increasing across categories—lower categories have lower signal cutoffs.

Next, I consider an environment with unbiased but noisier signals, where type distributions, conditional on scores, are mean-preserving spreads across categories, and the value function features a preference for or against dispersion. I show the optimal policy implies that the aggregate treatments will vary across categories, even though the unconditional type distributions are the same a priori. Hence an environment with ex ante symmetric types would endogenously give rise to aggregate inequality of treatments. For example, if the policy-maker has a preference for dispersion, then agents from categories with noisier type distributions will be favored for each position, creating aggregate inequality.

2 Literature

This paper relates to a growing literature which studies the effect and implementation of diversity and distributional objectives in matching and assignment models. Several recent papers consider affirmative action and diversity aspects in school choice and in contests. Abdulkadiroğlu and Sŏnmez (2003) study a model of controlled school choice, introducing type-specific quotas motivated by diversity concerns into a matching framework. Abdulkadiroğlu (2005) considers a two-sided matching setting and studies the existence of stable matching mechanisms with and without affirmative action. Kojima (2012) shows that quota- and priority-based affirmative action constraints may, in some cases, hurt all members of the targeted group and lead to Pareto worse outcome in any stable matching mechanism. Hafalir, Yenmez and Yildirim (2013) study the welfare effect of minority reserves, which are one possible implementation of affirmative action policies. Ehlers, Hafalir, Yenmez and Yildirim (2014) introduce a matching model of control constraints as soft bounds, and study fairness and welfare. Hafalir, Kojima and Yenmez (2019) develop a framework to study inter-district school choice, whereby students can be assigned to schools across districts, with one of the main policy objectives being to balance diversity, and study the diversity and efficiency properties of deferred acceptance and top trading cycles mechanisms. Fu (2007) models college admissions in an asymmetric all-pay auction model with complete information, where maximizing student efforts requires using a handicap that resembles affirmative action. This paper contributes to the literature by studying differential treatment, a broader class of policies than affirmative action, in the context of an assignment model. Moreover, this model explicitly makes a distinction between agents' payoff-relevant types and their signals, whereas the matching literature typically interprets signals as types. The latter distinction turns out to be critical for studying the efficiency implications of differential treatment.

More broadly, a large economics literature studies statistical discrimination, including papers by Phelps (1972) and Arrow (1973), which offered an alternative to the model of taste-based discrimination by Becker (1957). These seminal papers created two strands of literature: one where there are exogenous differences between different groups of agents, giving rise to discrimination; and another where differences between groups of agents arise endogenously in equilibrium, even though groups are ex ante identical. In the former strand, Chambers and Echenique (2019) study the connection between statistical discrimination and the statistical identification of workers' signals from skills. Lundberg and Startz (1983) study human capital investment in a model where different groups have more or less noisy signals of productivity. Moro and Norman (2003b) study human capital in a model with heterogeneous investment costs across groups. Cornell and Welch (1996) study labor tournaments where groups have access to more or fewer signals of their productivity, and the scarcity of jobs leads to differential treatment across groups. In the latter strand, Coate and Loury (1993) study noisily observed skill investments by workers, which lead to a multiplicity of equilibria, whereby some groups may invest more than others and thus endogenously be favored by employers, even though groups are ex ante identical. Several papers study discrimination which arises with interactions between groups. Moro and Norman (2004) study a model of human capital investment with endogenous wages, where groups may in equilibrium specialize in different jobs, leading to asymmetric outcomes that do not rely on a coordination failure within groups. Mailath, Samuelson and Shaked (2000) study discrimination that arises due to search frictions, even with perfect information about agents' types. These strands of the literature are summarized in a comprehensive survey by Fang and Moro (2011). This paper contributes to the literature in three main aspects. First, I focus on the problem of designing assignment policies which maximize efficiency, rather than on explaining discrimination per se, which is the main aim of statistical discrimination theories. I take differences in signalling technologies as given, and show that affirmative action may be an optimal policy response to differences across groups. Hence this paper provides an efficiency rationale for differential treatment.[1] Second, I study a general assignment model, with arbitrary treatments or positions, without making any specific parametric assumptions about the distributions of types and signals or the functional form of payoffs, in contrast to the existing literature on discrimination. I provide general results using tools from information economics, monotone comparative statics, and stochastic ordering theory. Third, in an application of my model I show that aggregate inequality across groups can arise even in the absence of human capital investment, search frictions, coordination failure, or interactions across groups, as a result of purely informational frictions.

A subset of the discrimination literature considers affirmative action as a possible policy remedy to discrimination. This literature generally finds that affirmative action may exacerbate differences across groups, for example in terms of human capital, and may lead to lower welfare (Coate and Loury, 1993; Moro and Norman, 2003b, 2004; Norman, 2003; Fang

[1] For example, the literature on labor discrimination seeks to explain why an employer may favor one group of workers over another, whereas I consider how an employer who understands that some groups have less favorable signalling technologies may optimally treat workers in a way that compensates for such differences.

and Norman, 2006). It is important to note, however, that affirmative action is distinct from merely banning statistical discrimination, which much of the literature focuses on. I consider a richer model of differential treatment, which applies to more general policies, beyond affirmative action or banning or allowing statistical discrimination. Differential treatment more generally considers the possibility of treating groups differently, hence affirmative action may itself entail a form of discrimination.

Other models of the effect of affirmative action include Chung (2000), who provides a theory of role models. Chan and Eyster (2003) study color-blind affirmative action in a university admissions model, where signals are directly payoff-relevant and the university has an explicit preference for diversity. They show that the optimal color-blind admissions policy may involve randomization, which allows the university to balance its preferences for diversity with its preference for higher-score students. Epple, Romano and Sieg (2008) study a related model with competing colleges which have a preference for diversity. Fryer and Loury (2007) consider a similar model with an added ex ante investment stage, where groups have exogenous differences in skill investment costs, and analyze the optimal timing of interventions. In contrast to this literature, the policy-maker in my model has no preference for diversity, and only seeks to maximize expected total surplus. Moreover, signals are distinct from types, and only the latter are directly payoff-relevant. Finally, I do not assume any a priori differences in the distributions of types across groups, and differences only arise due to signalling.

3 Model

A policy-maker wants to assign agents to treatments or positions. There is a set of agents N with cardinality n, and an ordered set of positions $P = (p_1, ..., p_m)$ with corresponding capacities $(k_1, ..., k_m) > 0$, where $\sum_j k_j \geq n$.[2]

Each agent has an unknown type $t_i \in T$, drawn from a distribution F, and a publicly observed category or characteristic $x_i \in X$. The policy-maker observes a noisy signal of each agent's type, $s_i \in S$, drawn from a distribution F_{t_i,x_i}. I assume S is an ordered convex set, T is an ordered set with discrete or continuous values, and X is any set of (possibly vector-valued) characteristics with discrete or continuous values.

[2]This is without loss of generality, in the sense that position p_1 can be a "null" position, with unlimited capacity, reflecting the possibility that agents are not assigned to any position.

A policy is a stochastic assignment of agents, mapping observables to positions, denoted by a function $\mathscr{P} : (X \times S)^n \to \Delta(P^n)$. I denote by $\mathbb{P}(p|x_i, s_i)$ the distribution of agent i's position, and by $p(x_i, s_i)$ the realized position. Every possible assignment generates some individual ex post surplus, given by $v : T \times P \to \mathbb{R}$, where v is increasing in both arguments.[3]

The policy-maker's objective is to design a policy that maximizes the expected total surplus from the assignment, subject to ex post feasibility:

$$\max_{\mathscr{P}} \quad \sum_i \mathbb{E}[v(t_i, p(x_i, s_i))] \quad = \quad \sum_i \sum_j \int_T v(t_i, p_j) dF(t_i|x_i, s_i) \cdot \mathbb{P}(p_j|x_i, s_i) \tag{1}$$
$$\text{s.t.} \quad |\{i : p(x_i, s_i) = p_j\}| \leq k_j$$

I make three standard assumptions to analyze this problem.

Assumption 1 (Supermodularity)**.** *The surplus v is supermodular (equivalently, has increasing differences) in (t, p): $v(t'', p'') + v(t', p') > v(t'', p') + v(t', p'')$ for all $t'' > t'$ and $p'' > p'$.*

Assumption 1 provides a form of complementarity between the agent's type and their position. Supermodularity of the surplus function is sufficient to guarantee that the first-best assignment features positive assortative matching: i.e. it is optimal to assign higher positions to higher types. Such assortative matching is common in all of the motivating examples for this paper, and supermodularity is commonly used to motivate this observation in many applications.

Assumption 2 (Continuity)**.** *The expected surplus $\mathbb{E}[v(t, p)|x, s]$ is continuous in s.*

Assumption 2 is a mild technical assumption for practical applications, as it only requires that the expected surplus for an agent with signal s varies continuously in s, holding the position p and the category x constant. This assumption is satisfied whenever the density $f(t|s)$ is uniformly continuous is s, which by the monotone convergence theorem implies that $\mathbb{E}[v(t, p)|x, s]$ is continuous in s. Moreover, the main intuitions of this paper do not crucially rely on continuity, and one can state analogous results in the absence of this assumption.

[3]Note that this surplus depends on types, whereas signals are not directly payoff-relevant. Alternatively, one can consider a model where signals are themselves productive, i.e. $V : S \times T \times P \to \mathbb{R}$. As will be clear from the subsequent analysis, my main results on differential treatment would continue to hold qualitatively. That is, the optimal policy has similar features and the main intuitions remain valid, but the amount of differential treatment is moderated to some extent. This setting offers a fruitful avenue for future research.

Assumption 3 (MLRP)**.** *The family of signal distributions* $\{F_{t,x}\}$ *satisfies the strict monotone likelihood ratio property in* t*: for* $s'' > s'$*, the ratio of the densities* $\frac{f_{t,x}(s'')}{f_{t,x}(s')}$ *is increasing in* t*, conditional on* x*, whenever it is well-defined.*[4]

Assumption 3 is a standard assumption in the literature. It ensures that there is an intuitive way to interpret signals and update beliefs regarding an agent's type, conditional on observing some signal, as in the following proposition.

Proposition 1 (Milgrom 1981)**.** *Under Assumption 3,* $F(\cdot|x_i, s_i = s'') \succ_{FOSD} F(\cdot|x_i, s_i = s')$ *for any* $s'' > s'$*.*

Proposition 1, due to Milgrom (1981), shows that observing a higher signal s is good news regarding the agent's type (conditional on any category x), i.e. a higher signal improves the distribution of t in the sense of first-order stochastic dominance. This proposition implies that $\mathbb{E}[v(t,p)|x,s] = \int_T v(t,p)dF(t|x,s)$ is increasing in s, for any x, p, since $v(t,p)$ is increasing in t and $F(t|x,s)$ is increasing in s in the first order. Furthermore, the proposition also implies that $\mathbb{E}[v(t,p'')|x,s] - \mathbb{E}[v(t,p')|x,s] = \int_T v(t,p'') - v(t,p')dF(t|x,s)$ is increasing in s, for any x and $p'' > p'$, because $v(t,p)$ is supermodular, hence $v(t,p'') - v(t,p')$ is an increasing function of t, and $F(t|x,s)$ is increasing in s in the first order.

The assumptions above are assumed to hold within categories. Next, I define a notion of comparability across different categories. In principle the distributions of types and signals may differ in arbitrary ways across categories, e.g. as functions of x, these distributions may be shifting or spreading, or may not be comparable using any of the standard stochastic orders at all. Indeed, for the main result of the paper it is sufficient to define comparability of the categories as in the definition below, which is much weaker than the usual notions of stochastic orders over distributions. First, this definition only imposes conditions on the expected values of the distributions, rather than the whole distribution functions. Second, these conditions allow for comparisons of distributions that cannot be ranked according to FOSD, SOSD, etc. Third, such conditions may be easier to test empirically, as they only require comparing conditional expectations.

Definition 1. *Two categories,* x *and* x'*, are* ***comparable*** *if* $\forall s \in S$ *either:*

[4]Alternatively, one can define MLRP as $f_{t'',x}(s'')f_{t',x}(s') > f_{t',x}(s'')f_{t'',x}(s')$ for all $t'' > t'$ and $s'' > s'$, such that $f_{t',x}(s') > 0$ or $f_{t'',x}(s') > 0$, holding x constant (cf. Athey (2002)). The latter is only slightly more general and makes no difference to the analysis, so I focus on the case where the ratio is well-defined.

(i) $\exists s' \in S$ *s.t.* $\mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s] = \mathbb{E}[v(t,p)|x',s'] - \mathbb{E}[v(t,p')|x',s']\, \forall p, p'$, *or*

(ii) $\forall s' \in S$: *either* $\forall p$ *and* $p' < p$ $\mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s] \geq \mathbb{E}[v(t,p)|x',s'] - \mathbb{E}[v(t,p')|x',s']$, *or* $\forall p$ *and* $p' < p$ $\mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s] \leq \mathbb{E}[v(t,p)|x',s'] - \mathbb{E}[v(t,p')|x',s']$.

This definition of comparability establishes an equivalence across categories, defined in terms of differences in expected surplus. For example, a signal s, for given category x, may be similar to a signal $s' \neq s$, for given category x', in the sense that they imply the same differences in expected values for different positions, or one implies higher or lower differences in expected values across positions. Such comparability is required to hold for any p and p', i.e. it is invariant across p.

One can define an alternative, more demanding notion of strong comparability, which implies the notion above and is easier to state (and perhaps to verify empirically, as it does not require comparing differences across positions). This stronger notion is not necessary for my main results, but it relates to several of the propositions that follow below, so I state it here for completeness.

Definition 2. *Two categories, x and x', are* ***strongly comparable*** *if* $\forall s \in S$ *either:*

(i) $\exists s' \in S$ *s.t.* $\mathbb{E}[v(t,p)|x,s] = \mathbb{E}[v(t,p)|x',s']\, \forall p$, *or*

(ii) $\forall s' \in S$: *either* $\mathbb{E}[v(t,p)|x,s] \geq \mathbb{E}[v(t,p)|x',s']\, \forall p$ *or* $\mathbb{E}[v(t,p)|x,s] \leq \mathbb{E}[v(t,p)|x',s']\, \forall p$.

The following examples illustrate the notions of comparability above.

Example 1. *Consider the following 4 categories, x_1, x_2, x_3, x_4. Note that x_1 and x_2 are comparable, x_1 and x_3 are strongly comparable, and x_1 and x_4 are not comparable.*

x_1	p_1	p_2	p_3
$\mathbb{E}[v(t,p)|x_1,s_1]$	1	3	6
$\mathbb{E}[v(t,p)|x_1,s_2]$	2	5	10
$\mathbb{E}[v(t,p)|x_1,s_3]$	3	7	13

x_2	p_1	p_2	p_3
$\mathbb{E}[v(t,p)|x_2,s_1]$	1	4	9
$\mathbb{E}[v(t,p)|x_2,s_2]$	4	8	14
$\mathbb{E}[v(t,p)|x_2,s_3]$	5	10	17

x_3	p_1	p_2	p_3
$\mathbb{E}[v(t,p)|x_3,s_1]$	2	5	10
$\mathbb{E}[v(t,p)|x_3,s_2]$	3	7	13
$\mathbb{E}[v(t,p)|x_3,s_3]$	4	9	16

x_4	p_1	p_2	p_3
$\mathbb{E}[v(t,p)|x_4,s_1]$	1	2	4
$\mathbb{E}[v(t,p)|x_4,s_2]$	2	3.9	10
$\mathbb{E}[v(t,p)|x_4,s_3]$	3	7	14

First, categories x_1 and x_2 are comparable because

$$\mathbb{E}[v(t,p')|x_1,s_2] - \mathbb{E}[v(t,p)|x_1,s_2] = \mathbb{E}[v(t,p')|x_2,s_1] - \mathbb{E}[v(t,p)|x_2,s_1]\,\forall p,p',$$
$$\mathbb{E}[v(t,p')|x_1,s_3] - \mathbb{E}[v(t,p)|x_1,s_3] = \mathbb{E}[v(t,p')|x_2,s_2] - \mathbb{E}[v(t,p)|x_2,s_2]\,\forall p,p',$$

and the remaining signals also satisfy the conditions for comparability.

Second, categories x_1 and x_3 are strongly comparable because $\mathbb{E}[v(t,p)|x_1,s_2] = \mathbb{E}[v(t,p)|x_2,s_1]\,\forall p$, and $\mathbb{E}[v(t,p)|x_1,s_3] = \mathbb{E}[v(t,p)|x_2,s_2]\,\forall p$, and the remaining signals also satisfy the conditions for strong comparability.

Finally, categories x_1 and x_4 are not comparable: notice that

$$\mathbb{E}[v(t,p_2)|x_4,s_2] - \mathbb{E}[v(t,p_1)|x_4,s_2] < \mathbb{E}[v(t,p_2)|x_1,s] - \mathbb{E}[v(t,p_1)|x_1,s]\,\forall s,$$
$$\mathbb{E}[v(t,p_3)|x_4,s_2] - \mathbb{E}[v(t,p_2)|x_4,s_2] > \mathbb{E}[v(t,p_3)|x_1,s] - \mathbb{E}[v(t,p_2)|x_1,s]\,\forall s,$$

hence the signal s_2 for category x_4 cannot be compared to any signal for category x_1.

3.1 Conditions for comparability

Comparability across categories is satisfied in many environments. I provide three propositions which characterize some natural settings where it holds: (i) when the set of positions is binary; (ii) when $v(\cdot)$ satisfies an additional condition, which ensures that categories are comparable for any $\{F_{t,x}\}$; and (iii) when $\{F_{t,x}\}$ satisfies an additional condition, which ensures that categories are comparable for any $v(\cdot)$.

Proposition 2. *Suppose $P = (p_1, p_2)$, with corresponding capacities k_1, k_2. Then all categories are comparable.*

With a binary set of positions, the assumptions made earlier ensure that categories are comparable. These assumptions ensure that $\mathbb{E}[v(t,p)|x,s]$ is increasing in s for any p, and moreover $\mathbb{E}[v(t,p_2)|x,s] - \mathbb{E}[v(t,p_1)|x,s]$ is also continuous and increasing in s. Therefore a violation of comparability requires at least 3 positions among which agents must be assigned.

There are several applications where the set of treatments that the policy-maker must assign is binary: in school and university admissions, the decision to admit or not admit a student is binary, and similarly hiring decisions in organizations are also binary. In such settings, categories are comparable without any further assumptions on $v(t,p)$ or $\{F_{t,x}\}$.

The next proposition provides another environment where comparability holds.

Proposition 3. *Suppose $v(t,p)$ is multiplicatively separable in v and p. Then all categories are strongly comparable (hence also comparable), for any family of signals $\{F_{t,x}\}$.*

Proposition 3 shows that comparability is satisfied for *any* family of distributions $\{F_{t,x}\}$, as long as $v(\cdot)$ is multiplicatively separable. This makes comparability a widely applicable concept. Such separable value functions are commonly used in auction theory, mechanism design, and other settings, as they provide easy ways to model complementarities and supermodularity. For example, the value function in a quasilinear environment is $v \cdot x$, where v is the agent's type and x is the allocation; the equivalent here is $v(t,p) = t \cdot p$.

Next I provide a definition of equivalence of signal structures, which yields another sufficient condition for categories to be comparable, for any $v(\cdot)$. This result provides an alternative approach to think about comparability, which complements the one in Proposition 3.

To state the definition, first let $\widetilde{v}_{p,x}(s) \equiv \mathbb{E}[v(t,p)|x,s]$ denote the expected value of $v(\cdot)$ with respect to $F(t|x,s)$, for any p and x. Then let $\widetilde{w}_{p,x}(v) \equiv \widetilde{v}_{p,x}^{-1}(s)$ denote the inverse, for any p and x.[5]

Definition 3. *The signal structures $F_{t,x}$ and $F_{t,x'}$ are **equivalent** with respect to $v(\cdot)$ if $\exists \delta$ s.t.*

$$\widetilde{w}_{p,x}(v) = \widetilde{w}_{p,x'}(v) + \delta$$

for any p, for all $v \in \mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)]$.
*The signals $F_{t,x}$ and $F_{t,x'}$ are **identical** with respect to $v(\cdot)$ if this holds for $\delta = 0$.*

In words, two signal structures are *equivalent* if the expected values as a function of signals are horizontal translations of each other, over their common image. That is, the expected value $\widetilde{v}_{p,x}(s)$ corresponding to a signal distribution $F_{t,x}$ is a shift to the left or right of the expected value $\widetilde{v}_{p,x'}(s)$ corresponding to a signal distribution $F_{t,x'}$. These expected value functions only need to agree with each other over the range where their images intersect. Analogously, two signal structures are *identical* with respect to $v(\cdot)$ if their expected values as a function of signals are exactly equal. This is the case, for example, if $F(t|x,s)$ is constant in x, i.e. t conditional on s is independent of x.

[5] This inverse is well-defined, because $v(t,p)$ is increasing in t and the expectation is taken with distributions $F(t|x,s)$ which are FOSD-increasing in s.

Example 2 below is an especially simple one, where the distributions of signals are horizontal translations of one another. But the equivalence can hold more generally—it would be sufficient, for example, that for every s there exists some $s' = s + \delta$ such that the Bayesian posteriors $F(t|x, s)$ and $F(t|x', s')$ are equal.

Example 2. *Let* $t \in \{t_L, t_H\}$ *and* $S = [0, 1]$*. Suppose*

$$F(s|t,x) = \begin{cases} \min\{3s, 1\} & \textit{for } t = t_L \\ \min\{(3s)^3, 1\} & \textit{for } t = t_H \end{cases}$$

$$F(s|t,x') = \begin{cases} \max\{0, 3(s - \frac{2}{3})\} & \textit{for } t = t_L \\ \max\{0, [3(s - \frac{2}{3})]^3\} & \textit{for } t = t_H. \end{cases}$$

Then $F_{t,x}$ *and* $F_{t,x'}$ *are equivalent, because a signal* s *for category* x *is equivalent to a signal* $s' = s + \frac{2}{3}$ *for category* x*'.*

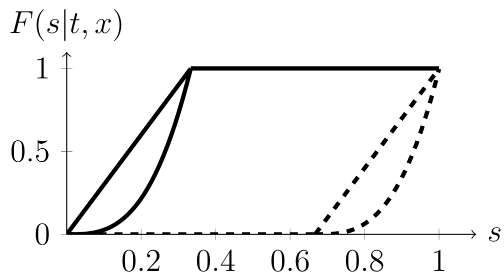

Figure 1: Signal distributions for x (solid) and x' (dashed)

The definition of signal equivalence above is useful because it provides another type of sufficient condition for categories to be comparable, without imposing additional conditions on $v(\cdot)$.

Proposition 4. *Suppose the family of signal distributions* $\{F_{t,x}\}$ *are (pairwise) equivalent. Then all categories are strongly comparable.*

Proposition 4 establishes a close connection between the equivalence of signals and the comparability of categories. Intuitively, if a signal s, for given category x, is equivalent to a signal s', for given category x', then the two categories are comparable. For the remainder of the analysis I will assume that all categories are comparable.

3.2 Optimal policy

Theorem 1. *If categories are comparable, the optimal policy $\mathscr{P}^*$ assigns agents to positions assortatively with respect to an index defined by $r(x,s) = \mathbb{E}[v(t,p_m)|x,s] - \mathbb{E}[v(t,p_{m-1})|x,s]$.*

In this theorem I first define an index function, given by $\mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s]$, for some $p > p'$, and show that this index induces a total preorder over $\{(x_i, s_i)\}$. I then show that this preorder is invariant with respect to p and p'. Hence without loss of generality I let $p = p_m$ and $p' = p_{m-1}$. Then I show that among policies that are deterministic (up to tie-breaking among index-equivalent agents), optimality implies that agents are assigned assortatively according to this index. Finally, I extend the result to stochastic policies, to show that there is no gain from randomization among non-payoff-equivalent agents; hence the optimal policy is essentially deterministic, and assigns agents to positions assortatively with respect to the index function, i.e. the optimal policy is weakly increasing in $r(x,s)$.

The Hungarian assignment algorithm and recent developments from the engineering and computer science literatures that build on it can be used to solve for the optimal assignment even in the absence of comparability. While these algorithms find the optimal assignment in a more general setting, they are less interpretable in terms of economic intuition and do not provide comparative statics, since the algorithms do not have an index implementation for the optimal assignment, unlike the result in Theorem 1. That is, the predictions of these algorithms are more of a black box. Theorem 1 and the definition of comparability suggest an interesting avenue for future work: one can decentralize this assignment model so that a different principal is in charge of each treatment or position, and decides whether each agent is admitted to it or not in order to maximize the set of types admitted to that position. Such a model would be useful to study decentralized college admissions, relative to the efficient assignment that is characterized in this paper.[6]

The proof of Theorem 1 also implies the following observation.

Corollary 1. *The optimal policy is deterministic, except for tie-breaking among payoff-equivalent agents.*

Moreover, the optimal assignment described in Theorem 1 can be implemented with category-

[6] Interestingly, the equilibrium of this decentralized version of the model will be efficient if and only if categories are comparable, suggesting that the concept of comparability has important implications beyond the scope of this paper. This is a promising direction that will be explored in subsequent research.

specific signal thresholds for each position. The proof of the result directly shows how to set these thresholds optimally.

Corollary 2. *For a set of agents with observables $\{(x_i, s_i)\}$, the optimal policy is implemented by category-specific signal cutoffs for each position:* $\bar{s}_p(x) \equiv \min\{s_i : \mathbb{P}^*(p|x, s_i) > 0\}$.

Theorem 1 shows that affirmative action arises endogenously in this environment, as the solution to the policy-maker's surplus maximization problem. In general, the category-specific cutoffs that implement $\mathscr{P}^*$ need not be the same across categories, and hence some categories of agents will be favored, relative to their signal realizations. In an education context, it may be efficient to admit students to schools or universities according to exam score cutoffs that are different across categories (e.g. across socio-economic status or some other observable covariate). Similarly, in a labor market context, it may be efficient to hire or promote employees according to performance thresholds that are different across categories (e.g. across gender or some other observable covariate).

Corollary 2 implies that the optimal policy is assortative with respect to the agent's index, $r(x, s)$, but in general need not be assortative with respect to signals. The optimal assignment is monotonic in signals within categories, but not across categories. That is, to maximize surplus the policy-maker may assign agents with lower signals to higher positions. Note that assortativity with respect to signals would be equivalent to the optimal signal threshold for each position being category-invariant, i.e. $\bar{s}_p(x)$ would be constant in x for each p. Example 3 below illustrates the non-monotonicity with respect to signals in a very simple setting.

Example 3. *Suppose there are 2 positions, p_1, p_2, with capacities $k_1 = 2$, $k_2 = 1$, and 3 agents with observables $(x_1, s_1), (x_1, s_2), (x_2, s_1)$, for some categories x_1, x_2 and some signals, $s_1 < s_2$, with corresponding expected values for the positions given by the following table:*

	p_1	p_2
$\mathbb{E}[v(t,p)\|x_1, s_1]$	*1*	*2*
$\mathbb{E}[v(t,p)\|x_1, s_2]$	*2*	*4*
$\mathbb{E}[v(t,p)\|x_2, s_1]$	*0*	*3*

The optimal assignment here is $p(x_2, s_1) = p_2$, $p(x_1, s_2) = p(x_1, s_1) = p_1$. So $p(x_2, s_1) > p(x_1, s_2)$ even though $s_1 < s_2$, and despite the fact that $\mathbb{E}[v(t, p_2)|x_2, s_1] < \mathbb{E}[v(t, p_2)|x_1, s_2]$.

Therefore the optimal policy is not increasing with respect to either the agents' signals or the agents' expected values.

The next result shows that the optimal policy is assortative with respect to signals only in a very special case, when the signal structures are identical with respect to $v(\cdot)$. More generally, one might only expect the optimal policy to be "quasi-assortative," in the sense that it assigns agents to policies according to their index, which only loosely corresponds to their signal realizations.

Theorem 2. *The optimal policy $\mathscr{P}^*$ can be implemented with category-invariant cutoffs for any $(k_1, ..., k_m)$ if and only if the signals $\{F_{t,x}\}$ are identical with respect to $v(\cdot)$.*
In particular, if $\{F_{t,x}\}$ are equivalent but not identical, $\exists$ categories x, x', capacities $(k_1, ..., k_m)$ and signals $s > s'$ such that $\mathbb{P}^*(x, s) < \mathbb{P}^*(x', s')$.

This result illustrates a *nominal* aspect of affirmative action: to maximize surplus the policy-maker must translate signals across categories. Since surplus is a function of types, not signals per se, the policy-maker maps signals into beliefs about types, and this mapping may be different across categories.[7] The signal structures across different categories may be inherently biased in favor of some categories, unless these signal structures are identical, which is unlikely to be the case in practice. If the distributions that generate these signals are not identical, then this translation will compensate agents from categories whose signals are inherently worse, giving rise to affirmative action.

The next result highlights a separate, *substantive* aspect of affirmative action: even if signals are unbiased across categories (i.e. a signal implies the same expected type for any category), the surplus-maximizing policy may feature affirmative action in favor of agents whose categories have noisier or less noisy signals. This is because the optimal policy is assortative with respect to $r(x, s)$, which need not be constant in x even if $\mathbb{E}[t|x, s]$ is constant in x. In particular, if v exhibits some intrinsic preference for or against variability, then agents whose expected types are equal, conditional on the same signal, will be ranked differently according to $r(x, s)$.

This intuition is quite general and can be formalized and illustrated in a more specific setting.

[7] The definition of comparability ensures that such a mapping is possible—if categories are not comparable, then the optimal policy may assign agents to positions in a way that is not assortative with respect to any index across categories.

To do so, I consider two special classes of surplus functions, which exhibit a preference for and against dispersion in types, akin to risk averse and and risk seeking preferences.

Definition 4. *A value function $v(t,p)$ has (strictly)* ***convex differences in*** *t if $v(t,p') - v(t,p)$ is a (strictly) convex function of t, for any $p' > p$. Analogously, $v(t,p)$ has (strictly)* ***concave differences in*** *t if $v(t,p') - v(t,p)$ is a (strictly) concave function of t, for any $p' > p$.*[8]

The next result uses these definitions to illustrate the substantive aspect of affirmative action.

Theorem 3. *Suppose $\{F_{s,x}\}$ is such that $\mathbb{E}[t|x,s]$ is constant in x, for every s, and for some x', x'', $F(t|x'',s)$ is a mean-preserving spread of $F(t|x',s)$.*
If $v(t,p)$ has convex differences in t, then $r(x'',s) \geq r(x',s)$ for all s, with a strict inequality when v has strictly convex differences in t. If $v(t,p)$ has concave differences in t, then $r(x'',s) \leq r(x',s)$ for all s, with a strict inequality if v has strictly concave differences in t.

Theorem 3 illustrates the substantive aspect of affirmative action: even if signals are unbiased, the noisiness of signals affects how agents are ranked across categories. Because these ranks differ across categories, there exist problems where the capacities are such that the optimal policy assigns some agents with lower signals to higher positions, based on the differences in ranks across categories. I.e. agents from categories with more or less noisy signals may have a higher or lower index, depending on whether v has convex or concave differences in type. Noisiness is formalized in terms of mean preserving spreads, which in this case is equivalent to the convex stochastic order and second-order stochastic dominance.

3.3 Applications and extensions

Theorem 1 provides a general characterization of the optimal policy which maximizes expected total surplus. Theorem 2 shows when the optimal policy features differential treatment. Theorem 3 shows that differential treatment can be optimal because of differences in the informativeness or noisiness of signal structures across characteristics, not just because of "biases" in these signals. These results are derived in a general, non-parametric setup. In the following subsections I consider some applications to more specific settings which may be of interest to policy-makers.

[8] One can also define analogous properties of the differences when T is a discrete space, but for ease of exposition I will focus on the case where T is continuous.

3.3.1 Affirmative action with biased signals

In some environments signalling technologies may be biased across categories. That is, for some categories the mapping between types and signals may be consistently higher, in some sense, thus making the informational content of signals biased across categories. To model such an application, I assume that signal structures can also be ordered across categories (whereas in my main analysis I only assumed that signals can be compared according to the MLRP *within* categories). In particular, the distributions of types conditional on some signal s may be comparable across characteristics x in the FOSD sense.

In the context of university admissions, for example, it may be the case that the distribution of abilities, conditional on some university entrance score s, is FOSD-decreasing in a socio-economic status covariate x. That is, conditional on the same exam score, a lower socio-economic status student would have a higher distribution of ability, if socio-economic status tends to raise students' exam scores in some sense. This may reflect the fact that in order to achieve the same exam score, a lower socio-economic status student may have to overcome larger obstacles, requiring higher ability; or it may reflect the beneficial effect of parental investments in education on exam performance.[9] This setting can be formalized with the following assumption.

Assumption 4. *Suppose X can be ordered in such a way that $F(\cdot|x', s) \succ_{FOSD} F(\cdot|x'', s)$ for $x' < x''$, for any s.*

Assumption 4 says that the set of possible characteristics X can be ordered in some way so that "higher" characteristics are more strongly associated with lower types, given the same signal. A sufficient condition for this assumption to hold would be that the family of distributions $\{F_{t,x}\}$ satisfies a version of the strict monotone likelihood ratio property in x,[10] which would imply Assumption 4, analogously to Proposition 1.

[9] The literature on human capital (see e.g. Todd and Wolpin, 2007; Heckman, 2011; Fiorini and Keane, 2014) documents that higher socio-economic status parents tend to contribute more time and money to their children's education, which would justify such an assumption. Such an assumption is also consistent with the idea of "belief flipping" in the theoretical models of Fryer (2007) and Bohren, Imas and Rosenberg (2018), and with the latter's experimental evidence of belief reversion.

[10] Specifically: the family of signal distributions $\{F_{t,x}\}$ satisfies the strict inverse monotone likelihood ratio property in x if X can be ordered so that for $x'' > x'$, the ratio of the densities $\frac{f_{t,x'}(s)}{f_{t,x''}(s)}$ is increasing in t, for any s, whenever it is well-defined.

Following a similar logic to that in Theorem 1, Assumption 4 implies that the optimal policy features a form of affirmative action, which favors lower categories.

Proposition 5. *If categories are comparable, the optimal policy $\mathscr{P}^*$ is implemented by category-specific cutoffs $\bar{s}_p(x)$ that are increasing in x.*

That is, conditional on the same signal s, a lower-category agent is assigned according to a lower cutoff for each position, compared to a higher-category agent, and hence is favored by the optimal assignment policy.[11] This is because the optimal policy assigns agents according to the ranking function $r(x, s)$, which is increasing in s and decreasing in x under the assumptions in this application of the model.[12] Thus Proposition 5 characterizes a particular form of affirmative action which maximizes surplus: agents are treated differently depending on their characteristics, with lower-characteristics agents being "favored" by implementing lower signal cutoffs for each position.

3.3.2 Optimal treatment inequality with unbiased signals

My results so far have considered how a policy-maker should use the signals and characteristics of the agents to implement an assignment that maximizes total surplus. In many empirically relevant settings I find that the optimal policy "favors" some groups, relative to their *signals*, in the sense that it treats signals differently across characteristics. But this need not imply that different groups receive different treatments relative to their *types*. For example, if signals for one group are mechanically shifted downwards, compared to another group, as in Example 2, it may be the case that the mapping from types to positions induced by $\mathscr{P}^*$ is in fact the same across X—that is, no group is actually favored when one considers how types are treated, even though $\mathscr{P}^*$ treats groups differentially in terms of signals.

However, in some cases $\mathscr{P}^*$ does induce unequal treatment across groups, with respect to types, not just signals. Interestingly, this can also be the case when signals are unbiased

[11] This favoritism may not strictly benefit lower-category agents, if the realizations of observables $\{(x_i, s_i)\}_i$ and the capacities for each position happen to be such that the differences in ranks are irrelevant around the cutoffs for each position. But more generally, because $r(x, s)$ is strictly decreasing in x, there exist some capacities and realizations of the observables such that the optimal policy strictly favors lower-category agents.

[12] The fact that categories are comparable ensures that such a ranking function is invariant with respect to the positions (p, p') with which $r(x, s)$ is defined, as in the proof of Theorem 1.

across characteristics, i.e. $\mathbb{E}[t|x,s]$ is constant in x. In this section I study such a setting as an application of the model.

Assumption 5. *Suppose $\{F_{s,x}\}$ is such that $\mathbb{E}[t|x,s]$ is constant in x, for every s, and for some x', x'', $F(t|x'',s)$ is a mean-preserving spread of $F(t|x',s)$.*

Assumption 5 is the same as the setting of Theorem 3, and represents an environment where signals are unbiased across categories, but their noisiness differs across categories. Signals are noisier for category x'', in the sense that the distribution of types conditional on a signal s is a mean-preserving spread of the distribution of types for category x', conditional on the same signal. In an education context, this could be the case if university entrance scores are more or less noisy measures of ability for some categories than others; that is, conditional on the same score, one group may have higher or lower dispersion in ability.

Proposition 6. *If $v(t,p)$ has convex [concave] differences in t, then $p^*(x'',s) \geq [\leq] p^*(x',s)$ for all s, with strict inequality for some capacities and realizations of the observables. Moreover, for all p, $\{s : \mathbb{P}^*(p|x'',s) > 0\} \leq [\geq] \{s : \mathbb{P}^*(p|x',s) > 0\}$ in the strong set order.*

Proposition 6 shows that aggregate inequality of treatments across groups can be optimal even when signals are unbiased, with identical prior distributions of types across groups. This is driven by the fact that the optimal policy assigns agents according to their expected incremental gains from higher treatments, not according to their expected types. Therefore the treatments of agents from categories x' and x'' differ in aggregate, with x'' receiving higher treatments when $v(t,p)$ has convex differences in t, because x'' has noisier signals and the policy-maker has a preference for dispersion. As a result, for any given treatment p, agents from category x'' who are assigned to p have lower signals, leading to aggregate inequality, despite the fact that the prior distribution of types is the same across categories.

In the context of university admissions this leads to the conclusion that average exam scores or expected abilities need not be equal across groups for each treatment. Rather, an efficient admission policy would equalize the expected incremental surplus from admission. Similarly, in the context of hiring and promotion policies in organizations, the result implies that past performance or expected productivity need not be equal across groups.

The optimal treatment inequality characterized in Proposition 6 can be illustrated with a parametric example.

Example 4. *Suppose a university is deciding which of* 2 *students to admit, i.e. the positions are* (p_1, p_2) *with capacities* $k_1 = k_2 = 1$. *The students have normally distributed abilities,* $t_i \sim N(0, 1)$, *and belong to categories* x^1 *and* x^2. *A category* x^j *student has a university admission score* $s_i = t_i + \varepsilon_i^j$, *where* $\varepsilon_i^j \sim N(0, \sigma_j)$ *is zero-mean, normally-distributed noise, with variance* $\sigma_2 > \sigma_1$, *for categories* x^1, x^2. *Note that* $F(t_2|x^2, s_2 = s)$ *is a mean-preserving spread of* $F(t_1|x^1, s_1 = s)$, *for any* s.

For any $v(t, p)$ *that has strictly convex differences in* t, *conditional on* $s_1 = s_2 = s$, *the university would strictly prefer to admit student* 2, *since* $r(x^2, s) > r(x^1, s)$, *and for any* s_2 *in some neighborhood below* s, *the university would prefer to admit student* 2. *That is, there exists some threshold* $\bar{s}_2(s_1) < s_1$ *such that for a given* s_1, *the university prefers to admit student* 2 *if and only if* $s_2 \geq \bar{s}_2(s_1)$.

Then student 2 *is admitted with probability* $\mathbb{P}(s_2 \geq \bar{s}_2(s_1)) = \int \mathbb{P}(s_2 - s_1 \geq \bar{s}_2(s_1) - s_1|s_1)dF(s_1)$. *Note that* $s_2 - s_1$ *is normally distributed, with mean* 0, *and for any* s_1, $\bar{s}_2(s_1) - s_1 < 0$. *Therefore the integrand is strictly larger than* $\frac{1}{2}$ *everywhere, and* $\mathbb{P}(s_2 \geq \bar{s}_2(s_1)) > \frac{1}{2}$. *That is, the optimal policy is more likely to admit student* 2 *than student* 1. *Moreover, given that the type distributions are the same ex ante, the expected ability of student* 2 *conditional on* 2 *being admitted is lower than the corresponding expected ability of student* 1 *conditional on* 1 *being admitted. Therefore aggregate inequality arises both in terms of admission scores, since* $\bar{s}_2(s_1) < s_1$, *and in terms of expected types who are admitted across different categories.*

4 Conclusions

I study a setting where a policy-maker wants to assign agents with unobservable types and observable signals and characteristics to different positions or treatments. I characterize the policy that maximizes expected total surplus: it assigns agents to positions monotonically with respect to an index function that measures the expected incremental gains from different treatments. However the optimal policy is in general not monotonic with respect to the agents' signals, or even the agents' expected types. Therefore the optimal policy will feature differential treatment in a variety of cases, where the distributions of signals conditional on types vary across characteristics.

I highlight two main intuitions for these results. First, to maximize surplus the policy-maker must translate signals into beliefs about types, and this translation need not be the same across characteristics if some groups of agents have different signal distributions—i.e. if signals are "biased" across characteristics. In this case the policy-maker may optimally favor groups whose signal distributions are worse. Second, differential treatment can be optimal even if signals are unbiased across characteristics, provided their informativeness differs across characteristics. That is, different groups may have more or less noisy signal distributions, and the optimal policy may favor some groups depending on whether the policy-maker's objective function has a preference for or against dispersion in types.

These results provide a novel efficiency-based rationale for affirmative action and other differential treatment policies in education and labor markets. Affirmative action has clear efficiency implications, not just ethical ones, which should inform decisions and policy-making. The model also highlights some important features for empirical research. First, an efficient assignment policy should equalize the expected incremental gains from treatments for the marginally treated agents. Second, treatment cutoffs need not in general be equal across characteristics, if the distributions of signals differ. Third, average signals and types within each treatment need not be equal across characteristics. These observations can be clearly mapped into measurable quantities in the contexts of university admissions and hiring and promotion policies in organizations.

5 References

Abdulkadiroğlu, Atila (2005) "College admissions with affirmative action," *International Journal of Game Theory*, Vol. 33, No. 4, pp. 535–549.

Abdulkadiroğlu, Atila and Tayfun Sŏnmez (2003) "School Choice: A Mechanism Design Approach," *American Economic Review*, Vol. 93, No. 3, pp. 729–747.

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

Athey, Susan (2002) "Monotone Comparative Statics under Uncertainty," *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 187–223.

Becker, Gary S (1957) *The economics of discrimination*: University of Chicago press.

Blume, Lawrence E (2006) “The dynamics of statistical discrimination,” *The Economic Journal*, Vol. 116, No. 515, pp. F480–F498.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg (2018) “The Dynamics of Discrimination: Theory and Evidence,” *Working Paper*.

Chambers, Christopher and Federico Echenique (2019) “A characterization of “Phelpsian” statistical discrimination,” *Working Paper*.

Chan, Jimmy and Erik Eyster (2003) “Does banning affirmative action lower college student quality?” *American Economic Review*, Vol. 93, No. 3, pp. 858–872.

Chung, Kim-Sau (2000) “Role models and arguments for affirmative action,” *American Economic Review*, Vol. 90, No. 3, pp. 640–648.

Coate, Stephen and Glenn C Loury (1993) “Will affirmative-action policies eliminate negative stereotypes?” *The American Economic Review*, pp. 1220–1240.

Cornell, Bradford and Ivo Welch (1996) “Culture, information, and screening discrimination,” *Journal of political Economy*, Vol. 104, No. 3, pp. 542–571.

Doğan, Battal (2016) “Responsive affirmative action in school choice,” *Journal of Economic Theory*, Vol. 165, pp. 69 – 105.

Eeckhout, Jan (2006) “Minorities and endogenous segregation,” *The Review of Economic Studies*, Vol. 73, No. 1, pp. 31–53.

Ehlers, Lars, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim (2014) “School choice with controlled choice constraints: Hard bounds versus soft bounds,” *Journal of Economic Theory*, Vol. 153, pp. 648 – 683.

Epple, Dennis, Richard Romano, and Holger Sieg (2008) “Diversity and affirmative action in higher education,” *Journal of Public Economic Theory*, Vol. 10, No. 4, pp. 475–501.

Fang, Hanming and Andrea Moro (2011) “Theories of statistical discrimination and affirmative action: A survey,” in *Handbook of social economics*, Vol. 1: Elsevier, pp. 133–200.

Fang, Hanming and Peter Norman (2006) “Government-Mandated Discriminatory Policies: Theory and Evidence,” *International Economic Review*, Vol. 47, No. 2, pp. 361–389.

Fiorini, Mario and Michael P Keane (2014) "How the allocation of children's time affects cognitive and noncognitive development," *Journal of Labor Economics*, Vol. 32, No. 4, pp. 787–836.

Fryer, Roland G (2007) "Belief flipping in a dynamic model of statistical discrimination," *Journal of Public Economics*, Vol. 91, No. 5-6, pp. 1151–1166.

Fryer, Roland G and Glenn C Loury (2007) *Valuing identity: The simple economics of affirmative action policies*: Mimeo.

Fu, Qiang (2007) "A Theory of Affirmative Action in College Admissions," *Economic Inquiry*, Vol. 44, No. 3, pp. 420–428.

Hafalir, Isa E., Fuhito Kojima, and M. Bumin Yenmez (2019) "Interdistrict School Choice: A Theory Of Student Assignment," *Working Paper*.

Hafalir, Isa E, M Bumin Yenmez, and Muhammed A Yildirim (2013) "Effective affirmative action in school choice," *Theoretical Economics*, Vol. 8, No. 2, pp. 325–363.

Heckman, James J (2011) "The economics of inequality: The value of early childhood education.," *American Educator*, Vol. 35, No. 1, p. 31.

Kojima, Fuhito (2012) "School choice: Impossibilities for affirmative action," *Games and Economic Behavior*, Vol. 75, No. 2, pp. 685 – 693.

Levin, Jonathan (2009) "The dynamics of collective reputation," *The BE Journal of Theoretical Economics*, Vol. 9, No. 1.

Liu, Qingmin, George J. Mailath, Andrew Postlewaite, and Larry Samuelson (2014) "Stable Matching With Incomplete Information," *Econometrica*, Vol. 82, No. 2, pp. 541–587.

Lundberg, Shelly J and Richard Startz (1983) "Private discrimination and social intervention in competitive labor market," *The American Economic Review*, Vol. 73, No. 3, pp. 340–347.

Mailath, George J, Larry Samuelson, and Avner Shaked (2000) "Endogenous inequality in integrated labor markets with two-sided search," *American Economic Review*, Vol. 90, No. 1, pp. 46–72.

Milgrom, Paul R (1981) "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, pp. 380–391.

Moro, Andrea and Peter Norman (2003a) "Affirmative action in a competitive economy," *Journal of Public Economics*, Vol. 87, No. 3-4, pp. 567–594.

——— (2003b) "Empirical implications of statistical discrimination on the returns to measures of skills," *Annales d'Economie et de Statistique*, pp. 399–417.

——— (2004) "A general equilibrium model of statistical discrimination," *Journal of economic theory*, Vol. 114, No. 1, pp. 1–30.

Norman, Peter (2003) "Statistical discrimination and efficiency," *The Review of Economic Studies*, Vol. 70, No. 3, pp. 615–627.

Pai, Malesh and Rakesh Vohra (2014) "The Design of Affirmative Action Policy: A Mechanism Design Approach," *Working Paper*.

Phelps, Edmund S (1972) "The statistical theory of racism and sexism," *The american economic review*, pp. 659–661.

Todd, Petra E and Kenneth I Wolpin (2007) "The production of cognitive achievement in children: Home, school, and racial test score gaps," *Journal of Human capital*, Vol. 1, No. 1, pp. 91–136.

A Proofs

Proposition 2

Proof. From Proposition 1 it follows that $\mathbb{E}[v(t,p)|x,s]$ is increasing in s for any p, and moreover $\mathbb{E}[v(t,p_2)|x,s] - \mathbb{E}[v(t,p_1)|x,s]$ is also increasing in s. Consider any category x and signal s. For any category x', since $\mathbb{E}[v(t,p_2)|x',s] - \mathbb{E}[v(t,p_1)|x',s]$ is continuous and increasing in s, either $\exists s'$ such that $\mathbb{E}[v(t,p_2)|x',s'] - \mathbb{E}[v(t,p_1)|x',s'] = \mathbb{E}[v(t,p_2)|x,s] - \mathbb{E}[v(t,p_1)|x,s]$, or $\mathbb{E}[v(t,p_2)|x',s'] - \mathbb{E}[v(t,p_1)|x',s'] > \mathbb{E}[v(t,p_2)|x,s] - \mathbb{E}[v(t,p_1)|x,s]$ for all s', or $\mathbb{E}[v(t,p_2)|x',s'] - \mathbb{E}[v(t,p_1)|x',s'] < \mathbb{E}[v(t,p_2)|x,s] - \mathbb{E}[v(t,p_1)|x,s]$ for all s'. In all 3 cases, the condition for comparability is satisfied for all p, p', hence x and x' are comparable. $\square$

Proposition 3

Proof. Let $v(t,p) = v_1(t) \cdot v_2(p)$ and consider any family of signal structures $\{F_{t,x}\}$. Consider two categories, x and x', and an arbitrary signal s. Then

$$\mathbb{E}[v(t,p)|x,s] = \mathbb{E}[v_1(t)|x,s] \cdot v_2(p).$$

Consider $\{\mathbb{E}[v_1(t)|x',s] : s \in S\}$, i.e. the range of possible values $\mathbb{E}[v_1(t)|x',s]$ over all s. Because $\mathbb{E}[v(t,p)|x,s]$ is continuous in s, this set is an interval. There are two cases.

Case 1: $\mathbb{E}[v_1(t)|x,s] \in \{\mathbb{E}[v_1(t)|x',s] : s \in S\}$. Then $\exists s' \in S$ s.t. $\mathbb{E}[v_1(t)|x,s] = \mathbb{E}[v_1(t)|x',s']$. Hence $\exists s'$ s.t. $\forall p$:

$$\mathbb{E}[v(t,p)|x,s] = \mathbb{E}[v_1(t)|x,s] \cdot v_2(p) = \mathbb{E}[v_1(t)|x',s'] \cdot v_2(p) = \mathbb{E}[v(t,p)|x',s'].$$

Case 2: $\mathbb{E}[v_1(t)|x,s] \notin \{\mathbb{E}[v_1(t)|x',s] : s \in S\}$. If $\mathbb{E}[v_1(t)|x,s] <$ [resp. $>$] $\{\mathbb{E}[v_1(t)|x',s] : s \in S\}$, then $\forall s'$ and $\forall p$ we have

$$\mathbb{E}[v(t,p)|x,s] = \mathbb{E}[v_1(t)|x,s] \cdot v_2(p) < \text{[resp. } > \text{]}\, \mathbb{E}[v_1(t)|x',s'] \cdot v_2(p) = \mathbb{E}[v(t,p)|x',s']$$

In both cases we conclude that x and x' are strongly comparable. $\square$

Proposition 4

Proof. Consider two categories, x and x', and an arbitrary signal s.

$\{F_{t,x}\}$ and $\{F_{t,x'}\}$ are equivalent, so $\exists \delta$ s.t.

$$\widetilde{w}_{p,x}(v) = \widetilde{w}_{p,x'}(v) + \delta$$

for any p, for all $v \in \text{Im}[\widetilde{v}_{p,x}(s)] \cap \text{Im}[\widetilde{v}_{p,x'}(s)]$.

There are 2 cases to consider.

Case 1: $\text{Im}[\widetilde{v}_{p,x}(s)] \cap \text{Im}[\widetilde{v}_{p,x'}(s)] = \varnothing$. Because $\widetilde{v}_{p,x}(s)$ and $\widetilde{v}_{p,x'}(s)$ are continuous, the intermediate value theorem implies that either $\widetilde{v}_{p,x}(s) > \widetilde{v}_{p,x'}(s')$ $\forall s, s'$, or $\widetilde{v}_{p,x}(s) < \widetilde{v}_{p,x'}(s')$ $\forall s'$. This holds for any arbitrary p, hence part 2 of the definition of strong comparability is satisfied.

Case 2: $\mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)] \neq \varnothing$. Take any $v \in \mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)]$, and take δ s.t.

$$\widetilde{w}_{p,x}(v) = \widetilde{w}_{p,x'}(v) + \delta.$$

Let $s' = s - \delta$. Then

$$\mathbb{E}[v(t,p)|x,s] = \mathbb{E}[v(t,p)|x', s-\delta] = \mathbb{E}[v(t,p)|x', s']$$

holds $\forall v \in \mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)]$. Hence part 1 of the definition of strong comparability is satisfied for all such v.

Next, consider $v < \mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)]$. Because $\widetilde{v}_{p,x}(s)$ is continuous and increasing in s, this implies that $\forall s' \in S$ we have $\mathbb{E}[v(t,p)|x',s'] > \mathbb{E}[v(t,p)|x,s]$ or $\mathbb{E}[v(t,p)|x',s'] < \mathbb{E}[v(t,p)|x,s]$, $\forall p$. Analogously, the same holds for $v > \mathrm{Im}[\widetilde{v}_{p,x}(s)] \cap \mathrm{Im}[\widetilde{v}_{p,x'}(s)]$. Hence part 2 of the definition of strong comparability is satisfied for all such v.

In both cases we conclude that x and x' are strongly comparable. $\square$

Theorem 1

Proof. Consider any positions p, p' with $p > p'$ and let

$$r_{p,p'}(x,s) = \mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s]$$

be the index of an agent with observables (x,s).

Note that $r_{p,p'}(\cdot)$ induces a total preorder on $X \times S$, with totally ordered equivalence classes, denoted by C_l. Then (x_i,s_i) and (x_j,s_j) belong to the same equivalence class iff $r_{p,p'}(x_i,s_i) = r_{p,p'}(x_j,s_j)$, and $(x_i,s_i) \in C_l$ and $(x_j,s_j) \in C_{l'}$ with $l < l'$ iff $r_{p,p'}(x_i,s_i) < r_{p,p'}(x_j,s_j)$.

Furthermore, for any positions $\hat{p}$ and $\hat{p}'$ with $\hat{p} > \hat{p}'$, let $r_{\hat{p},\hat{p}'}(x,s) = \mathbb{E}[v(t,\hat{p})|x,s] - \mathbb{E}[v(t,\hat{p}')|x,s]$ be an alternative index, which also induces a total preorder on $X \times S$. Because all categories x and x' are comparable, by definition we have that $r_{p,p'}(x_i,s_i) \geq r_{p,p'}(x_j,s_j)$ if and only if $r_{\hat{p},\hat{p}'}(x_i,s_i) \geq r_{\hat{p},\hat{p}'}(x_j,s_j)$. Hence the equivalence classes defined by $r_{p,p'}(\cdot)$ and $r_{\hat{p},\hat{p}'}(\cdot)$ are the same, and total preorders induced by $\mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s]$ are invariant with respect to (p,p'). So without loss of generality let the index be defined by

$$r(x,s) = \mathbb{E}[v(t,p_m)|x,s] - \mathbb{E}[v(t,p_{m-1})|x,s].$$

First, I show that in any deterministic policy $\mathscr{P}$, if agents are not assigned to policies in weakly increasing order of their index, the policy is suboptimal.

Consider any (x_i, s_i) and (x_j, s_j) with $r(x_i, s_i) < r(x_j, s_j)$. Suppose a policy $\mathscr{P}$ assigns $\mathbb{P}(x_j, s_j) = p'$ and $\mathbb{P}(x_i, s_i) = p''$ for some $p' < p''$.

Since $r(x_i, s_i) < r(x_j, s_j)$, we have $\mathbb{E}[v(t, p'')|x_i, s_i] - \mathbb{E}[v(t, p')|x_i, s_i] < \mathbb{E}[v(t, p'')|x_j, s_j] - \mathbb{E}[v(t, p')|x_j, s_j]$. Switching the assignment of agents i and j, holding all other assignments fixed, produces another feasible policy, with strictly higher total surplus, since

$$\mathbb{E}[v(t, p'')|x_i, s_i] + \mathbb{E}[v(t, p')|x_j, s_j] < \mathbb{E}[v(t, p'')|x_j, s_j] + \mathbb{E}[v(t, p')|x_i, s_i],$$

where the left-hand side is the i and j surplus under $\mathscr{P}$, and the right-hand side is the i and j surplus under the alternative policy.

Second, I show that there is no gain from randomization of non-index-equivalent agents within each position.

Consider any policy that randomizes among non-equivalent agents. Take the largest position p'' such that $q_i \equiv \mathbb{P}[p(x_i, s_i) = p''] \in (0, 1)$ and $q_j \equiv \mathbb{P}[p(x_j, s_j) = p''] \in (0, 1)$, for some agents (x_i, s_i) and (x_j, s_j) with $r(x_i, s_i) < r(x_j, s_j)$. Optimality requires that agents are matched to a position with probability 1, so $\exists$ another position $p' < p''$ s.t. $q'_j \equiv \mathbb{P}[p(x_j, s_j) = p'] \in (0, 1)$. Let $\hat{q} = \min\{q_i, q_j, q'_j\}$, and consider an alternative policy that moves a probability mass $\hat{q}$ of agent j's assignment from position p' to position p'', and a probability mass $\hat{q}$ of agent i's assignment from position p'' to position p'. Such a policy is feasible, holding all else constant. Because $r(x_i, s_i) < r(x_j, s_j)$, we have $\mathbb{E}[v(t, p'')|x_i, s_i] - \mathbb{E}[v(t, p')|x_i, s_i] < \mathbb{E}[v(t, p'')|x_j, s_j] - \mathbb{E}[v(t, p')|x_j, s_j]$. Hence the alternative policy yields strictly larger expected total surplus, since

$$\hat{q}\,\mathbb{E}[v(t, p'')|x_i, s_i] + \hat{q}\,\mathbb{E}[v(t, p')|x_j, s_j] < \hat{q}\,\mathbb{E}[v(t, p'')|x_j, s_j] + \hat{q}\,\mathbb{E}[v(t, p')|x_i, s_i],$$

where the left-hand side is the relevant i and j surplus under $\mathscr{P}$, and the right-hand side is the relevant i and j surplus under the alternative policy. Therefore the optimal policy cannot randomize among agents with different indices who are assigned to the same position.

Hence the optimal policy is weakly increasing in $r(x, s)$ and assigns agents assortatively with respect to $r(x, s)$, up to randomization among r-equivalent agents. $\square$

Theorem 2

Proof. Proving sufficiency is straight-forward: if the signals $\{F_{t,x}\}$ are identical with respect to $v(\cdot)$, then for any $(k_1, ..., k_m)$ the optimal policy $\mathscr{P}^*$ can be implemented with category-invariant cutoffs.

Suppose $\{F_{t,x}\}$ are identical with respect to $v(\cdot)$. Then $\widetilde{w}_{p,x}(v) = \widetilde{w}_{p,x'}(v)$ for any x, x', hence $\widetilde{v}_{p,x}(s) = \widetilde{v}_{p,x'}(s)$ for any x, x'. Therefore $r(x, s) = r(x', s)$ for any x, x'. Hence the optimal policy $\mathscr{P}^*$ has $\mathbb{P}^*(x, s) = \mathbb{P}^*(x', s)$, and hence $\bar{s}_p(x) = \bar{s}_p(x')$.

Next, I prove necessity by contrapositive: if the signals $\{F_{t,x}\}$ are not identical, then there exist $(k_1, ..., k_m)$ such that the optimal policy $\mathscr{P}^*$ is implemented with cutoffs that vary across categories.

Suppose $\{F_{t,x}\}$ are not identical. Then for some x', x'' and p, $\exists v^* \in \text{Im}[\widetilde{v}_{p,x'}(s)] \cap \text{Im}[\widetilde{v}_{p,x''}(s)]$ such that $\widetilde{w}_{p,x'}(v^*) \neq \widetilde{w}_{p,x''}(v^*)$. Note that $\widetilde{v}_{p,x}(s)$ is continuous, so there exists a neighborhood around $\widetilde{w}_{p,x'}(v^*)$ and around $\widetilde{w}_{p,x''}(v^*)$ such that $\widetilde{w}_{p,x'}(v) \neq \widetilde{w}_{p,x''}(v)$ in the neighborhood. Consider the largest such neighborhood of signals, denoted (s_{min}, s_{max}), where $\widetilde{v}_{p,x'}(s) \neq \widetilde{v}_{p,x''}(s)$ for any $s \in (s_{min}, s_{max})$. W.l.o.g. suppose $\widetilde{v}_{p,x'}(s) > \widetilde{v}_{p,x''}(s)$ and $r(x', s) > r(x'', s)$ for all $s \in (s_{min}, s_{max})$.

Let $k_m = 1 - F_{t,x'}(s_{min}) + \sum_{x \neq x'}[1 - F_{t,x}(s_{max})]$.

Then by Theorem 1, for $s \in (s_{min}, s_{max})$ the optimal policy $\mathscr{P}^*$ assigns $p^*(x', s) = p_m$, while $p^*(x'', s) < p_m$. Hence $\mathscr{P}^*$ is implemented by cutoffs $\bar{s}_{p_m}(x') < \bar{s}_{p_m}(x'')$, i.e. the optimal cutoffs vary across categories.

Finally, the above implies that in particular if $\{F_{t,x}\}$ are equivalent, but not identical, then there exist capacities such that the optimal cutoffs vary across categories. □

Theorem 3

Proof. Suppose $\{F_{s,x}\}$ is such that $\mathbb{E}[t|x, s]$ is constant in x, for every s, and for some x', x'', $F(t|x'', s)$ is a mean-preserving spread of $F(t|x', s)$.

Suppose $v(t, p)$ has convex [strictly convex] differences in t. Then

$$\begin{aligned} r(x'', s) &= \mathbb{E}[v(t, p_m)|x'', s] - \mathbb{E}[v(t, p_{m-1})|x'', s] \\ &= \int_t v(t, p_m) - v(t, p_{m-1}) dF(t|x'', s) \\ &\geq [>] \int_t v(t, p_m) - v(t, p_{m-1}) dF(t|x', s) \\ &= r(x', s) \end{aligned}$$

where the [strict] inequality follows from the fact that $v(t, p_m) - v(t, p_{m-1})$ is a convex function of t and $F(t|x'', s)$ is a mean-preserving spread of $F(t|x', s)$.

The proof that $r(x'', s) \leq [<] r(x', s)$ if $v(t, p)$ has concave [strictly concave] differences in t is analogous. □

Proposition 5

Proof. First, by Theorem 1, the optimal policy $\mathscr{P}^*$ is assortitative with respect to the index $r(x, s) \equiv \mathbb{E}[v(t, p_m)|x, s] - \mathbb{E}[v(t, p_{m-1})|x, s]$.

Second, $r(x, s) = \int_T [v(t, p_m) - v(t, p_{m-1})] dF(t|x, s)$ is increasing in s, because $v(t, p_m) - v(t, p_{m-1})$ is increasing in t, and under Assumption 3 $F(t|x, s)$ is FOSD-increasing in s.

Third, $r(x, s) = \int_T [v(t, p_m) - v(t, p_{m-1})] dF(t|x, s)$ is decreasing in x, because $v(t, p_m) - v(t, p_{m-1})$ is increasing in t, and under Assumption 4 $F(t|x, s)$ is FOSD-decreasing in x.

That is, for any s and $x'' > x'$, $r(x'', s) < r(x', s)$. Therefore $p^*(x', s) \geq p^*(x'', s)$, for any s (with strict inequality for some capacities). If $\mathscr{P}^*$ involves some randomization due to tie-breaking, this inequality is in terms of FOSD of the distributions $\mathbb{P}^*(p|x, s)$, and the support of $\mathbb{P}^*(p|x, s)$, as a function of (x, s), can be compared in the strong set order.

Hence for any position p,

$$\bar{s}_p(x'') \equiv \min\{s_i : \mathbb{P}^*(p|x'', s_i) > 0\} \geq \bar{s}_p(x') \equiv \min\{s_i : \mathbb{P}^*(p|x', s_i) > 0\},$$

hence $\bar{s}_p(x)$ is increasing in x, for any p. □

Proposition 6

Proof. Suppose $v(t, p)$ has convex differences in t (the proof with concave differences is analogous). By Theorem 3, under Assumption 5, $r(x'', s) \geq r(x', s)$ for all s. Therefore we have to consider 2 cases.

First, if $\mathscr{P}^*$ does not involve randomization due to tie-breaking, because $\mathscr{P}^*$ is monotonic in $r(x, s)$, we have that $p^*(x'', s) \geq p^*(x', s)$, with a strict inequality for some capacities and realizations of the observables.

Second, if $\mathscr{P}^*$ involves randomization due to tie-breaking, since (x'', s) and (x', s) are not payoff-equivalent, it must be the case that $|\operatorname{supp}[p^*(x'', s)] \cap \operatorname{supp}[p^*(x', s)]| \leq 1$, and moreover $\min\{\operatorname{supp}[p^*(x'', s)]\} \geq \max\{\operatorname{supp}[p^*(x', s)]\}$, and $\operatorname{supp}[p^*(x'', s)] \geq \operatorname{supp}[p^*(x', s)]$ in the strong set order. Then $p^*(x'', s)$ FOSD-dominates $p^*(x', s)$. Thus in both cases we have that $p^*(x'', s) \geq p^*(x', s)$, which proves the first part of the proposition.

Finally, consider the sets of agents assigned to any position p from each category. Since $p^*(x'', s) \geq p^*(x', s)$ for each s, we have that $\{s : \mathbb{P}^*(p|x'', s) > 0\} \leq \{s : \mathbb{P}^*(p|x', s) > 0\}$ in the strong set order, for any p. $\square$