Preliminary draft, not for circulation

Affirmative Action with Multidimensional Identities*

February 14, 2022

Jean-Paul Carvalho, *University of Oxford*
Bary Pradelski, *CNRS*
Cole Williams, *University of Vienna*

Abstract

Affirmative action policies are widely employed in college admissions, hiring, and other decisions to reduce underrepresentation of disadvantaged groups. Existing policies have a major limitation in that they treat each identity dimension (e.g. race, gender) as independent. We find that generically only *intersectional* policies, which account for interactions between identity dimensions, can eliminate underrepresentation among intersectional groups. Under certain conditions, every nonintersectional policy worsens representativeness for at least one intersectional group. Nonintersectional policies can eliminate underrepresentation along each identity dimension, despite preserving (and potentially disguising) underrepresentation among intersectional groups. We demonstrate how existing models of affirmative action can be extended to multidimensional identities and construct intersectional policies that eliminate underrepresentation.

Key words: Affirmative action, education, inequality, underrepresentation, identity

JEL classification: J7, I24, D02

Carvalho: Department of Economics and New College, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK, jean-paul.carvalho@economics.ox.ac.uk. *Pradelski*: Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France, bary.pradelski@cnrs.fr. *Williams*: Department of Economics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, cole.williams@univie.ac.at. We are grateful for comments and suggestions by Andrew Newman, Ludvig Sinander, and Peyton Young. All errors are ours. Williams acknowledges financial support from the Austrian Science Fund FWF under project number I 3487 and FG 6-G.

1 Introduction

This paper examines how multidimensional identities affect the design of affirmative action policies. We study a decision maker who must select a subset of applicants from an applicant pool. Examples include decisions regarding college admissions, hiring, lending, government contracting, and selection of electoral candidates by political parties. We depart from existing work by assuming each applicant has a multidimensional identity (e.g. race, gender, caste, socioeconomic class). We then show that standard affirmative action policies cannot eliminate underrepresentation when identities are multidimensional and construct policies that can do so.

Policies aimed at reducing the underrepresentation of disadvantaged groups have been employed in numerous countries, including caste-based quotas (reservation) in India, race and gender representation requirements in hiring, promotion, and procurement in South Africa, and reduced university entry requirements for ethnic minorities (*youhui zhengce*) in China. In the United States, affirmative action emerged from the Civil Rights movement and originally targeted racial discrimination. Title VII legislation expanded the set of protected categories, banning discrimination on the basis of race, skin color, religion, sex, and national origin. Protection for women was strengthened with Executive Order 11375 in 1967 and the Equal Employment Act of 1972. Today, the representation of groups defined by race, gender, and other identity characteristics is a major factor in college admissions, private and public sector hiring, government contracts, lending, and many other areas (Holzer and Neumark, 2000; Fryer and Loury, 2005). The incorporation of these characteristics into the decision-making process is meant to reduce bias in evaluating candidates, as well as adjust for socioeconomic disadvantages faced by groups (Chetty et al., 2014) and various forms of feedback through which underrepresentation can beget underrepresentation (Loury, 1977; Coate and Loury, 1993; Borjas, 1992; Athey et al., 2000).

Affirmative action policies have done much to reduce the underrepresentation of women and minorities in universities and the professions (e.g. Leonard, 1984; Bagde et al., 2016). Nevertheless, we show that existing affirmative action policies suffer from a major design flaw. Most, if not all, policies are formulated separately for each identity dimension (e.g. race and gender), often by different committees or organizations. At times, only one dimension is treated. Describing the European Union's gender policies, Skjeie (2015) states: "The dominant equality notion is mainly one-dimensional. What have recently been termed 'gender+' equality policies – i.e. policies which address gender inequalities in relation to other inequalities – are rather few and far between" [p.

79]. This failure to properly account for the multidimensionality of identity is not limited to policy. The economics literature on structural inequality focuses almost exclusively on unidimensional notions of identity (see reviews by Croson and Gneezy, 2009; Altonji and Blank, 1999; Fang and Moro, 2011). Conventions for collecting and reporting data are likewise reductive. Even where data are disaggregated based on identity categories such as race and gender, there is seldom information on the economic performance of the intersectional groups (e.g. black women). As multiple dimensions of identity are bundled in each person, there are connections between identity dimensions that cannot be neglected without producing analytical and policy errors. This paper examines the nature and severity of these errors.

The obvious reason for the reductive approach to assessing/addressing underrepresentation is simplicity. It is common, and often necessary, to reduce a complex problem to several parts. Problems arise, however, when the connections between the parts are neglected (Saari, 2015, 2018), as when affirmative action policies fail to account for connections between different dimensions of identity. This point has been long understood by scholars outside of economics; it is the central theme of the literature on intersectionality (e.g. Crenshaw, 1989). However, the multidimensionality of identity is largely unexplored by economists.

In our model, a decision maker must select a subset of applicants from an applicant pool. Each applicant has a score (e.g. test score) and a multidimensional identity. Score distributions can vary across intersectional groups due to socioeconomic disadvantages, bias, and other factors. We conceptualize an affirmative action policy as follows. The decision maker adjusts each applicant's score as a function of their identity and then accepts every applicant with an adjusted score above some threshold level. A nonintersectional policy is one in which scores are adjusted independently along each identity dimension (e.g. race, sex). This is the conventional way of formulating affirmative action policies by admissions committees, employers, lenders, and other decision makers facing such screening problems. In contrast, an intersectional policy applies a potentially different adjustment for each intersectional group.

We ask the following fundamental question: Can a nonintersectional policy achieve a *representative outcome* in which each intersectional group is represented according to its population share? We find that, generically, nonintersectional policies cannot do so, whereas intersectional policies can (Section 3.1). Nonintersectional policies can only achieve a representative outcome in special environments, where inequality/bias has a nonintersectional structure, i.e., is independent across identity dimensions (Section 3.2). Due to negative spillovers across identity dimensions, they can

fail to generate monotone improvements in representativeness, i.e., improve the representativeness of every intersectional group (Section 3.3). Nonintersectional policies can, however, achieve a *reductive representative outcome* in which there is proportional representation along every identity dimension (Section 3.4). If data are gathered in this reductive manner, it could thus give the false impression that structural inequality has been eliminated, whereas some intersectional groups continue to be underrepresented. Finally, we show how our analysis can be incorporated into existing work by extending an important paper on the design of affirmative action policies by Fryer and Loury (2013) to multidimensional identities (Section 3.5).

1.1 Related Literature

There are two related (and overlapping) strands of literature in economics. The first deals with the causes of intergroup inequality, including the socioeconomic environment (Chetty et al., 2014), taste-based discrimination (Becker, 1957), statistical discrimination (Phelps, 1972; Arrow, 1973; Chambers and Echenique, 2021), intergenerational transfers of human capital (Becker and Tomes, 1979; Loury, 1977, 1981; Borjas, 1992), norms (Akerlof and Kranton, 2000; Young, 2015; Bertrand et al., 2015; Eguia, 2017), learning (Chung, 2000; Fernández, 2013), peer effects, and local complementarities in education (Borjas, 1992; Benabou, 1993; Chaudhuri and Sethi, 2008). The second deals with affirmative action policies for reducing intergroup inequality (e.g. Coate and Loury, 1993; Loury, 2009; Goldin and Rouse, 2000; Fershtman and Pavan, 2021). Coate and Loury (1993) analyze the effects of affirmative action under statistical discrimination, famously showing that a 'patronizing equilibrium' can arise in which groups have proportional representation but one (disadvantaged) group has lower levels of skill formation. Fryer and Loury (2013) analyze the conditions under which it is efficient to grant disadvantaged minorities preferential access to positions rather than subsidize skill development. When affirmative action policies can be written based on identity (sighted), preferential access is more efficient. We extend Fryer and Loury's model, showing how to construct such policies when individuals have multidimensional identities. There is also an emerging interdisciplinary literature on algorithmic fairness which deals with reducing bias in machine learning and algorithmic decision-making (e.g. Kleinberg et al., 2016; Kleinberg and Raghavan, 2018; Kleinberg et al., 2018; Chouldechova and Roth, 2018; Rambachan and Roth, 2019; Raji et al., 2020). In particular, Kleinberg et al. (2016) provide an impossibility theorem in which three fairness conditions for algorithmic classification of individuals cannot be jointly achieved.

Much has been learned from this body of work. However, these analyses treat identity as unidimensional, whereas human identity is a higher-dimensional object describing one's race, gender, class, and many other characteristics. Notable exceptions are described below. The simple extension of affirmative action policies derived through undimensional analysis to a multidimensional setting is to independently formulate an intervention along each identity dimension and then check for proportional representation along each dimension (e.g. race, gender). This is the current practice. But this approach does not properly account for the multidimensionality of identity, because it ignores interactions between identity dimensions and neglects the fundamental unit of analysis when identity is multidimensional: the intersectional group.

In a recent article, Small and Pager (2020) encourage economists studying discrimination to draw on approaches from sociology and other disciplines, especially the notion of institutional discrimination. This paper draws on the concept of intersectionality introduced by Crenshaw (1989) in a critique of the unidimensional notions of identity that dominated legal doctrine and politics around anti-discrimination. Based on the unique experiences of black women, Crenshaw (1989) argued that an individual's experience is not the sum of their race and gender. Intersectionality has been an influential approach to studying discrimination and structural inequality outside of economics (see Cooper, 2016; Collins and Bilge, 2020). Collins and Bilge (2020) define the approach as follows: "As an analytical tool, intersectionality views categories of race, class, gender, sexuality, nation, ability, ethnicity, and age—among others—as interrelated and mutually shaping one another" [p. 1].

Through our analysis of affirmative action with multidimensional identities, we arrive at a mathematical characterization of the problems with unidimensional notions of identity and the gains from switching to the intersectional group as the unit of analysis. In economics, there are few examples of work on multidimensional identity. These include Sen (2006) on how drawing from multidimensional identities can reduce conflict, Meyer and Strulovici (2012) on comparing economic outcomes when inequality is multidimensional, Akerlof (2017) on how individuals choose to value different dimensions of their identity, and Elu and Loubert (2013) on how returns to schooling in sub-Saharan Africa depend on the interaction between gender and ethnicity. To our knowledge, the first analysis of intersectional policies in economics is by Carvalho and Pradelski (2021) who study a specific inequality-generating mechanism and show that subsidies along one identity dimension will alter representation along other identity dimensions. They also characterize systems of intersectional self-financing subsidies and role-model policies that achieve representative outcomes.

Finally, Mehrotra et al. (2022) study a model of selection for a specific inequality-generating mechanism (multiplicative bias) and show that a particular policy, i.e., nonintersectional minimal quotas, cannot achieve efficiency.

Our paper reveals that the interaction across identity dimensions poses a far more general problem. We demonstrate that *all* nonintersectional policies fail to achieve a representative outcome for an arbitrary number of identity dimensions and for generic inequality-generating mechanisms. We also show how to construct intersectional policies that achieve a representative outcome in this more general environment.

2 The Model

Consider a decision maker (e.g. college, employer) who must select a subset of applicants from an applicant pool. The applicant pool has unit mass and the decision maker accepts a share $\alpha \in (0,1)$ of the applicants and rejects the rest. Each applicant has a score x belonging to an open interval $X \subset \mathbb{R}$ and a multidimensional identity described by a vector of group characteristics $g \in G = \{0,1\}^n$, with $n \geq 2$.[1] While an individual's full identity is an n-dimensional object $g = (g_i)_{i=1}^n$, we can also express an individual's identity in a reductive manner in terms of one identity dimension: all individuals with entry $g_j = 1$ belong to category j (e.g. all women). We employ the natural partial order over groups whereby $g' \leq g$ if and only if $g'_i \leq g_i$ for each entry $i = 1, ..., n$. The joint distribution over characteristics and scores p is assumed to belong to the subspace $P \subset \Delta(X \times G)$ for which the conditional probabilities $p(\cdot|g)$ are mutually absolutely continuous with respect to the Lebesgue measure and we endow P with the weak* topology. We denote the marginal probability of belonging to group g, $p(X \times \{g\})$, simply by $p(g)$.

With the goal of achieving a more representative accepted class (e.g. student body, employee pool), the decision maker sets a *policy* $\mathbf{q} : X \times G \to X$ mapping an applicant's score and identity group characteristics to an adjusted score $y = \mathbf{q}(x, g)$. Let q_g denote $\mathbf{q}(\cdot, g)$ for a given identity g. We impose structure on the space of policies $\mathbf{Q}$ by supposing there to be a family of increasing bijections $Q \subset X^X$ such that $\mathbf{q} \in \mathbf{Q}$ if and only if $q_g \in Q$ for all $g \in G$. We require the set of functions to be *rich*, meaning that for $x, y \in X$ there exists a function $q \in Q$ satisfying $q(x) = y$, and *commutative* whereby $q, q' \in Q$ implies $q \circ q' = q' \circ q$. Two exemplar policy spaces are the following.

[1] Examples in this paper involve binary characteristics for illustration only. Our coding can accommodate more realistic non-binary characteristics by interpreting each entry as an indicator variable for a characteristic.

- **Additive**. $X = \mathbb{R}$ and $\mathbf{q} \in \mathbf{Q}$ if and only if $\mathbf{q}(x, g) \equiv x + \delta_g$ for some $\delta_g \in \mathbb{R}$.
- **Multiplicative**. $X = \mathbb{R}_{>0}$ and $\mathbf{q} \in \mathbf{Q}$ if and only if $\mathbf{q}(x, g) \equiv \theta_g \cdot x$ for some $\theta_g \in \mathbb{R}_{>0}$.

Given the policy, the decision maker sets an acceptance rule $f : X \to \{0, 1\}$ which is a monotone function mapping *adjusted* scores to the decision to either accept $f(y) = 1$ or reject $f(y) = 0$ an applicant, subject to the capacity constraint $E(f(y)) = \alpha$.[2]

Definition 1. A policy is **nonintersectional** if $g = g' + g''$ implies $q_g = q_{g'} \circ q_{g''}$. Otherwise, a policy is **intersectional**.

A nonintersectional policy treats each identity dimension as independent: applying q_g to the scores for members of group g is the same as iteratively applying q_{e_i} for each category i to which they belong, where e_i denotes the ith standard basis vector. For example, if the policy is additive then $\delta_g = \sum_{i=1}^{n} g_i \cdot \delta_{e_i}$ and if it is multiplicative then $\theta_g = \prod_{i=1}^{n} \theta_{e_i}^{g_i}$. Observe that a nonintersectional policy normalizes $q_{\mathbf{0}}(x) = x$ for all $x \in X$ since $g = g + \mathbf{0}$ implies $q_g = q_g \circ q_{\mathbf{0}}$. Appendix B offers a more general definition of nonintersectionality and proves that this normalization comes without loss in generality.

Example 1. Consider a simplified example of multidimensional identity: male $g = (0, \cdot)$, female $g = (1, \cdot)$, white $g = (\cdot, 0)$, black $g = (\cdot, 1)$. Suppose the policy is additive, boosting the scores for females by a and that of black individuals by b:

$$\delta_{(0,0)} = 0, \;\; \delta_{(1,0)} = a,$$
$$\delta_{(0,1)} = b, \;\; \delta_{(1,1)} = a + b.$$

This policy is nonintersectional. If instead the policy additionally lifts the scores of black females by $c \neq 0$ so that $\delta_{(1,1)} = a + b + c$, then the policy is intersectional.

3 Results

Rather than studying a decision maker who admits the most qualified applicants within certain representation constraints, we ask an even more fundamental question: Can proportional representation of identity groups even be achieved with the nonintersectional policies currently in use?

[2] In our construction, an individual's score x is adjusted based on their identity and then a (uniform) acceptance rule is applied to each adjusted score y. Equivalently, a different acceptance rule could be applied to *unadjusted* scores for each group g. Hence our results apply to all acceptance rules, not just policies for adjusting scores.

For a given policy and decision rule, let $p(g|f=1)$ denote the probability that an individual belongs to intersectional group g given that they have been accepted, i.e. $p(g|f=1) = \frac{E(f(y)|g)p(g)}{E(f(y))}$.

Definition 2. An outcome is *representative* if the representation of each intersectional group is equal to its population share: $p(g|f=1) = p(g)$ for each $g \in G$.

3.1 Nonintersectional policies do not eliminate underrepresentation

We demonstrate that the inherent constraints on nonintersectional policies prevent their achieving a representative state. We employ the topological notion of genericity whereby a property is generic of a set if it holds on a dense open subset.

Theorem 1. *For generic distributions $p \in P$:*

(a) There exists an intersectional policy that yields a representative outcome.

(b) There does not exist a nonintersectional policy that yields a representative outcome.

We relegate the technical details to Lemmas A.1 and A.2 in Appendix A. The rest of the proof, including the construction of intersectional policies, is presented here.

Proof of Theorem 1. We begin with part (b). For a given policy, the acceptance rule constitutes a cutoff x^* for which $f(y) = 1$ if and only $y \geq x^*$. The ensuing outcome is representative if

$$\left(p(q_g(x) \geq x^*|g) - \alpha\right) p(g) = 0 \text{ for all } g \in G. \tag{1}$$

Suppose $p(g) > 0$ for all $g \in G$. Let $F_g(x)$ be the CDF of a type g's score and define $\beta_g \equiv F_g^{-1}(1-\alpha)$ to be the $(1-\alpha)$th score quantile. Condition (1) becomes

$$q_g(\beta_g) = x^* \text{ for all } g \in G. \tag{2}$$

Let $\mathbf{0}$ and $\mathbf{1}$ be the vector of zeros and ones respectively. For a representative nonintersectional policy, condition (2) requires $q_{\mathbf{0}}(\beta_{\mathbf{0}}) = \beta_{\mathbf{0}} = x^*$ which pins down the admissions rule. The condition further requires $q_{e_i}(\beta_{e_i}) = \beta_{\mathbf{0}}$ for all $1 \leq i \leq n$. When policies are rich and commutative, this uniquely determines each q_{e_i} (see Lemma A.1). Condition (2) also requires $q_{\mathbf{1}}(\beta_{\mathbf{1}}) = (q_{e_1} \circ \cdots \circ q_{e_n})(\beta_{\mathbf{1}}) = \beta_{\mathbf{0}}$ or equivalently $\beta_{\mathbf{1}} = (q_{e_n}^{-1} \circ \cdots \circ q_{e_1}^{-1})(\beta_{\mathbf{0}})$. For $n \geq 2$, the set of distributions for which this

equality fails to hold and $p(g) > 0$ for all $g \in G$ is open and dense in P (see Lemma A.2). This proves part (b).

Turning to part (a), for a given x^* the richness assumption provides the existence of a policy satisfying (2) for every $p \in P$. Such a policy evidently satisfies the capacity constraint since $E(f(y)) = E(E(f(y)|g))$ and $E(f(y)|g) = \alpha$ for all $g \in G$. From the conclusion of part (b), such a policy must be intersectional on an open dense subset of P. □

Hence an affirmative action policy can achieve a representative outcome when written on the basis of the intersectional groups. The system of equations (2) shows precisely how to construct an intersectional policy that eliminates underrepresentation: the scores of each intersectional group must be adjusted so that they are equal at the $(1-\alpha)$th quantile. We demonstrate that generically this cannot be achieved by a nonintersectional policy based reductively on the identity dimensions such as race, sex, and socioeconomic class. Note that the space of permissible nonintersectional policies is large and far more general than the additive and multiplicative policies used as examples above. For example, if $h : X \to (0, 1)$ is any continuous and strictly increasing function (e.g. a continuous cdf), then the functions $h^{-1}(h(x)^a)$ with $a \in \mathbb{R}_{>0}$ form a rich and commutative family and can be used to construct a policy on X. There are many other nonintersectional policies that follow a simple functional form, as well as ones taking even more complicated forms that would be difficult to describe. What Theorem 1 says is that, regardless of the family of functions used to construct the policies Q, as long as the policy can be sensibly applied nonintersectionally (independently across identity dimensions), then generically the outcome will not be representative.

We can also show that there does not exist a sequence of nonintersectional policies that comes arbitrarily close to a representative outcome. The representativeness induced by a policy can be described by the vector $\rho = (\rho_g)_{g\in G} \equiv (|p(g|f = 1) - p(g)|)_{g\in G}$, where $\rho \in Z \subset \mathbb{R}^n$. We can define a measure of representativeness as follows: $W : Z \to \mathbb{R}$. All we require is that W is continuous, strictly increasing $\frac{\partial W}{\partial \rho_g} > 0$ for all $g \in G$, and $W(\mathbf{0}) = 0$. Then the following result proceeds from Theorem 1:

Corollary 1. *Generically, for each distribution p, there exists a threshold $\bar{W} > 0$ such that $W(\rho) > \bar{W}$ whenever the policy is nonintersectional.*

The proof is in Appendix A. Hence, generically, the outcome of every nonintersectional policy is bounded away from a representative outcome.

3.2 When do nonintersectional policies perform well?

A deeper point revealed in the proof of Theorem 1 is that a nonintersectional policy can only achieve a representative outcome if the score distributions themselves have a specific "nonintersectional" relationship. The result can be stated as follows in its most general, albeit abstract, form:

Proposition 1. *A nonintersectional policy achieves a representative outcome if and only if the function $q \in Q$ mapping $q(\beta_g) = \beta_{g'}$ also maps $q(\beta_{g-g'}) = \beta_{\mathbf{0}}$ for all groups $g' \leq g$.*

Proof. First, assume condition (2) holds so that $q_g(\beta_g) = q_{g'}(\beta_{g'})$ for any two groups. If additionally $g' \leq g$, then $q_g = q_{g'} \circ q_{g-g'}$, implying $q_{g-g'}(\beta_g) = \beta_{g'}$. From (2) and the normalization of $q_{\mathbf{0}}$, $q_{g-g'}(\beta_{g-g'}) = q_{\mathbf{0}}(\beta_{\mathbf{0}}) = \beta_{\mathbf{0}}$. Because the function $q \in Q$ mapping $q(\beta_g) = \beta_{g'}$ is unique (Lemma A.1), it is equal to $q_{g-g'}$ and the desired conclusion holds.

Now, assume $q(\beta_g) = \beta_{g'}$ implies $q(\beta_{g-g'}) = \beta_{\mathbf{0}}$ for all groups $g' \leq g$. Then defining the functions $q_g(\beta_g) = \beta_{\mathbf{0}}$ for all $g \in G$, we want to show that these functions collectively define a nonintersectional policy. To prove this, take any two groups with $g' \leq g$ and observe that because $q_g(\beta_g) = q_{g'}(\beta_{g'})$ we have $q_{g'}^{-1} \circ q_g(\beta_g) = \beta_{g'}$. By assumption, $q_{g'}^{-1} \circ q_g(\beta_{g-g'}) = \beta_{\mathbf{0}}$ and by definition $q_{g-g'}(\beta_{g-g'}) = \beta_{\mathbf{0}}$. As the function $q \in Q$ mapping $q(\beta_{g-g'}) = \beta_{\mathbf{0}}$ is unique (Lemma A.1) we have $q_{g'}^{-1} \circ q_g = q_{g-g'}$ and thus $q_g = q_{g'} \circ q_{g-g'}$, implying that the policy $\mathbf{q}$ defined by $\mathbf{q}(x, g) = q_g(x)$ for all $x \in X$ and $g \in G$ is nonintersectional. □

To interpret this condition, consider the case of sex and race as coded in Example 1. The condition in Proposition 1 means, among other things, that the same adjustment required to equalize the scores of black females and white females at the $(1-\alpha)$th quantile is required to equalize the scores of black males and white males at the $(1-\alpha)$th quantile: $q(\beta_{(1,1)}) = \beta_{(1,0)}$ implies $q(\beta_{(0,1)}) = \beta_{(0,0)}$. That is, *differences in scores based on race are independent of sex*. Therefore, for a nonintersectional policy to achieve a representative outcome, the score distributions must exhibit a kind of independence across identity dimensions. The following example further illustrates this point.

Example 2 (Human Capital and Bias)**.** Suppose that each individual's human capital $\hat{x}$ is drawn independently from a normal distribution with mean $\hat{\mu}$. The decision maker interprets scores in a biased manner on the basis of an individual's group affiliation. Specifically, an individual with human capital $\hat{x}$ is ascribed a *biased score* $x = \hat{x} - b$, where the bias b is drawn independently for a member of group g from a normal distribution with mean μ_g and a variance that is common to all groups. Thus, the biased scores for members of group g are normally distributed with mean

$\hat{\mu} - \mu_g$ and variance denoted by σ^2. Letting Φ denote the standard normal distribution, then the $(1-\alpha)$th quantile for each biased score distribution is defined as the number β_g satisfying $\Phi\left(\frac{\beta_g - \hat{\mu} + \mu_g}{\sigma}\right) = 1 - \alpha$. Simplifying this equation, we obtain $\beta_g = \Phi^{-1}(1-\alpha)\,\sigma + \hat{\mu} - \mu_g$.

Supposing the decision maker uses an additive policy, each $q \in Q$ can be expressed as $q(x) = x + \delta$ for some real number δ. From this and Theorem 1, if a policy $\mathbf{q}$ achieves a representative outcome, then $q_g(\beta_g) = \beta_g + \delta_g$ is constant across groups, which is equivalent to requiring $\delta_g - \mu_g = \delta_{g'} - \mu_{g'}$ for all $g, g' \in G$. Moreover, if the policy is nonintersectional, then for any $g' \leq g$ we have $\delta_g = \delta_{g'} + \delta_{g-g'}$. Taken together, these conditions provide the following equalities

$$\mu_g - \mu_{g'} = \delta_g - \delta_{g'} = \delta_{g-g'} = \mu_{g-g'} - \mu_{\mathbf{0}}.$$

Thus, the bias to which groups are subject must itself take a specific additive, nonintersectional form. This is summarized by the following result.

Corollary 2. *For Example 2, a nonintersectional policy achieves a representative outcome if and only if $g = g' + g''$ implies $\mu_g + \mu_{\mathbf{0}} = \mu_{g'} + \mu_{g''}$ for all groups.*

3.3 Monotone Improvements

We know that nonintersectional policies generically cannot achieve a representative outcome. We now weaken the requirement and ask whether there are nonintersectional policies that can at least improve representativeness for all intersectional groups. Let $p^0(g|f=1)$ be the representation of each group g in the absence of any policy. A group is *overrepresented* without a policy if $p^0(g|f=1) > p(g)$ and *underrepresented* if $p^0(g|f=1) < p(g)$. We can then define an improvement in representativeness as follows:

Definition 3. A *monotone improvement* occurs if $|p(g|f=1) - p(g)| < |p^0(g|f=1) - p(g)|$ for all $g \in G$.

Hence, a monotone improvement in representativeness is one that brings all identity groups closer to proportional representation. What we find is that a nonintersectional policy can only yield a monotone improvement in representativeness if underrepresentation/overrepresentation among intersectional groups has a kind of monotonic structure.

Proposition 2. *Suppose groups $\mathbf{0}$ and g are overrepresented and there is a group $g' \leq g$ such that g' and $g - g'$ are underrepresented. Then no nonintersectional policy yields a monotone improvement.*

Proof. Toward a contradiction, suppose $\mathbf{q}$ is nonintersectional and delivers a monotone improvement, groups $\mathbf{0}$ and g are overrepresented, and groups g' and $g'' = g - g'$ are underrepresented. Before introducing a policy, the acceptance rule admits a student if and only if their score x exceeds the value x_0^* equating $\sum_{g \in G} p(g)(1 - F_g(x_0^*)) = \alpha$. After introducing the policy, the admissions cutoff shifts to x^* equating $\sum_{g \in G} p(g)(1 - F_g(q_g^{-1}(x^*))) = \alpha$. A monotone improvement requires $q_{\tilde{g}}(x_0^*) < x^*$ for any overrepresented group and $q_{\tilde{g}}(x_0^*) > x^*$ for any underrepresented group. The new cutoff must exceed the initial one $x^* > x_0^*$ as group zero is overrepresented and $q_{\mathbf{0}}(x_0^*) = x_0^*$. Thus we have

$$q_{g'}(x_0^*) > x^* > x_0^*$$

which, because $q_{g''}$ is increasing implies

$$q_{g''}(q_{g'}(x_0^*)) > q_{g''}(x^*) > q_{g''}(x_0^*).$$

Because $q_g = q_{g'} \circ q_{g''}$ and g'' is underrepresented the preceding inequalities imply $q_g(x_0^*) > q_{g''}(x_0^*) > x^*$. But then the policy cannot generate a monotone improvement since g is overrepresented, a contradiction. □

To make the above condition concrete, return to Example 2 with two-dimensional identities, $n = 2$. Suppose groups $\mathbf{0}$ and $\mathbf{1}$ are overrepresented at the expense of groups $(1, 0)$ and $(0, 1)$ who are underrepresented, for instance when $\mu_{(1,0)} = \mu_{(0,1)} < \mu_{\mathbf{0}} = \mu_{\mathbf{1}}$. Then there is no nonintersectional policy that reduces underrepresentation of group $(0, 1)$ without increasing underrepresentation of group $(1, 0)$, and vice versa. The reason is that nonintersectional policies fail to account for negative spillovers across identity dimensions which can rule out monotone improvements.

There are plausible conditions under which the ordering of representation in Proposition 2 holds. For example, in the Dutch parliament in 2013 white males and minority females were overrepresented, while white females and minority males were underrepresented (Celis et al., 2014). In many cases, minority females face a double disadvantage. But one reason why the ordering in Proposition 2 (and the Dutch parliament) can arise is the measurement of representation in a reductive manner, based on the identity dimensions (e.g., race, gender) and not the intersectional groups. The selection of a minority female increases representativeness along two dimensions (race/ethnicity and gender) and is thus a double improvement according to measures of representation based on the identity dimensions (see Muegge and Erzeel, 2016). However, this still leaves underrepresentation at the intersectional level. We turn our attention to this issue in the following subsection.

3.4 Reductive Representation and Hidden Inequality

Structural inequality is an important concept because inequality structured by race, sex, and other identity characteristics often goes unnoticed when focusing on aggregate measures of income and wealth inequality. When identities are multidimensional, the natural unit of analysis is the intersectional group. Accordingly, we have defined a representative outcome as proportional representation across intersectional groups. However, this is not the standard measure in current practice. Admissions and hiring committees tend to reduce the dimensionality of the problem and pursue the following (nonintersectional) objective of proportional representation across identity dimensions:

Definition 4. A *reductive representative outcome* is one in which $p(g_i = 1|f = 1) = p(g_i = 1)$ for each identity dimension $i = 1, \ldots, n$.

We now ask whether a nonintersectional policy can at least achieve this lower-dimensional objective.

Theorem 2. *For every $p \in P$, there is a nonintersectional policy that achieves a reductive-representative outcome.*

Proof. The goal is to show that there is a nonintersectional policy $\mathbf{q}$ and an acceptance rule f achieving a reductive-representative outcome, $p(g_i = 1|f = 1) = p(g_i = 1)$ for all $i = 1, ..., n$, subject to the capacity constraint $p(f = 1) = \alpha$. From Bayes' rule

$$p(g_i = 1|f = 1) = \frac{p(f = 1|g_i = 1)p(g_i = 1)}{p(f = 1)}.$$

Thus, to prove Theorem 2 it is necessary and sufficient to find a nonintersectional policy and acceptance rule so that the participation of each dimension in the support is equal to the population participation rate, i.e.

$$p(f = 1|g_i = 1) = p(f = 1) = \alpha, \text{ for all } i = 1, ..., n \text{ with } p(g_i) > 0.$$

We first make a few observations and introduce helpful notation.

Claim 1. *It is without loss of generality to assume scores belong to* $(0, 1)$.

To see why this is true, consider the primitives X and Q and let $h : (0, 1) \to X$ be strictly increasing and continuous. Then let $\tilde{Q}$ be the set of functions on $(0, 1)$ defined by $\tilde{q}(x) = h^{-1}(q(h(x))$ for

each $q \in Q$. Each $\tilde{q}$ takes a score $x \in (0,1)$, maps it to X via h, adjusts it according to q, and then maps it back to $(0,1)$ via h^{-1}. Note, assuming scores to belong to $(0,1)$ simplifies notation, but is not essential to the argument.

Claim 2. *Each function $q \in Q$ can be continuously extended to a function $[0,1] \to [0,1]$ by defining $q(0) = 0$ and $q(1) = 1$. Let Q' denote the continuous extensions of the functions in Q.*

As each $q \in Q$ is an increasing bijection from $(0,1)$ to itself, it is clear that $\lim_{x \to 0^+} q(x) = 0$ and $\lim_{x \to 1^-} q(x) = 1$. Notice that, although the functions in Q' are no longer commutative, the composition $q \circ q'$ is well-defined and continuous for all q and q' in Q'.

Claim 3. *Let $\bar{Q} = Q' \cup \{0,1\}$ where 0 and 1 denote the constant functions $x \mapsto 0$ and $x \mapsto 1$ respectively. $\bar{Q}$ can be parameterized by an isomorphism $q \mapsto q(\cdot|\theta)$ such that*

1. *$\theta \in [0,1]$*
2. *$q(x|\theta)$ is continuous in θ for all $x \in [0,1]$ and strictly increasing in θ for all $x \in (0,1)$.*
3. *$q(x|0) = 0$ and $q(x|1) = 1$ for all $x \in [0,1]$.*

The proof of this claim follows from defining $\Theta : \bar{Q} \to [0,1]$ by $\Theta(q) = q(1/2)$ and then setting $q(\cdot|\theta)$ equal to the function $q \in \bar{Q}$ satisfying $\Theta(q) = \theta$. From Lemma A.1 part (b), it is clear that each function $q \in \bar{Q}$ is assigned a distinct parameter. Also, $\bar{Q}$ holds the richness property on $[0,1]$ and so for each parameter $\theta \in [0,1]$ there is a $q \in \bar{Q}$ satisfying $\theta = \Theta(q)$. As $\theta' > \theta$ is equivalent to $q(1/2|\theta') > q(1/2|\theta)$, Lemma A.1 implies that $q(x|\theta') > q(x|\theta)$ for all $x \in (0,1)$; thus $q(x|\theta)$ is strictly increasing in θ for all $x \in (0,1)$. The richness property also provides that the function $q(x|\theta)$ is continuous in θ. Finally, the functions satisfying $q(1/2) = 0$ and $q(1/2) = 1$ are the constant functions $x \mapsto 0$ and $x \mapsto 1$, respectively.

We now turn to the main argument. Denote $\boldsymbol{\theta} = (\theta_i)_{i=0}^n$ with $\theta_i \in [0,1]$ for all $i = 0, 1, ..., n$. Set $g_0 = 1$ whenever it appears below. For $i = 0, 1, ..., n$, define

$$q^i(\cdot|g, \boldsymbol{\theta}) = q\left(\cdot|g_i\theta_i + (1 - g_i)\frac{1}{2}\right)$$

$$q^{-i}(\cdot|g, \boldsymbol{\theta}) = \underbrace{q^0(\cdot|g, \boldsymbol{\theta}) \circ q^1(\cdot|g, \boldsymbol{\theta}) \circ ... \circ q^n(\cdot|g, \boldsymbol{\theta})}_{\text{composition of all } j \neq i}.$$

Let $q(\cdot|i,g,\boldsymbol{\theta})$ denote the composition $q^i(q^{-i}(\cdot|g,\boldsymbol{\theta})|g,\boldsymbol{\theta})$. Fix $x^* \in (0,1)$ and define

$$\bar{R}_i(\boldsymbol{\theta}) = \sum_{g \in G} \Pr\left(q(x|i,g,\boldsymbol{\theta}) \geq x^*|g\right) \Pr(g|g_i = 1) \text{ for } i = 0,1,...,n. \tag{3}$$

When $\boldsymbol{\theta}$ is interior, $\bar{R}_i$ is equal to the participation rate for members of category $i = 1,...,n$ and $\bar{R}_0$ is equal to the total participation rate.

Notice that $q(x|i,g,\theta_i = 0,\boldsymbol{\theta}_{-i}) = 0$ and $q(x|i,g,\theta_i = 1,\boldsymbol{\theta}_{-i}) = 1$ for all $\boldsymbol{\theta}_{-i} \in [0,1]^n$ and $x \in [0,1]$. Thus, $\bar{R}_i(\theta_i = 0,\boldsymbol{\theta_{-i}}) = 0$ and $\bar{R}_i(\theta_i = 1,\boldsymbol{\theta}_{-i}) = 1$ for all $\boldsymbol{\theta}_{-i} \in [0,1]^n$. Since each $\bar{R}_i(\boldsymbol{\theta})$ is continuous, we can apply the Poincaré-Miranda Theorem[3] which provides that for a given $\alpha \in (0,1)$ there exists a value $\boldsymbol{\theta}^* \in [0,1]^{n+1}$ such that $\bar{R}_i(\boldsymbol{\theta}^*) = \alpha$ for all $i = 0,1,...,n$.

To complete the proof, note that each θ_i^* must lie in the interior $(0,1)$ since $\bar{R}_i(\boldsymbol{\theta}^*) = \alpha \in (0,1)$ for all $i = 0,1,...,n$. Therefore $q(\cdot|\theta_i^*) \in Q'$ and also the restriction $q(\cdot|\theta_i^*)|_{(0,1)} \in Q$. Thus, the nonintersectional policy $\mathbf{q}$ with $q_{e_i} = q(\cdot|\theta_i^*)|_{(0,1)}$ for $i = 1,...,n$ and the threshold $x^{**} = q^{-1}(x^*|\theta_0^*)$ achieve a reductive-representative outcome and satisfy the capacity constraint. □

While Theorem 2 may lend some support to nonintersectional policies, it raises the problem of hidden inequality. Suppose an admissions/hiring committee designs a nonintersectional policy to achieve a reductive representative outcome, while gathering data on underrepresentation along each identity dimension. It is conventional to gather and analyze data in such a nonintersectional manner. On this basis, the committee might conclude that underrepresentation has been eliminated. But this is only true for each identity dimension. According to Theorem 1, generically, at least one intersectional group will remain underrepresentated. Thus, the reductive approach to structural inequality only goes part of the way. Moreover, it can create a false impression of having eliminated structural inequality.

3.5 Fryer and Loury (2013) with Multidimensional Identities

An important paper by Fryer and Loury (2013) studies the design of affirmative action policies in a two-stage environment, which we shall describe below. We will show how to extend their analysis to multidimensional identities and apply our results. In doing so, we alter their notation to clarify the connection to our work. Among other things, this provides an example of how the score distributions in our analysis can be made at least partially endogenous.

[3]For a simple statement of the Poincaré-Miranda Theorem, see Theorem 1 in Fonda and Gidoni (2016).

In stage 1, individuals decide whether or not to invest in skills, $s \in \{0, 1\}$. An individual's cost of acquiring skills is a draw from a probability distribution which depends on their social identity $i \in G = \{A, B\}$. Let $H_g(c)$ and $h_g(c)$ be the cost distribution and density for members of identity group $g \in G$ taking full support on the same interval for all groups. Group B is disadvantaged in the sense that $h_B(c)/h_A(c)$ is strictly increasing in c. In stage 2, each individual's productivity x is realized. Given skill s, the distribution of productivity is given by the distribution $F_s(x)$ with support $\mathbb{R}$. We assume F_1 to have first order stochastic dominance over F_0. That is, the likelihood of a high productivity draw is higher for those who invested in skills at stage 1. To simplify the exposition, further assume the support of the cost distributions H_g contains the interval $[0, E(x|s = 1) - E(x|s = 0)]$. After observing their productivity, individuals can purchase one of a fixed number of production opportunities (slots) at the market-clearing price x^*. Under *laissez-faire*, all individuals with $x \geq x^*$ will purchase a slot and produce. Because group B is disadvantaged in skill acquisition at stage 1, it will be underrepresented among those with production opportunities at stage 2. Fryer and Loury (2013) show that when *sighted* affirmative action policies are permitted, the most efficient policy that achieves a representative outcome is one which subsidizes the purchase of slots by members of the disadvantaged group at stage 2.[4]

Our model has the same deep structure as the Fryer-Loury model. Rather than individuals purchasing slots if their productivity satisfies $x \geq x^*$, with the productivity distributions varying between groups, a decision maker accepts applicants whose scores satisfy $x \geq x^*$, with the score distributions varying among groups. The subsidies to groups in their model are the same as the score adjustments in ours. Hence we can extend the Fryer-Loury model to multidimensional identities, $g \in G = \{0, 1\}^n$, and apply our results. Following the original formulation, let the affirmative action policy be additive. If a member of group g buys a slot, they receive a payment equal to their productivity, minus the cost x^*, net of the subsidy/tax δ_g, i.e. $x + \delta_g - x^*$. Normalizing the outside option to zero, such an individual purchases a slot if and only if $x + \delta_g - x^* \geq 0$.

- An individual in group g with skill s buys a slot with probability $\Pr(x \geq x^* - \delta_g|s) = 1 - F_s(x^* - \delta_g)$.
- The expected payoff to acquiring skills, $s = 1$, is $\int_{x^*-\delta_g}^{\infty} (x + \delta_g - x^*) \mathrm{d}F_1(x) - c$.
- The expected payoff to not acquiring skills, $s = 0$, is $\int_{x^*-\delta_g}^{\infty} (x + \delta_g - x^*) \mathrm{d}F_0(x)$.

[4]When policies are constrained to be *blind* (cannot be written on the basis of identity), then the most efficient policy that achieves a representative outcome is one which subsidizes skill acquisition at stage 1.

To simplify notation, let $t_g = x^* - \delta_g$ denote the threshold that the productivity of members of group g must exceed to buy a slot. As in Fryer and Loury (2013), we can write the benefit to skill formation as

$$\begin{aligned} B(t_g) &= \int_{t_g}^{\infty} (x - t_g) \mathrm{d}F_1(x) - \int_{t_g}^{\infty} (x - t_g) \mathrm{d}F_0(x) \\ &= \int_{t_g}^{\infty} (1 - F_1(x)) \mathrm{d}x - \int_{t_g}^{\infty} (1 - F_0(x)) \mathrm{d}x \\ &= \int_{t_g}^{\infty} (F_0(x) - F_1(x)) \mathrm{d}x. \end{aligned}$$

Note $\frac{\partial}{\partial t_g} B(t_g) = -(F_0(t_g) - F_1(t_g)) < 0$.

Given cost c, a group g member invests in skills if and only if $B(t_g) \geq c$. The probability that they invest in skills is therefore $H_g(B(t_g))$. The participation rate of group g members, i.e. the share of group g members who purchase a slot, is therefore:

$$\begin{aligned} R_g(t_g) &= H_g(B(t_g))(1 - F_1(t_g)) + (1 - H_g(B(t_g)))(1 - F_0(t_g)) \\ &= H_g(B(t_g))(F_0(t_g) - F_1(t_g)) + 1 - F_0(t_g). \end{aligned}$$

Computing the change in participation as a result of changing t_g, we obtain

$$\begin{aligned} \frac{\partial R_g}{\partial t_g} &= -h_g(B(t_g))(F_0(t_g) - F_1(t_g))^2 + H_g(B(t_g))(f_0(t_g) - f_1(t_g)) - f_0(t_g) \\ &= -h_g(B(t_g))(F_0(t_g) - F_1(t_g))^2 - (1 - H_g(B(t_g)))f_0(t_g) - H_g(B(t_g))f_1(t_g) < 0. \end{aligned}$$

Noting that $\lim_{t_g \to -\infty} F_s(t_g) = 0$ and $\lim_{t_g \to \infty} F_s(t_g) = 1$, we can also see that $\lim_{t_g \to -\infty} R_g(t_g) = 1$ and $\lim_{t_g \to \infty} R_g(t_g) = 0$. Thus, there exists a unique t_g^* satisfying $R_g(t_g^*) = \alpha$ for each $g \in G$. To achieve a representative outcome, consider the policy with $\delta_{\mathbf{0}} = 0$, the threshold x^* equal to $t_{\mathbf{0}}^* \equiv R_{\mathbf{0}}^{-1}(\alpha)$, and with $\delta_g = R_{\mathbf{0}}^{-1}(\alpha) - R_g^{-1}(\alpha)$ for all g. With such a policy, a member of group g buys a slot with probability α for each group $g \in G$ and thus the capacity constraint is also satisfied.

Now consider nonintersectional policies which requires $\delta_g = \delta_{g'} + \delta_{g''}$ for all groups g, g', and g'' satisfying $g = g' + g''$. Recall that such a policy normalizes $\delta_{\mathbf{0}} = 0$ so that the cutoff x^* must equal $t_{\mathbf{0}}^*$. A nonintersectional policy achieves a representative outcome if and only if $g = g' + g''$ implies

$$t_g^* = x^* - \delta_g = x^* - \delta_{g'} - \delta_{g''} = t_{g'}^* + t_{g''}^* - t_{\mathbf{0}}^*.$$

By the same reasoning as Theorem 1, this equality generically does not hold.

Formally, to connect with Theorem 1, we treat the distributions F_s as fixed and let P be the set of joint distributions over group identity and investment costs such that $p(\cdot|g) \sim H_g$ is continuous. Then the conclusion of Theorem 1 holds. In particular, if subsidies are calculated independently for each identity dimension, this will generically leave one or more intersectional groups underrepresented. To eliminate all underrepresentation, the subsidies will have to be intersectional, i.e. computed separately for each intersectional group. Precisely how to compute the intersectional subsidies is given by (2) in the proof of Theorem 1 and in this specific case by the system derived above: $\delta_g = R_{\mathbf{0}}^{-1}(\alpha) - R_g^{-1}(\alpha)$ for all $g \in G$. Thus, our analysis shows how existing work can be extended to examine the effects of multidimensional identities.

4 Conclusion

The economics literature on intergroup inequality and affirmative action treats identity as unidimensional. This paper has shown that when identities are multidimensional, structural inequality cannot be eliminated using conventional nonintersectional policies. Nonintersectional policies may not even achieve an approximately representative outcome. But they can always achieve a reductive representative outcome in which underrepresentation is eliminated along each identity dimension. Of course, the simplicity argument for nonintersectional policies remains, as the number of intersectional groups grows rapidly in the number of identity dimensions. Our framework is flexible and can be extended in a number of directions, including new ways of making the score distributions endogenous. Our work also points to a potentially fruitful empirical research program on how multidimensional identities shape the evolution of intergroup inequality and the effectiveness of affirmative action policies.

A Appendix: Technical details for Theorem 1

Throughout, $Q \subset X^X$ is maintained to be a rich and commutative family of increasing bijections. Recall that Q is *rich* if for each pair $x, y \in X$ there is a function $q \in Q$ satisfying $q(x) = y$ and *commutative* if $q, q' \in Q$ implies $q \circ q' = q' \circ q$.

The first lemma characterizes several useful properties of the set Q.

Lemma A.1. *The family of functions Q holds the following properties.*

(a) Each $q \in Q$ is continuous.

(b) The function mapping x to y is unique for each $x, y \in X$.

(c) Q contains the identity function.

(d) Q is closed under composition: $q, q' \in Q$ implies $q \circ q' \in Q$.

(e) Q contains its inverses: $q \in Q$ implies $q^{-1} \in Q$.

Proof. (a) Each $q \in Q$ is continuous because it is an increasing bijection.

(b) Suppose there are two functions q and q' in Q satisfying $q(x_0) = q'(x_0)$ for some $x_0 \in X$. For any $x \in X$, richness provides the existence of a function q'' in Q satisfying $q''(x_0) = x$. Thus $q''(q(x_0)) = q''(q'(x_0))$ and commutativity implies $q(q''(x_0)) = q'(q''(x_0))$ and so $q(x) = q'(x)$ for all $x \in X$.

(c) To prove that Q contains an identity function, for a given $x_0 \in X$ there is a function $q \in Q$ satisfying $q(x_0) = x_0$. For an arbitrary $x \in X$ and a function $q' \in Q$ satisfying $q'(x_0) = x$

$$x = q'(x_0) = q'(q(x_0)) = q(q'(x_0)) = q(x)$$

and thus $q(x) = x$ for all $x \in X$.

(d) To verify that Q has the closure property, for any $q, q' \in Q$ and $x_0 \in X$, there is a function $q'' \in Q$ for which $q(q'(x_0)) = q''(x_0)$. For any $x \in X$ and $q''' \in Q$ for which $q'''(x) = x_0$ the commutativity property provides that $q'''(q(q'(x)) = q'''(q''(x))$ and thus $q(q'(x)) = q''(x)$ for all $x \in X$. Therefore, for every $q, q' \in Q$ there exists $q'' \in Q$ such that $q \circ q' = q''$.

(e) To show that Q contains its inverses, for a function $q \in Q$ and a point $x_0 \in X$, there is another function $q' \in Q$ satisfying $q'(q(x_0)) = x_0$. The closure property (d) implies $q \circ q' \in Q$ which must mean that $q \circ q'$ is the identity function as (b) and (c) provide that it is the unique function in Q admitting a fixed point. Thus, $q' = q^{-1}$. □

Observe that by taking together properties (c)-(e), the commutativity assumption, and the associativity of functional composition, $(Q, \circ)$ takes the form of an abelian group. Two additional conclusions follow immediately from this lemma. From (a) and (b), the functions in Q are ordered in the sense that $q(x) > q'(x)$ for some $x \in X$ implies $q(x') > q'(x')$ for all $x' \in X$. Since Q contains the identity function, this conclusion further implies that each function either increases scores, decreases scores, or leaves them constant, i.e. for all $q \in Q$ the sign of $q(x) - x$ is constant.

Lemma A.2. *Assume Q to be rich and commutative. The subset of p for which $\beta_{\mathbf{1}} \neq (q_{e_n}^{-1} \circ \cdots \circ q_{e_1}^{-1})(\beta_{\mathbf{0}})$ (with $q_{e_i}(\beta_{e_i}) = \beta_{\mathbf{0}}$ for $1 \leq i \leq n$) and $p(g) > 0$ for all $g \in G$ is open and dense in P.*

Proof. First, let us show that the subset $A = \left\{p \in P : \beta_{\mathbf{1}} = (q_{e_n}^{-1} \circ \cdots \circ q_{e_1}^{-1})(\beta_{\mathbf{0}})\right\}$ is closed. Toward a contradiction, let $\{p_\gamma\}_{\gamma\in\Gamma}$ be a net in A converging to $p \in A^c$. Let $(\beta_g)_{g\in G}$ be defined with respect to p and $(\beta_g^\gamma)_{g\in G}$ be defined with respect to p_γ for each $\gamma \in \Gamma$. By the definition of weak* convergence, for any $\epsilon > 0$,

$$F_g^\gamma(\beta_g - \epsilon) \to F_g(\beta_g - \epsilon) < 1 - \alpha,\ F_g^\gamma(\beta_g + \epsilon) \to F_g(\beta_g + \epsilon) > 1 - \alpha.$$

Hence, β_g^γ converges to β_g for all $g \in G$. Thus, there exists γ_0 such that $\gamma \geq \gamma_0$ implies $\beta_{\mathbf{1}}^\gamma \neq \sum_{i=1}^n \beta_{e_i}^\gamma - (n-1)\beta_{\mathbf{0}}^\gamma$ contradicting the assumption that $\{p_\gamma\}_{\gamma\in\Gamma}$ is a net in A. Furthermore, $B = \{p \in P : p(g) = 0 \text{ for some } g \in G\}$ is closed and thus $(A \cup B)^c$ is open.

Finally, to show $(A \cup B)^c$ is dense in P, let $p \in A \cup B$ and let $\{p_\gamma\}_{\gamma\in(0,1)}$ be a net with $F_g^\gamma = F_g$ for all $\gamma \in (0,1)$ and $g \neq \mathbf{1}$, $p_\gamma(g) = p(g)\gamma + p'(g)(1-\gamma)$ where $p'(g) > 0$ for all $g \in G$, $F_{\mathbf{1}}^\gamma = F_{\mathbf{1}}\gamma + F_{\mathbf{1}}'(1-\gamma)$ where $F_{\mathbf{1}}(\beta_{\mathbf{1}}) \neq F_{\mathbf{1}}'(\beta_{\mathbf{1}})$.[5] As $\{p_\gamma\}_{\gamma\in(0,1)}$ is a net in $(A \cup B)^c$ that converges to p, it follows that p is in the closure of $(A \cup B)^c$. As the choice of $p \in A \cup B$ was arbitrary, $(A \cup B)^c$ is dense in P. □

Proof of Corollary 1. Toward a contradiction, suppose there is a distribution p for which no nonintersectional policy achieves a representative outcome, but there is a sequence of nonintersectional policies $(\mathbf{q}^m)_m$ with corresponding score thresholds $(x^m)_m$ such that, for all $\epsilon > 0$ there is an index m_ϵ for which $W(\rho(\mathbf{q}^m, x^m)) < \epsilon$ if $m \geq m_\epsilon$. Continue to denote the $1-\alpha$th score quantile by $\beta_g = F_g^{-1}(1-\alpha)$ for each group $g \in G$. It must be that $x^m \to x^* \equiv \beta_{\mathbf{0}}$ or else $\rho_{\mathbf{0}}$ is bounded away from zero. Similarly, $q_{e_i}^m(\beta_{e_i}) \to x^*$ for each $i = 1, ..., n$ or else some ρ_{e_i} is bounded away from zero. But then, letting $\mathbf{q}$ be the nonintersectional policy satisfying $q_{e_i}(\beta_{e_i}) = x^*$ for all $i = 1, ..., n$, we have that $(\mathbf{q}^m, x^m) \to (\mathbf{q}, x^*)$, and thus the nonintersectional policy $\mathbf{q}$ achieves a representative outcome with score threshold x^*, a contradiction. □

B Appendix: Generalizing

In this appendix, we show that our definition of nonintersectional policies does not rely on the particular labels used for the groups. We generalize by defining a policy $\mathbf{q}$ to be **nonintersectional***

[5] Notice that P includes the family of normal distributions $\{\mathcal{N}(\mu, 1)\}_{\mu\in\mathbb{R}}$ for the conditional distributions $p(\cdot|g)$ and so we can find such an $F_{\mathbf{1}}'$.

if $q_g \circ q_{g'} = q_{g \vee g'} \circ q_{g \wedge g'}$[6]. First, we show how applying a normalization produces the simpler definition used in the text.

Proposition B.1 (Normalize)**.** *Let $g^* \in G$ be a group. If q is a nonintersectional* policy, then so is the policy r defined by $r_g \equiv q_{g^*}^{-1} \circ q_g$ for all $g \in G$.*

Proof.

$$\begin{aligned} r_g \circ r_{g'} &= q_{g^*}^{-1} \circ q_{g^*}^{-1} \circ q_g \circ q_{g'} = q_{g^*}^{-1} \circ q_{g^*}^{-1} \circ q_{g \vee g'} \circ q_{g \wedge g'} \\ &= r_{g \vee g'} \circ r_{g \wedge g'}. \end{aligned}$$

Notice that for group g^*, $r_{g^*}(x) = (q_{g^*}^{-1} \circ q_{g^*})(x) = x$ for all $x \in X$. Thus, if $g^* = \mathbf{0}$ then for $g' \leq g$, $r_{g - g'} \circ r_{g'} = r_g \circ r_{\mathbf{0}} = r_g$. □

Next, we show that if a policy is nonintersectional*, then it remains so if we relabel the groups.

Proposition B.2 (Relabel)**.** *Suppose the policy q is nonintersectional*. Let g^* be a group and relabel all groups according to the mapping $h(g) = |g - g^*|$. Then the policy r defined by $r_{h(g)} = q_g$ for all $g \in G$ is likewise nonintersectional*.*

Proof. First observe that from the definition of a nonintersectional* policy: $q_{g - g_i e_i} \circ q_{g_i e_i} = q_g \circ q_{\mathbf{0}}$ for $1 \leq i \leq n$. Repeated application of this observation yields

$$q_g = q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ q_{g_1 e_1}) \circ \cdots \circ (q_{\mathbf{0}}^{-1} \circ q_{g_n e_n}). \tag{4}$$

It is enough to prove that the claim is true for $g^* = e_i$ for $1 \leq i \leq n$. For simplicity of notation but without loss of generality, consider $g^* = e_1$. As $h(\mathbf{0}) = e_1$ and $h(e_1) = \mathbf{0}$, we have $r_{e_1} = q_{\mathbf{0}}$ and $r_{\mathbf{0}} = q_{e_1}$. For $i > 1$, $h(e_1 + e_i) = e_i$ and so

$$\begin{aligned} r_{e_i} = q_{e_1 + e_i} = q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ q_{e_1}) \circ (q_{\mathbf{0}}^{-1} \circ q_{e_i}) &= q_{\mathbf{0}}^{-1} \circ q_{e_1} \circ q_{e_i} \\ &= q_{\mathbf{0}}^{-1} \circ r_{\mathbf{0}} \circ q_{e_i} \end{aligned}$$

and so $q_{\mathbf{0}}^{-1} \circ q_{e_i} = r_{\mathbf{0}}^{-1} \circ r_{e_i}$. Denoting $h(g) = (h_1, \ldots, h_n)$, for $i > 1$, $h_i = g_i$ and $h_1 = 1 - g_1$. Furthermore, $q_{g_1 e_1} = r_{h_1 e_1}$. Combining these observations, (4) can be written

$$\begin{aligned} r_{h(g)} = q_g &= q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ r_{h_1 e_1}) \circ (r_{\mathbf{0}}^{-1} \circ r_{h_2 e_2}) \circ \cdots \circ (r_{\mathbf{0}}^{-1} \circ r_{h_n e_n}) \\ &= r_{\mathbf{0}} \circ (r_{\mathbf{0}}^{-1} \circ r_{h_1 e_1}) \circ (r_{\mathbf{0}}^{-1} \circ r_{h_2 e_2}) \circ \cdots \circ (r_{\mathbf{0}}^{-1} \circ r_{h_n e_n}). \end{aligned} \tag{5}$$

[6]Equivalently, $g = g' + g''$ implies $q_g \circ q_{\mathbf{0}} = q_{g'} \circ q_{g''}$

Since $h : G \to G$ is a bijection, for all groups $h' \in G$ there is a group $g \in G$ such that $h' = h(g)$; hence, $r_{h'}$ can be formulated as in (5). Using this formulation, it follows immediately that r is nonintersectional*. □

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753.

Akerlof, R. (2017). Value formation: The role of esteem. *Games and Economic Behavior 102*, 1–19.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. *Handbook of Labor Economics 3*, 3143–3259.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, pp. 3–33. Princeton, NJ: Princeton University Press.

Athey, S., C. Avery, and P. Zemsky (2000). Mentoring and diversity. *American Economic Review 90*(4), 765–786.

Bagde, S., D. Epple, and L. Taylor (2016). Does affirmative action work? Caste, gender, college quality, and academic success in India. *American Economic Review 106*(6), 1495–1521.

Becker, G. (1957). *The Economics of Discrimination*. Chicago, IL: Chicago University Press.

Becker, G. S. and N. Tomes (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy 87*(6), 1153–1189.

Benabou, R. J. (1993, August). Workings of a city: Location, education, and production. *Quarterly Journal of Economics 108*(3), 619–652.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *Quarterly Journal of Economics 130*(2), 571–614.

Borjas, G. J. (1992, February). Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics 107*(1), 123–150.

Carvalho, J.-P. and B. Pradelski (2021). Identity and underrepresentation: Interactions between race and gender. Working paper.

Celis, K., S. Erzeel, L. Muegge, and A. Damstra (2014). Quotas and intersectionality: Ethnicity and gender in candidate selection. *International Political Science Review 35*, 41–54.

Chambers, C. P. and F. Echenique (2021). A characterisation of 'Phelpsian' statistical discrimination. *The Economic Journal 131*(637), 2018–2032.

Chaudhuri, S. and R. Sethi (2008). Statistical discrimination with peer effects: Can integration eliminate negative stereotypes? *Review of Economic Studies 75*(2), 579–596.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *Quarterly Journal of Economics 129*(4), 1553–1623.

Chouldechova, A. and A. Roth (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*.

Chung, K.-S. (2000). Role models and arguments for affirmative action. *American Economic Review 90*(3), 640–648.

Coate, S. and G. C. Loury (1993). Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, 1220–1240.

Collins, P. H. and S. Bilge (2020). *Intersectionality*. John Wiley & Sons.

Cooper, B. (2016). Intersectionality. In L. Disch and M. Hawkesworth (Eds.), *The Oxford Handbook of Feminist Theory*, pp. 385–406. Oxford, UK: Oxford University Press.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 139–168.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–74.

Eguia, J. X. (2017). Discrimination and assimilation at school. *Journal of Public Economics 156*, 48–58.

Elu, J. U. and L. Loubert (2013). Earnings inequality and the intersectionality of gender and ethnicity in Sub-Saharan Africa: The case of Tanzanian manufacturing. *American Economic Review 103*(3), 289–292.

Fang, H. and A. Moro (2011). Theories of statistical discrimination and affirmative action: A survey. In M. O. J. Jess Benhabib and A. Bisin (Eds.), *Handbook of Social Economics*, pp. 133–200. Elsevier.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review 103*(1), 472–500.

Fershtman, D. and A. Pavan (2021). "soft" affirmative action and minority recruitment. *American Economic Review: Insights 3*(1), 1–18.

Fonda, A. and P. Gidoni (2016). Generalizing the poincaré–miranda theorem: the avoiding cones condition. *Annali di Matematica Pura ed Applicata (1923-) 195*(4), 1347–1371.

Fryer, R. G. and G. C. Loury (2005). Affirmative action in winner-take-all markets. *The Journal of Economic Inequality 3*(3), 263–280.

Fryer, R. G. and G. C. Loury (2013). Valuing diversity. *Journal of Political Economy 121*(4), 747–774.

Goldin, C. and C. Rouse (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American Economic Review 90*(4), 715–741.

Holzer, H. and D. Neumark (2000). Assessing affirmative action. *Journal of Economic Literature 38*(3), 483–568.

Kleinberg, J., J. Ludwig, S. Mullainathan, and C. R. Sunstein (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis 10*, 113–174.

Kleinberg, J., S. Mullainathan, and M. Raghavan (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

Kleinberg, J. and M. Raghavan (2018). Selection problems in the presence of implicit bias. *arXiv preprint arXiv:1801.03533*.

Leonard, J. S. (1984). The impact of affirmative action on employment. *Journal of Labor Economics 2*(4), 439–463.

Loury, G. C. (1977). A dynamic theory of racial income differences. In P. Wallace and A. LaMond (Eds.), *Women, Minorities, and Employment Discrimination*, pp. 86–153. D.C Heathand Co, Lexington MA.

Loury, G. C. (1981). Intergenerational transfers and the distribution of earnings. *Econometrica 49*(4), 843–867.

Loury, G. C. (2009). *The Anatomy of Racial Inequality*. Harvard University Press.

Mehrotra, A., B. Pradelski, and N. Vishnoi (2022). Selection in the presence of implicit bias: The advantage of intersectional constraints. Working paper.

Meyer, M. and B. Strulovici (2012). Increasing interdependence of multivariate distributions. *Journal of Economic Theory 147*(4), 1460–1489.

Muegge, L. M. and S. Erzeel (2016). Double jeopary or multiple advantage? Intersectionality and political representation. *Parliamentary Affairs 69*, 499–511.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–661.

Raji, I. D., T. Gebru, M. Mitchell, J. Buolamwini, J. Lee, and E. Denton (2020). Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 145–151.

Rambachan, A. and J. Roth (2019). Bias in, bias out? evaluating the folk wisdom. *arXiv preprint arXiv:1909.08518*.

Saari, D. G. (2015). Social science puzzles: A systems analysis challenge. *Evolutionary and Institutional Economics Review 12*(1), 123–139.

Saari, D. G. (2018). *Mathematics motivated by the social and behavioral sciences*. SIAM.

Sen, A. (2006). *Identity and Violence: The Illusion of Destiny*. New York, NY: W.W. Norton.

Skjeie, H. (2015). Gender equality and nondiscrimination: How to tackle multiple discrimination effectively. In F. Bettio and S. Sansonetti (Eds.), *Visions for Gender Equality*, pp. 79–82. Luxembourg: European Union.

Small, M. L. and D. Pager (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives 34*(2), 49–67.

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics 7*(1), 359–387.