Affirmative Action with Multidimensional Identities*

April 6, 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 predominantly treat each identity dimension (e.g., race, gender, caste) as independent. We find that generically such nonintersectional policies cannot eliminate underrepresentation. Under certain conditions, every nonintersectional policy worsens the representativeness of at least one intersectional group. Accounting for interactions between identity dimensions, we construct intersectional policies that achieve a representative outcome. Nonintersectional policies can, however, eliminate underrepresentation along each identity dimension, despite preserving (and potentially disguising) underrepresentation among intersectional groups.

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

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 Piotr Dworczak, Jon Eguia, Shengwu Li, Avital Livny, Andrew Newman, Paula Onuchic, 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 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). With multidimensional identities, we show that standard affirmative action policies cannot eliminate underrepresentation 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, gender, 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 reproduces itself (Loury, 1977; Borjas, 1992; Coate and Loury, 1993; 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 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).

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. 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.

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, gender). This is the conventional way of formulating affirmative action policies by admissions committees, employers, lenders, and other decision makers facing such selection 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., they can reduce representativeness for some intersectional groups (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. Recent advances include the design of reserve systems for the allocation of resources (Dur et al., 2018; Sönmez and Yenmez, 2022). 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 basic 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, genderuality, 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. In computer science, Flanigan et al. (2021) develop an algorithm for selecting citizens' assemblies when identities are multidimensional. Agents do not have scores; instead, the focus is on the trade-off between giving each member of the population an equal likelihood of being assigned to a panel and satisfying quotas defined in terms of the identity dimensions. 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 $g = (g_1, \dots, g_n)$, we can also express an individual's identity in a reductive manner in terms of one identity dimension: all individuals with entry $g_i = 1$ belong to category i (e.g., all women). 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 score distributions $F_g(\cdot) \equiv p(\cdot|g)$ are continuous and have full support on X. We denote the marginal probability of belonging to group g, $p(X \times \{g\})$, simply by $p(g)$.[2]

With the goal of achieving a more representative accepted class (e.g., student body, employee pool), the decision maker sets a *policy* $\mathbf{q} = (q_g)_{g \in G}$ such that $q_g : X \to X$ maps the score x of an applicant from group g to an adjusted score $q_g(x)$. We impose structure on the space of policies by supposing that there is a family of increasing bijections $Q \subset X^X$ such that $\mathbf{q} \in Q^{|G|}$. We require the set of functions to be (i) *rich*, meaning that for all $x, y \in X$ there exists a function $q \in Q$ satisfying

[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.

[2]We endow $\mathbb{R}$ with the usual topology, $\Delta(X \times G)$ with the weak* topology, and both $X \subset \mathbb{R}$ and $P \subset \Delta(X \times G)$ with their respective relative topologies.

$q(x) = y$, and (ii) *commutative*, meaning that $q, q' \in Q$ implies $q \circ q' = q' \circ q$. Two exemplar policy spaces are described by the following.

- **Additive**. For each $q \in Q$ there exists $\delta \in \mathbb{R}$ such that $q(x) = x + \delta$ for all $x \in X = \mathbb{R}$.
- **Multiplicative**. For each $q \in Q$ there exists $\theta \in \mathbb{R}_{>0}$ such that $q(x) = \theta \cdot x$ for all $x \in X = \mathbb{R}_{>0}$.

Given the policy, the decision maker sets an *acceptance threshold* $x^* \in X$ whereby applicants whose adjusted scores exceed the threshold $q_g(x) \geq x^*$ are accepted and all others $q_g(x) < x^*$ are rejected, subject to the capacity constraint $\sum_{g \in G} \Pr(q_g(x) \geq x^* | g) p(g) = \alpha$.[3]

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 C 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 women 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 women by $c \neq 0$ so that $\delta_{(1,1)} = a + b + c$, then the policy is intersectional.

[3]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 $q_g(x)$. 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.

3 Results

Our analysis answers the following questions: Can proportional representation of identity groups be achieved with nonintersectional policies (as is current practice)? If not generally, under what conditions? Can nonintersectional measures of inequality disguise or even worsen underrepresentation of some intersectional groups?

For a given policy and decision rule, let $\hat{p}(g)$ denote the probability that an individual belongs to intersectional group g given that they have been accepted, i.e., $\hat{p}(g) = \frac{\Pr(q_g(x) \geq x^* | g) p(g)}{\Pr(q_g(x) \geq x^*)}$.

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

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.[4]

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

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

(b) There exists an intersectional 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 (a). The ensuing outcome is representative if

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

Suppose $p(g) > 0$ for all $g \in G$. Recall that $F_g(x)$ is 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

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

[4]Recall that a subset $A \subset S$ is *dense* in a set S if its closure equals the set: $\bar{A} = S$ (see Aliprantis and Border, 2006).

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 (a).

Turning to part (b), 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 $\Pr(q_g(x) \geq x^*|g) = \alpha$ for all $g \in G$ and thus $\sum_{g \in G} \Pr(q_g(x) \geq x^*|g)p(g) = \alpha$. 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 designed on the basis of the intersectional groups. Condition (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, gender, 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 $\rho = (\rho_g)_{g \in G} \equiv (|\hat{p}(g) - p(g)|)_{g \in G}$, so that $\rho \in [0, 1]^n$. Letting $\|\cdot\|$ denote the Euclidean norm, the following result proceeds from Theorem 1:

Corollary 1. *Generically, for each distribution p, there exists a number $c > 0$ such that $\|\rho\| > c$ whenever the policy is nonintersectional.*

The proof is relegated to 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. This can be formalized as follows. Recall that $\beta_g \equiv F_g^{-1}(1-\alpha)$ is the $(1-\alpha)$th score quantile.

Definition 3. The environment exhibits *independence* across identity dimensions if the function $q \in Q$ mapping $q(\beta_g) = \beta_{g'}$ also maps $q(\beta_{g-g'}) = \beta_{\mathbf{0}}$ for all groups $g' \leq g$.[5]

To interpret this condition, consider the case of gender and race as coded in Example 1. The condition in Definition 3 means that *differences in scores based on race are independent of gender*. That is, the same adjustment required to equalize the scores of black women and white women at the $(1-\alpha)$th quantile is required to equalize the scores of black men and white men at the $(1-\alpha)$th quantile: $q(\beta_{(1,1)}) = \beta_{(1,0)}$ implies $q(\beta_{(0,1)}) = \beta_{(0,0)}$.

Proposition 1. *A nonintersectional policy can achieve a representative outcome if and only if the environment exhibits independence across identity dimensions.*

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

[5] As usual, for two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\mathbf{a} \leq \mathbf{b}$ if and only if $a_i \leq b_i$ for each dimension $i = 1, \ldots, n$.

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 $q_g(x)$ for all $x \in X$ and $g \in G$ is nonintersectional. $\square$

The following example illustrates:

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 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 x_0^* be the acceptance threshold and $\hat{p}_0(g) = \frac{(1-F_g(x_0^*))p(g)}{\sum_{g' \in G}(1-F_{g'}(x_0^*))p(g')}$ be the representation of group g in the absence of any policy. A group is *overrepresented* without a policy if $\hat{p}_0(g) > p(g)$ and *underrepresented* if $\hat{p}_0(g) < p(g)$. We can then define an improvement in representativeness as follows:

Definition 4. A *monotone improvement* occurs if $|\hat{p}(g) - p(g)| < |\hat{p}_0(g) - p(g)|$ for all $g \in G$.

Hence, a monotone improvement in representativeness is one that brings all intersectional groups closer to proportional representation. We find that a nonintersectional policy can only yield a monotone improvement in representativeness if underrepresentation/overrepresentation among intersectional groups follows a certain ordering, as implied by the following proposition.

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 men and minority women were overrepresented, while white women and minority men were underrepresented (Celis et al., 2014). While in many cases minority women face a double disadvantage, one reason why the ordering in Proposition 2 (and the Dutch parliament) can arise is because of prior affirmative action policies. If the measurement of representation is reductive, that is, based on the identity dimensions and not the intersectional groups, the selection of a minority woman increases representativeness along two dimensions (race/ethnicity and gender) and is thus a double improvement (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, gender, and other identity characteristics often goes unnoticed when focusing on aggregate measures of income and wealth inequality. When identities are multidimensional, the basic 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 5. A *reductive representative outcome* is one in which $\hat{p}(g_i) = p(g_i)$ for each identity dimension $i = 1, \ldots, n$.

We now ask whether a nonintersectional policy can at least achieve this reductive objective.

Theorem 2. *For a nonempty open subset of distributions in P, there exists a nonintersectional policy that achieves a reductive representative outcome.*

The proof is relegated to Appendix B. 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 will only be 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 and can create a false impression of having eliminated structural inequality. In fact, the problem could be even worse. A reductive representative policy could actually increase the underrepresentation of an underrepresented group. The following example illustrates.

Example 3. Building on Examples 1 and 2, consider a population in which ten percent of individuals are black, the remainder are white, and each racial group is evenly split between men and women. Suppose the biased scores for each group are normally distributed with a mean of zero for black and white women, a mean of 0.25 for black men, a mean of one for white men, and a variance of one for all groups. Assume that the decision maker has the capacity to admit half of the applicants.

Without an affirmative action policy, the decision maker would admit roughly a third of both black and white women, about 42 percent of black men, and about 70 percent of white men. Figure 1 plots the acceptance rates for women, black individuals, and black men for the nonintersectional policy lifting the scores of women by $a = 0.925\lambda$ and lifting the scores for black individuals by $b = 0.375\lambda$ so that the degree of intervention is measured by $\lambda \in [0, 1]$. When $\lambda = 1$ the policy achieves a reductive representative outcome and when $\lambda = 0$ there is no intervention.

While moving toward the reductive representative outcome has the desirable effect of increasing the representation for women and black individuals, only attending to these two dimensions hides the fact that the policy *reduces* the representation of black men. This is due to the rivalry inherent in representation. Even though black men are recipients of affirmative action, the increased acceptance of women leads the score admission threshold to increase by an even larger amount, crowding out the benefit to black men.

3.5 Fryer and Loury (2013) with Multidimensional Identities

Fryer and Loury (2013) study 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.

In stage 1, individuals decide whether or not to invest in skills, $s \in \{0, 1\}$. An individual's cost of

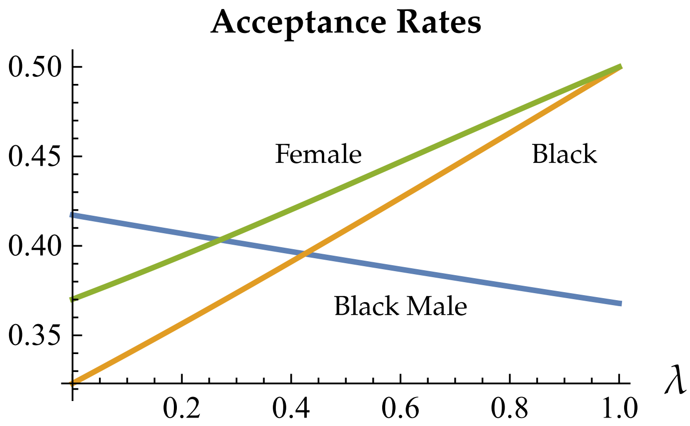

Figure 1: Group acceptance rates moving from no intervention ($\lambda = 0$) to the nonintersectional policy achieving a reductive representative outcome ($\lambda = 1$).

acquiring skills is a draw from a distribution which depends on their social identity $g \in G = \{A, B\}$. 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.[6]

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

[6]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.

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 $\int_{x^*-\delta_g}^{\infty}(x + \delta_g - x^*)\mathrm{d}F_0(x)$.

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 acceptance 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 acceptance 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 require $\delta_g = \delta_{g'} + \delta_{g''}$ 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 condition (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 incorporate 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 generically cannot be eliminated using conventional nonintersectional policies, even approximately. For an open set of conditions, a reductive representative outcome can be achieved in which underrepresentation is eliminated along each identity dimension. However, underrepresentation at the intersectional level will persist. 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 an empirical research program on how multidimensional identities shape the evolution of intergroup inequality and the effectiveness of affirmative action policies. A potentially fruitful way to proceed both theoretically and empirically is to examine the manipulability of identity, either through some form of "passing" or misrepresentation.[7] The design challenge would be to identify affirmative action policies that are robust to identity manipulation.

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.

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media.

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.

[7] See for example Cassan (2015) on the manipulation of caste identity to benefit from land redistribution. Note that the definition of identity groupings and the competition for policy attention is the theme of a vast literature in sociology, e.g., Brekhus et al. (2010); Schroer (2019).

Brekhus, W. H., D. L. Brunsma, T. Platts, and D. Priya (2010). On the contributions of cognitive sociology to the sociological study of race. *Sociology Compass 4*, 61–76.

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

Cassan, G. (2015). Identity-based policies and identity manipulation: Evidence from colonial Punjab. *American Economic Journal: Economic Policy 7*(4), 103–31.

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.

Dur, U., S. D. Kominers, P. A. Pathak, and T. Sönmez (2018). Reserve design: Unintended consequences and the demise of Boston's walk zones. *Journal of Political Economy 126*(6), 2457–2479.

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.

Flanigan, B., P. Gölz, A. Gupta, H. Brett, and A. D. Procaccia (2021). Fair algorithms for selecting citizens' assemblies. *Nature 596*, 548–552.

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.

Schroer, M. (2019). Sociology of attention: fundamental reflections on a theoretical program. In W. H. Brekus and G. Ignatow (Eds.), *The Oxford Handbook of Cognitive Sociology*. Oxford, UK: Oxford Handbooks.

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.

Sönmez, T. and M. B. Yenmez (2022). Affirmative action in India via vertical, horizontal, and overlapping reservations. *Econometrica*.

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

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, $q' = q^{-1}$. $\square$

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 $A = \{p \in P : \beta_{\mathbf{1}} = (q_{e_n}^{-1} \circ \cdots \circ q_{e_1}^{-1})(\beta_{\mathbf{0}})\}$ 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 $F_{\mathbf{1}}(\beta_{\mathbf{1}}) \neq F_{\mathbf{1}}'(\beta_{\mathbf{1}})$.[8] 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. $\square$

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 $\{\mathbf{q}_m\}$ with corresponding score thresholds $\{x_m\}$ such that, for all $\epsilon > 0$ there is an index m_ϵ satisfying $\|\rho(\mathbf{q}_m, x_m)\| < \epsilon$ if $m \geq m_\epsilon$. Continue to denote the $1-\alpha$th score quantile by

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

$\beta_g = F_g^{-1}(1-\alpha)$ 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, denoting $\mathbf{q}_m = (q_g^m)_{g\in G}$, it must also be that $q_{e_i}^m(\beta_{e_i}) \to x^*$ for each $i = 1, \dots, n$ or else some ρ_{e_i} is bounded away from zero. But then, letting $\mathbf{q}$ be the unique nonintersectional policy satisfying $q_{e_i}(\beta_{e_i}) = x^*$ for all $i = 1, \dots, n$, we have that $\sup_{x\in X} \|\mathbf{q}_m(x) - \mathbf{q}(x)\| \to 0$ and $x_m \to x^*$, and thus the nonintersectional policy $\mathbf{q}$ achieves a representative outcome with score threshold x^*, a contradiction. □

B Appendix: Theorem 2

To prove Theorem 2, we (i) restate the problem in simpler terms, (ii) provide a sufficient condition on a distribution $\bar{p}$ guaranteeing that each p in a neighborhood of $\bar{p}$ has a nonintersectional policy that achieves a reductive representative outcome, and (iii) give a simple example of one such distribution $\bar{p}$ satisfying the condition.

There are settings in which the nonintersectional policy achieving a reductive representative outcome is easily computed. Building on Example 1, suppose scores are normally distributed with mean μ_g, variance one, and the policy adds a to the scores of women and b to the scores of black individuals. Letting R_i denote the acceptance rate for $i = 1$ women and $i = 2$ black individuals, the total acceptance rate can be written simply as

$$S = \frac{1}{4}\left(1 - \Phi(x^* - \mu_{\mathbf{0}})\right) + \frac{1}{2}R_1 + \frac{1}{2}R_2 + \frac{1}{4}\left(1 - \Phi(x^* - \mu_{\mathbf{1}} - a - b)\right)$$

A reductive representative outcome requires $R_1 = R_2 = S = \alpha$. Rearranging the above expression, these equalities imply $\Phi(x^* - \mu_{\mathbf{0}}) = \Phi(x^* - \mu_{\mathbf{1}} - a - b)$ and thus $\mu_{\mathbf{0}} = \mu_{\mathbf{1}} + a + b$. Writing out the expressions for R_1 and R_2, one also finds that $\mu_{(1,0)} + a = \mu_{(0,1)} + b$. Thus the unique nonintersectional policy that achieves a reductive representative outcome is characterized by

$$a = \frac{\mu_{\mathbf{0}} - \mu_{(1,0)} + \mu_{(0,1)} - \mu_{(1,1)}}{2} \quad \text{and} \quad b = \frac{\mu_{\mathbf{0}} + \mu_{(1,0)} - \mu_{(0,1)} - \mu_{(1,1)}}{2}.$$

Inputting these values for a and b guarantees $R_1 = R_2 = S$ for all threshold values and thus a straightforward application of the intermediate value theorem provides that a unique threshold x^* equates each of these functions with α.

B.1 Proof of Theorem 2

It is useful to parameterize the functions in Q. To do this, fix some $x_0 \in X$ and let $r(\cdot|\theta)$ be the function $q \in Q$ for which $q(x_0) = \theta$. Using Lemma A.1, it is straightforward to prove that $r(x|\theta)$

is continuous in θ for all $x, \theta \in X$. Notice that the mapping $q \mapsto \theta$ represents an isomorphism: For each $q \in Q$ there is a unique $\theta \in \Theta$ satisfying $q(\cdot) = r(\cdot|\theta)$ and for each $\theta \in X$ there is a unique $q \in Q$ satisfying $r(\cdot|\theta) = q(\cdot)$.

If a policy is nonintersectional, then it is determined by the score adjustments q_{e_i} for $i = 1, ..., n$. Using our parameterization, a nonintersectional policy $\mathbf{q}$ can be characterized by the vector $\boldsymbol{\theta} \in X^n$ satisfying $q_{e_i} = r(\cdot|\theta_i)$ for all $i = 1, ..., n$. For a given group g, we can write the function adjusting its scores in terms of the parameterization explicitly as $r_g(\cdot|\boldsymbol{\theta}) = r(\cdot|\theta_1 \cdot g_1 + (1 - g_1) \cdot x_0) \circ \cdots \circ r(\cdot|\theta_n \cdot g_n + (1 - g_n) \cdot x_0)$. We can therefore write the acceptance rate for individuals belonging to dimension $i = 1, \ldots, n$ when the score threshold is x^* as

$$R_i(\boldsymbol{\theta}, x^*) = \sum_{g \in G} (1 - F_g(r_g^{-1}(x^*|\boldsymbol{\theta})))p(g|g_i = 1).$$

The total acceptance rate is

$$S(\boldsymbol{\theta}, x^*) = \sum_{g \in G} (1 - F_g(r_g^{-1}(x^*|\boldsymbol{\theta})))p(g).$$

The goal is to find a vector $(\boldsymbol{\theta}, x^*) \in X^{n+1}$ satisfying $R_i(\boldsymbol{\theta}, x^*) = \alpha$ for $i = 1, ..., n$ and $S(\boldsymbol{\theta}, x^*) = \alpha$.

Let $t(\boldsymbol{\theta})$ be the unique threshold satisfying $S(\boldsymbol{\theta}, t(\boldsymbol{\theta})) = \alpha$. Suppose that for a distribution $\bar{p} \in P$ there are two vectors $\mathbf{a}, \mathbf{b} \in X^n$ satisfying

$$R_i(a_i, \boldsymbol{\theta}_{-i}, t(a_i, \boldsymbol{\theta}_{-i})) < \alpha < R_i(b_i, \boldsymbol{\theta}_{-i}, t(b_i, \boldsymbol{\theta}_{-i})) \text{ for all } \boldsymbol{\theta}_{-i} \in \times_{j \neq i} [a_j, b_j]. \tag{3}$$

For example, the following are two natural conditions that guarantee (3) is satisfied.

1. There exist $\mathbf{a}, \mathbf{b} \in X^n$ for which $R_i(\mathbf{a}, t(\mathbf{a})) < \alpha < R_i(\mathbf{b}, t(\mathbf{b}))$ for all $i = 1, \ldots, n$.
2. $R_i(\boldsymbol{\theta}, t(\boldsymbol{\theta}))$ is decreasing in $\boldsymbol{\theta}_{-i}$ for all $\boldsymbol{\theta} \in [\mathbf{a}, \mathbf{b}]$.

When (3) holds, the Poincaré-Miranda Theorem provides that there exists a vector $\boldsymbol{\theta}^* \in [\mathbf{a}, \mathbf{b}]$ satisfying $R_i(\boldsymbol{\theta}^*, t(\boldsymbol{\theta}^*)) = \alpha$ for all $i = 1, \ldots, n$. Thus, the nonintersectional policy with $q_{e_i}(\cdot) = r(\cdot|\theta_i^*)$ for all $i = 1, \ldots, n$ and the threshold $x^* = t(\boldsymbol{\theta}^*)$ achieves a reductive representative outcome.[9]

To complete the proof, we show that for any distribution $\bar{p}$ satisfying (3), each p in a neighborhood of a $\bar{p}$ likewise satisfies (3) and then demonstrate the existence of a $\bar{p}$ satisfying (3).

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

Lemma B.1. *Suppose that for $\bar{p} \in P$ there exist $\mathbf{a}, \mathbf{b} \in X^n$ such that (3) is satisfied. Then (3) is satisfied by all p in a neighborhood of $\bar{p}$.*

Proof. Let us explicitly include the distribution p as an argument in the functions so that $t(\boldsymbol{\theta}, p)$ satisfies $S(\boldsymbol{\theta}, p, t(\boldsymbol{\theta}, p)) = \alpha$ when the distribution is p.

We can first verify that $t(\boldsymbol{\theta}, p)$ is continuous in $p \in P$ for all $\boldsymbol{\theta} \in X^n$ by noting that

$$\arg\max_{x^*[0,1]} \begin{cases} -\alpha^2 & \text{if } x^* = 0 \\ -(S(\boldsymbol{\theta}, p, x^*) - \alpha)^2 & \text{if } 0 < x^* < 1 \\ -\alpha^2 & \text{if } x^* = 1 \end{cases}$$

is a singleton and applying the Berge Maximum Theorem (see Aliprantis and Border, 2006, Theorem 17.31).

Next, since $A_{-i} = \times_{j \neq i}[a_i, b_i]$ is compact, by a second application of the Berge Maximum Theorem

$$m_i(\theta_i, p) = \max_{\boldsymbol{\theta}_{-i} \in A_{-i}} R_i\left(\boldsymbol{\theta}, p, t(\boldsymbol{\theta}, p)\right)$$

is continuous in θ_i and p. Because $m_i(a_i, \bar{p}) < \alpha < m_i(b_i, \bar{p})$ there is a neighborhood U_i of $\bar{p}$ such that, if p is in this neighborhood, then $m_i(a_i, p) < \alpha < m_i(b_i, p)$. Thus, for all $p \in \cap_{i=1}^n U_i$, (3) is satisfied. □

Lemma B.2. *There exists a distribution $\bar{p} \in P$ satisfying (3).*

Proof. Consider a distribution $\bar{p}$ for which $\bar{p}(g \in \{e_i\}_{i=1}^n) + \bar{p}(g = \mathbf{0}) = 1$, $\bar{p}(g = \mathbf{0}) \in (0, 1)$, and the score distributions are the same for all groups $F_g = F$ for all $g \in G$. The acceptance rates simplify to $R_i(\boldsymbol{\theta}, t(\boldsymbol{\theta})) = 1 - F(r_{e_i}^{-1}(t(\boldsymbol{\theta})|\boldsymbol{\theta}))$. Given a threshold, let $R_0(t) \equiv 1 - F(t(\boldsymbol{\theta}))$ denote the acceptance rate of group $\mathbf{0}$. The total acceptance rate likewise simplifies to

$$S(\boldsymbol{\theta}, t(\boldsymbol{\theta})) = \sum_{i=1}^n R_i(\boldsymbol{\theta}, t(\boldsymbol{\theta}))p(e_i) + R_0(t(\boldsymbol{\theta}))p(\mathbf{0}).$$

An increase in the adjusted scores $\boldsymbol{\theta} \leq \boldsymbol{\theta}'$ implies an increase in the score threshold $t(\boldsymbol{\theta}) \leq t(\boldsymbol{\theta}')$. Because the score distributions are the same for all groups, for any $\mathbf{a} = (a, \ldots, a)$ and $\mathbf{b} = (b, \ldots, b)$ with $a < x_0 < b$ we have $R_i(\mathbf{a}, t(\mathbf{a})) < R_0(\mathbf{a}, t(\mathbf{a})) < R_i(\mathbf{b}, t(\mathbf{b}))$ and thus $R_i(\mathbf{a}, t(\mathbf{a})) < \alpha < R_i(\mathbf{b}, t(\mathbf{b}))$ for all $i = 1, \ldots, n$. Finally, because $\boldsymbol{\theta}_{-i}$ only enters R_i through the threshold t, it follows that $R_i(\boldsymbol{\theta}, t(\boldsymbol{\theta}))$ is decreasing in $\boldsymbol{\theta}_{-i}$ for all $i = 1, \ldots, n$ and $\boldsymbol{\theta} \in X^n$. □

C 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*** if $q_g \circ q_{g'} = q_{g \vee g'} \circ q_{g \wedge g'}$.[10] First, we show how applying a normalization produces the simpler definition used in the text.

Proposition C.1 (Normalize)**.** *Let $g^* \in G$ be a group. If $\mathbf{q}$ 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 C.2 (Relabel)**.** *Suppose the policy $\mathbf{q}$ is nonintersectional*. Let g^* be a group and relabel all groups according to the mapping $h(g) = |g - g^*|$. Then the policy $\mathbf{r}$ 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

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

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

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

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

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 $\mathbf{r}$ is nonintersectional*. $\square$