Affirmative Action with Multidimensional Identities

Jean-Paul Carvalho* Bary Pradelski† Cole Williams‡

Studying the design of affirmative action policies when identities are multidimensional, we provide a formal demonstration of the importance of *intersectionality*. Prevailing affirmative action policies are based only on one identity dimension (e.g., race, gender, socioeconomic class). We find that any such nonintersectional policy can almost never achieve a representative outcome. In fact, nonintersectional policies often increase the underrepresentation of underrepresented groups in a manner undetected by standard measures. Examples based on race and gender reveal significant hidden inequality arising from nonintersectional policies. We show how to construct intersectional policies that achieve proportional representation.

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

1. Introduction

Affirmative action policies are widely employed in college admissions, hiring, lending, government contracting, and selection of electoral candidates by political parties. This has been the subject of considerable controversy. For example, the Supreme Court of the United States has recently ruled that race-based college admission policies are unlawful, even for affirmative action, thus reversing the landmark *Grutter vs. Bollinger* decision from 2003 (Supreme Court of the United States 2023). Our concern in this paper is not whether affirmative action should or should not be practiced. Instead, we are concerned with the proper design of affirmative action policies in practice, whether current policies can meet stated objectives, and how they might be adapted to do so. Affirmative action policies today aim to reduce underrepresentation along a number of identity dimensions (e.g., race, gender, caste, socioeconomic class). Prevailing policies, however, are formulated independently for each identity dimension. Studying a decision maker who must select a subset of applicants from an applicant pool, we demonstrate the impossibility of achieving proportional representation using nonintersectional policies (i.e., dimension-by-dimension adjustments). We also show how to construct intersectional policies that achieve proportional representation. Thus, we provide a formal

* Department of Economics, University of Oxford

† Maison Française, CNRS and Department of Economics, University of Oxford

‡ Department of Economics, Durham University Business School

characterization of the importance of *intersectionality* in the design of affirmative action policies and contribute to an active and important debate in academia and policy circles.

Existing approaches to inequality go beyond aggregate measures of income and wealth inequality, accounting for structural inequality based on various identity characteristics. 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 (Bowen and Bok 1998, 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, Bordalo et al. 2016, Coffman et al. 2021).

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). We show, however, that affirmative action policies that are formulated separately for each identity dimension (e.g., race and gender) suffer from a design flaw. Such nonintersectional policies are standard. They can be found, for example, where monitoring and regulation of representation along each identity dimension is the responsibility of different committees or organizations. An extreme case is where only one dimension is considered. 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). This is at odds with the structural inequality approach which disaggregates income and wealth inequality to examine differences between sociodemographic groups. When identity is multidimensional, the basic unit of analysis is the intersectional group. Hence, the structural inequality approach would suggest going further than the identity dimensions and examining representation among the intersectional groups.

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). Moreover, the failure of nonintersectional policies to achieve a representative outcome can be significant. In some cases, negative spillovers across identity dimensions provide that any nonintersectional policy that changes the representation of all intersectional groups *must* reduce the representativeness of some 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. In simple theoretical examples based on race and gender, we show how nonintersectional policies can exacerbate and hide underrepresentation at the intersectional level. Finally, we show how score distributions can be made endogenous by extending an important paper on the design of affirmative action policies by Fryer and Loury (2013) to multidimensional identities (Section 3.5).

There are a number of practical implications of these results for organizations and policymakers. Organizations meet periodically to evaluate diversity outcomes. Within firms, hiring committees and diversity, equity and inclusion (DEI) officers evaluate hiring and promotion decisions. University admissions committees evaluate the composition of incoming classes of undergraduate and graduate students and track the educational outcomes (grades and dropout rates) of various groups. Though we are unaware of systematic evidence on this issue, it appears that these evaluations are typically made based on tables reporting outcomes for each identity dimension (e.g., race, gender). The data for each intersectional group is typically not compiled, let alone reported, and thus not used in the evaluation process. According to our results, affirmative action policies can worsen the underrepresentation of some underrepresented groups. In addition, the negative consequences of such policies may go undetected by organizations, given the way they typically compile and report data. Thus, if organizations want to reduce the underrepresentation of certain groups, our analysis suggests that data on representation within the organization should be compiled for each intersectional group.

Another practical implication of our results is for reservation policies. For example, India reserves seats for certain districts in state legislatures for disadvantaged (scheduled) castes and tribes; legislative seats in other districts are reserved for women. What our analysis shows is that unless the reservations are defined based on the intersectional groups (not in a undimensional way), proportional representation of groups defined by gender *and* caste cannot generically be achieved. For example, men from disadvantaged castes could be made worse off by the reservation system. In fact, Celis et al. (2014) find that separate quotas for women and ethnic minorities in the selection of candidates for political office in Belgium and the Netherlands led to ethnic minority women being represented in larger numbers than ethnic minority men. The authors attribute this to the strategic choices of party leaders aiming to satisfy both quotas independently.

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, Chen and Li 2009), 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. 2017, Kleinberg and Raghavan 2018, Kleinberg et al. 2018, Chouldechova and Roth 2018, Rambachan and Roth 2020, Raji et al. 2020). In particular, Kleinberg et al. (2017) provide an impossibility result 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 unidimensional 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 often the case in 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, Sgroi et al. (2021) and Hong et al. (2022) on ingroup bias and redistributive preferences with multidimensional identities, and Shayo (2009) and Akerlof (2017) on how individuals choose to value different dimensions of their identity. Carvalho and Pradelski (2022) introduce the concept of intersectionality to economic theory. They study a specific inequality-generating mechanism, where individuals care about the representation of their group. In this context, they 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. Recent work by Aygun and Bó (2021), Pathak et al. (2021), and Sönmez and Yenmez (2022) identifies the specific challenges of combining distinct reservation systems when there are multidimensional identities. Because the affirmative action policies we study are far more flexible than reservation systems, our results highlight the design challenges that come from multidimensional identity per se. 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 tradeoff 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 than prior analysis suggests for specific inequality-generating mechanisms (cf. Carvalho and Pradelski 2022, Mehrotra et al. 2022). 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. Going beyond the specific examples in the existing literature is not a simple matter of generalization. It requires significant technical and conceptual advances including a general definition of the space of intersectional and non-intersectional policies and appropriate topological arguments. Moreover, through the general framework we introduce we are able to derive a number of results not seen in the literature to date. These results have important implications for the design of real-world affirmative action policies, including policies based on race and gender in the United States. Not only do we demonstrate the generic impossibility of non-intersectional policies achieving a representative outcome, we also show how badly non-intersectional policies fail, and the (non-generic) conditions under which they perform well. For example, the setting of Mehrotra et al. (2022) falls into one of our non-generic classes, so that a (different) non-intersectional policy can be designed to achieve a representative state. While the examples we study throughout the paper are based on race and gender, the analysis is far more general, applying to other identity dimensions such as caste, socioeconomic class, and geographic origin.

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 interval $X \subset \mathbb{R}$ and a multidimensional identity described by a vector of group characteristics $\mathbf{g} \in G = \{0,1\}^n$, with $n \geq 2$.[1] While an individual's full identity is an n-dimensional object $\mathbf{g} = (g_1, ..., g_n)$, 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 $F_{\mathbf{g}}(\cdot) \equiv p(\cdot|\mathbf{g})$ are continuous and have full support on X. We denote the marginal probability of belonging to group $\mathbf{g}$, $p(X \times \{\mathbf{g}\})$, simply by $p(\mathbf{g})$.[2]

With the goal of achieving a more representative accepted class (e.g., student body, employee pool), the decision maker sets a *policy* $\mathbf{q} = (q_{\mathbf{g}})_{\mathbf{g} \in G}$ such that $q_{\mathbf{g}} : X \to X$ maps the score x of an applicant from group $\mathbf{g}$ to an adjusted score $q_{\mathbf{g}}(x)$. We impose structure on the space of policies

[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 (for definitions see Chapters 2 and 15 of Aliprantis and Border 2006).

by supposing that there is a family of increasing bijections $Q \subset X^X$ such that $\mathbf{q} \in Q^{|G|}$. First, we require the set of functions to be *rich*, that is, policies are powerful in the sense that they can transform any score to any other score. Formally, for all $x, y \in X$ there exists a function $q \in Q$ satisfying $q(x) = y$. Second, we require that the set of functions is *commutative*, formally, $q, q' \in Q$ implies $q \circ q' = q' \circ q$. As we shall see this assumption ensures that each individual is assigned a unique score, even when policies are applied independently across identity dimensions.

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 $q_{\mathbf{g}}(x) \geq x^*$ are accepted and all others $q_{\mathbf{g}}(x) < x^*$ are rejected, subject to the capacity constraint $\sum_{\mathbf{g} \in G} \Pr(q_{\mathbf{g}}(x) \geq x^* | \mathbf{g}) p(\mathbf{g}) = \alpha$.[3]

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

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

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

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

[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 $q_{\mathbf{g}}(x)$. Equivalently, a different monotone acceptance rule could be applied to *unadjusted* scores for each group $\mathbf{g}$.

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.

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

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

3. Results

Our analysis answers the following questions: Can proportional representation of identity groups be achieved with nonintersectional policies (as is current practice)?[4] If not generally, under what conditions? Can nonintersectional measures of inequality disguise or even worsen underrepresentation of some intersectional groups?

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.[5]

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 1 and 2 in Appendix B. The rest of the proof, including the construction of intersectional policies, is presented here.

[4] While prior work focuses on the tradeoff between selecting applicants with the highest scores and achieving proportional representation (e.g. Chan and Eyster 2003), we ask the more primitive question of whether proportional representation is even achievable using standard nonintersectional policies.

[5] 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).

Proof. We begin with part (a). The ensuing outcome is representative if

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

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

$$q_{\mathbf{g}}(\beta_{\mathbf{g}}) = x^* \text{ for all } \mathbf{g} \in G. \tag{2}$$

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 $q_{\mathbf{e}_i}(\beta_{\mathbf{e}_i}) = \beta_{\mathbf{0}}$ for all $1 \leq i \leq n$. When policies are rich and commutative, this uniquely determines each $q_{\mathbf{e}_i}$ (see Lemma 1). Condition (2) also requires $q_{\mathbf{1}}(\beta_{\mathbf{1}}) = (q_{\mathbf{e}_1} \circ \cdots \circ q_{\mathbf{e}_n})(\beta_{\mathbf{1}}) = \beta_{\mathbf{0}}$ or equivalently $\beta_{\mathbf{1}} = (q_{\mathbf{e}_n}^{-1} \circ \cdots \circ q_{\mathbf{e}_1}^{-1})(\beta_{\mathbf{0}})$. For $n \geq 2$, the set of distributions for which this equality fails to hold and $p(\mathbf{g}) > 0$ for all $\mathbf{g} \in G$ is open and dense in P (see Lemma 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 $\Pr(q_{\mathbf{g}}(x) \geq x^*|\mathbf{g}) = \alpha$ for all $\mathbf{g} \in G$ and thus $\sum_{\mathbf{g} \in G} \Pr(q_{\mathbf{g}}(x) \geq x^*|\mathbf{g}) p(\mathbf{g}) = \alpha$. From the conclusion of part (a), such a policy must be intersectional on an open dense subset of P. $\square$

This demonstrates that generically proportional representation 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 $a > 0$ 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.

However, 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. This can be done simply by choosing the $\alpha p(g)$

applicants with the highest scores from each intersectional group. That the solution is so simple weighs in favor of intersectional policies, since it means that they are not necessarily more complex than nonintersectional policies. Note, however, that the number of intersectional groups grows exponentially in the number of group characteristics. For example, with binary characteristics the number of intersectional groups is 2^n. Therefore, computational complexity of even this simple operation can explode as the number of identity dimensions grows.

In fact, using the arguments underpinning Theorem 1, the results readily generalize in various ways. Firstly, while Theorem 1 considers proportional representation to be the target, it readily extends to any feasible target.

Definition 3. *Given a feasible target representation, that is,* $\boldsymbol{w} = (w_{\mathbf{g}})_{\mathbf{g}\in G} \in (0, \frac{1}{\alpha}]$ *such that* $\sum_{\mathbf{g}\in G} w_{\mathbf{g}} p(\mathbf{g}) = 1$*, an outcome is* $\boldsymbol{w}$*-representative if the representation of each intersectional group is equal to its* $\boldsymbol{w}$*-weighted population share:* $\hat{p}(\mathbf{g}) = w_{\mathbf{g}} p(\mathbf{g})$ *for all* $\mathbf{g} \in G$*.*

From this definition, if for each group we substitute $\tilde{\beta}_{\mathbf{g}} \equiv F_{\mathbf{g}}^{-1}(1 - w_{\mathbf{g}}\alpha)$ in place of $\beta_{\mathbf{g}}$ within the proof of Theorem 1, we obtain the following corollary.

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

(a) There does not exist a nonintersectional policy that yields a $\boldsymbol{w}$*-representative outcome.*

(b) There exists an intersectional policy that yields a $\boldsymbol{w}$*-representative outcome.*

While our results so far illustrate the limitations of nonintersectional policies, the result we have proved is actually even stronger. Even if a policy is almost intersectional, it cannot achieve proportional representation.

Definition 4. *A policy is* ***imperfectly intersectional*** *if there are distinct groups* $\mathbf{g}, \mathbf{g}', \mathbf{g}'' \in G$ *satisfying* $\mathbf{g} = \mathbf{g}' + \mathbf{g}''$ *and for which* $q_{\mathbf{g}} = q_{\mathbf{g}'} \circ q_{\mathbf{g}''}$*.*

Corollary 2. *For generic distributions* $p \in P$*, there does not exist an imperfectly intersectional policy that yields a representative outcome.*

That is, even if the policymaker can fashion the affirmative action policy to treat nearly all intersectional groups separately, if there is even one group whose treatment is nonintersectional, then the outcome is typically not representative. While the proof of Theorem 1 implies the desired conclusion for $\mathbf{g}=\mathbf{1}$, the conclusion equally holds for all other groups $\mathbf{g}$. Adapting the argument from Theorem 1 provides that, for distinct groups $\mathbf{g}'\leq\mathbf{g}$, the set of distributions $A(\mathbf{g},\mathbf{g}')\subset P$ for which $q_{\mathbf{g}}(\beta_{\mathbf{g}})=q_{\mathbf{g}-\mathbf{g}'}(\beta_{\mathbf{g}-\mathbf{g}'})=x^*$ implies $q_{\mathbf{g}'}\circ q_{\mathbf{g}-\mathbf{g}'}(\beta_{\mathbf{g}})\neq x^*$ is open and dense in P, implying that $\bigcup_{\mathbf{g}\in G}\bigcup_{\mathbf{g}'\leq\mathbf{g}}A(\mathbf{g},\mathbf{g}')$ is open and dense in P.

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 $\boldsymbol{\rho}=(\rho_{\mathbf{g}})_{\mathbf{g}\in G}\equiv(|\hat{p}(\mathbf{g})-p(\mathbf{g})|)_{\mathbf{g}\in G}$, so that $\rho\in[0,1]^n$. A representative outcome is thus given by $\boldsymbol{\rho}=\mathbf{0}$. Letting $\|\cdot\|$ denote the Euclidean norm, the following result proceeds from Theorem 1:

Corollary 3. *Generically, for each distribution p, there exists a number $c>0$ such that $\|\boldsymbol{\rho}\|>c$ whenever the policy is nonintersectional.*

The proof is relegated to Appendix B. 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 $\beta_{\mathbf{g}}\equiv F_{\mathbf{g}}^{-1}(1-\alpha)$ is the $(1-\alpha)$th score quantile.

Definition 5. *The environment exhibits independence across identity dimensions if the function $q\in Q$ mapping $q(\beta_{\mathbf{g}})=\beta_{\mathbf{g}'}$ also maps $q(\beta_{\mathbf{g}-\mathbf{g}'})=\beta_{\mathbf{0}}$ for all groups $\mathbf{g}'\leq\mathbf{g}$.*[6]

To interpret this condition, consider the case of gender and race as coded in Example 1. The condition in Definition 5 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 $q(\beta_{(0,1)})=\beta_{(0,0)}$.

[6]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 $i=1,...,n$.

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

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

The following theoretical 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 $\mathbf{g}$ from a normal distribution with mean $\mu_{\mathbf{g}}$ and a variance that is common to all groups. Thus, the biased scores for members of group $\mathbf{g}$ are normally distributed with mean $\hat{\mu} - \mu_{\mathbf{g}}$ 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 $\beta_{\mathbf{g}}$ satisfying $\Phi\left(\frac{\beta_{\mathbf{g}} - \hat{\mu} + \mu_{\mathbf{g}}}{\sigma}\right) = 1 - \alpha$. Simplifying this equation, we obtain $\beta_{\mathbf{g}} = \Phi^{-1}(1-\alpha)\,\sigma + \hat{\mu} - \mu_{\mathbf{g}}$.*

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 $q_{\mathbf{g}}(\beta_{\mathbf{g}}) = \beta_{\mathbf{g}} + \delta_{\mathbf{g}}$ is constant across groups, which is equivalent to requiring $\delta_{\mathbf{g}} - \mu_{\mathbf{g}} = \delta_{\mathbf{g}'} - \mu_{\mathbf{g}'}$ for all $\mathbf{g}, \mathbf{g}' \in G$. Moreover, if the policy is nonintersectional, then for any $\mathbf{g}' \leq \mathbf{g}$ we have $\delta_{\mathbf{g}} = \delta_{\mathbf{g}'} + \delta_{\mathbf{g}-\mathbf{g}'}$. Taken together, these conditions provide the following equalities

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

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

Corollary 4. *For Example 2, a nonintersectional policy achieves a representative outcome if and only if* $\mathbf{g} = \mathbf{g}' + \mathbf{g}''$ *implies* $\mu_{\mathbf{g}} + \mu_{\mathbf{0}} = \mu_{\mathbf{g}'} + \mu_{\mathbf{g}''}$ *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 $\hat{p}_0(\mathbf{g}) = \frac{(1-F_{\mathbf{g}}(x_0^*))p(\mathbf{g})}{\sum_{\mathbf{g}' \in G}(1-F_{\mathbf{g}'}(x_0^*))p(\mathbf{g}')}$ be the representation of group $\mathbf{g}$ in the absence of any policy. A group is *overrepresented* without a policy if $\hat{p}_0(\mathbf{g}) > p(\mathbf{g})$ and *underrepresented* if $\hat{p}_0(\mathbf{g}) < p(\mathbf{g})$. We can then define an improvement in representativeness as follows:

Definition 6. *A* monotone improvement *occurs if* $|\hat{p}(\mathbf{g}) - p(\mathbf{g})| < |\hat{p}_0(\mathbf{g}) - p(\mathbf{g})|$ *for all* $\mathbf{g} \in G$.

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* $\mathbf{g}$ *are overrepresented and there is a group* $\mathbf{g}' \leq \mathbf{g}$ *such that* $\mathbf{g}'$ *and* $\mathbf{g} - \mathbf{g}'$ *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 $\mathbf{g}$ are overrepresented, and groups $\mathbf{g}'$ and $\mathbf{g}'' = \mathbf{g} - \mathbf{g}'$ 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 $\sum_{\mathbf{g} \in G} p(\mathbf{g})(1 - F_{\mathbf{g}}(x_0^*)) = \alpha$. After introducing the policy, the admissions cutoff shifts to x^* equating $\sum_{\mathbf{g} \in G} p(\mathbf{g})(1 - F_{\mathbf{g}}(q_{\mathbf{g}}^{-1}(x^*))) = \alpha$. A monotone improvement requires $q_{\tilde{\mathbf{g}}}(x_0^*) < x^*$ for any overrepresented group and $q_{\tilde{\mathbf{g}}}(x_0^*) > x^*$ 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

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

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

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

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

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 7. *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.

Proposition 3. *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 D. While Proposition 3 may lend some support to nonintersectional policies, it raises the problem of hidden inequality, that is, inequality between intersectional groups that goes unnoticed. 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* 71 *percent of white men. Now consider a nonintersectional policy that lifts the scores of women by* 0.925 *and the scores of black individuals by* 0.375*, so that a reductive representative outcome is achieved. While the policy brings the acceptance rate for white women and white men roughly in line with the capacity, i.e,* 50 *percent, the acceptance rate for black men falls from* 42 *to* 37 *percent, counterbalanced by an increase for black women from* 32 *to* 63 *percent. See Figure 1.*

While the reductive representative outcome has the desirable effect of increasing the representation for women and black individuals, only attending to these two dimensions (for examples, in reviews of admissions and hiring practices that only report representation by race and gender) 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

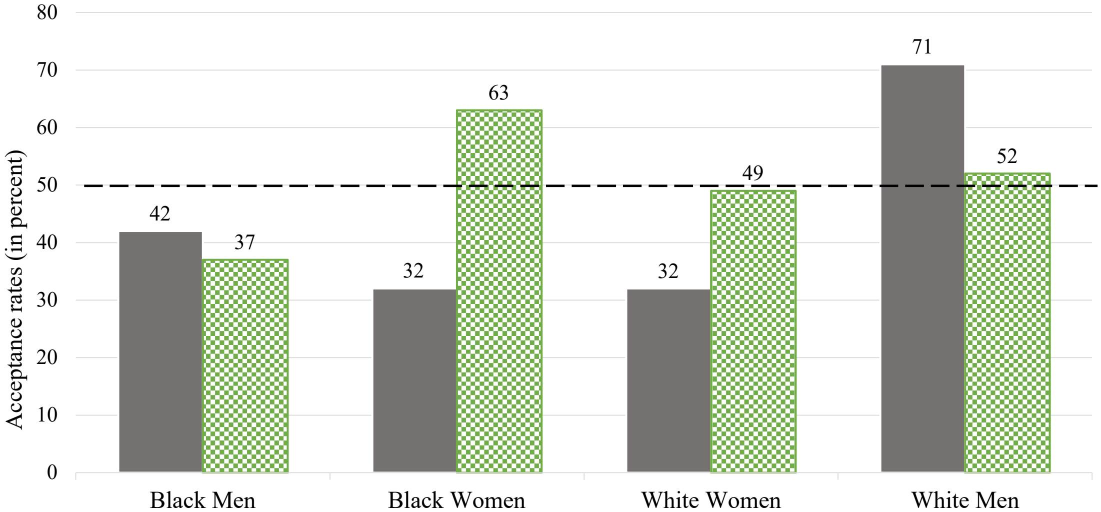

Figure 1 **Group acceptance rates without an affirmative action policy (gray, solid) and with a nonintersectional policy achieving a reductive representative outcome (green, chequered). Black dashed line shows the rate achieving a representative outcome.**

out the benefit to black men. Figure 2 further illustrates how the nonintersectional policy achieves proportional representation along the dimension of race (10*—*90*) and gender (*50*—*50*), but reduces the representation of black men from* 4.2 *percent to* 3.7 *percent. This reduction in representation is due to the resulting overrepresentation of black women, with an increase from* 3.2 *to* 6.3 *percent. Of course, this is one example, and we can construct others in which minority women face a double disadvantage.*

Finally, note that our example does not fulfil the conditions of Proposition 2, so these conditions are not necessary for there to be no nonintersectional policy that yields monotone improvements in representation among the intersectional groups. Of course, even if there exist nonintersectional policies that yield reductive representative outcomes and do not increase underrepresentation of any underrepresented intersectional group, we know by Theorem 1 that they cannot achieve a representative outcome.

3.5. Endogenous Score Distributions

We will now show how the score distributions in our model can be made endogenous by adopting the approach of Fryer and Loury (2013), who study the design of affirmative action policies in a two-stage environment described below. Their central finding is that restricting affirmative action policies to be race-blind is inefficient and also fundamentally changes the structure of the optimal policy. We extend their analysis to multidimensional identities and apply our results. In doing so, we alter their notation to clarify the connection to our work.

Sometimes making key features of a model endogenous produces non-generic outcomes. Here, non-generic score distributions could be produced and given these score distributions it could be

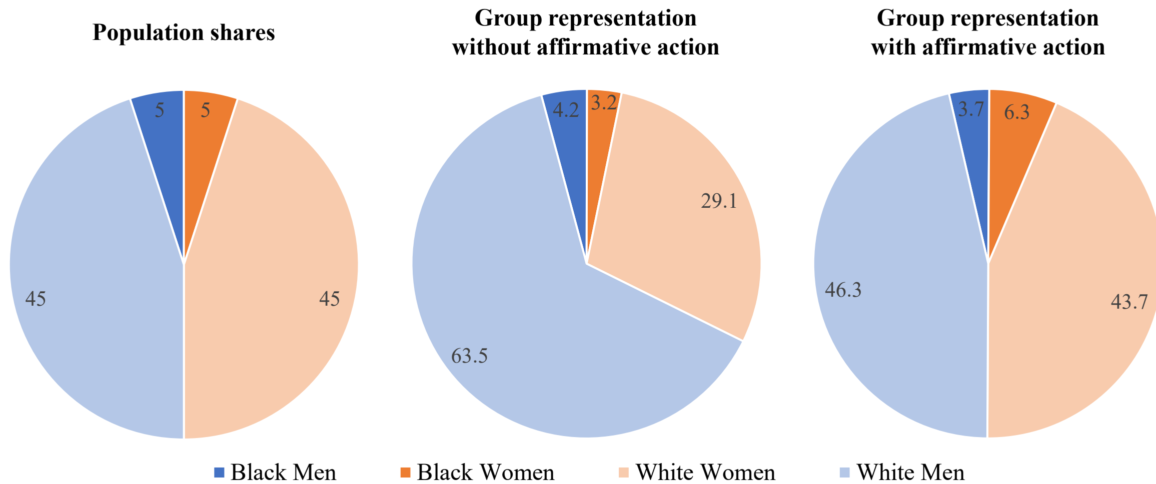

Figure 2 **A nonintersectional affirmative action policy that achieves proportional representation along the dimensions of race and gender, but reduces the representation of black men.**

that nonintersectional policies can achieve proportional representation. We find that not to be the case. Using our results, we show that when score distributions are endogenous and even when affirmative action policies are permitted to be "sighted" and conditioned on race, nonintersectional policies will fail to achieve diversity objectives.

In stage 1 of the Fryer-Loury model, 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 distribution which depends on their social identity $\mathbf{g} \in G = \{A, B\}$. Let $H_{\mathbf{g}}(c)$ and $h_{\mathbf{g}}(c)$ be the cost distribution and density for members of identity group $\mathbf{g} \in G$ 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 $H_{\mathbf{g}}$ 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.[7]

[7]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.

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, $\mathbf{g} \in G = \{0,1\}^n$, and apply our results. Following the original formulation, let the affirmative action policy be additive. If a member of group $\mathbf{g}$ buys a slot, they receive a payment equal to their productivity, minus the cost x^*, net of the subsidy/tax $\delta_{\mathbf{g}}$, i.e., $x + \delta_{\mathbf{g}} - x^*$. Normalizing the outside option to zero, such an individual purchases a slot if and only if $x + \delta_{\mathbf{g}} - x^* \geq 0$.

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

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

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

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

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

$$\begin{aligned} R_{\mathbf{g}}(t_{\mathbf{g}}) &= H_{\mathbf{g}}(B(t_{\mathbf{g}}))(1 - F_1(t_{\mathbf{g}})) + (1 - H_{\mathbf{g}}(B(t_{\mathbf{g}})))(1 - F_0(t_{\mathbf{g}})) \\ &= H_{\mathbf{g}}(B(t_{\mathbf{g}}))(F_0(t_{\mathbf{g}}) - F_1(t_{\mathbf{g}})) + 1 - F_0(t_{\mathbf{g}}). \end{aligned}$$

Computing the change in acceptance as a result of changing $t_{\mathbf{g}}$, we obtain

$$\begin{aligned} \frac{\partial R_{\mathbf{g}}}{\partial t_{\mathbf{g}}} &= -h_{\mathbf{g}}(B(t_{\mathbf{g}}))(F_0(t_{\mathbf{g}}) - F_1(t_{\mathbf{g}}))^2 + H_{\mathbf{g}}(B(t_{\mathbf{g}}))(f_0(t_{\mathbf{g}}) - f_1(t_{\mathbf{g}})) - f_0(t_{\mathbf{g}}) \\ &= -h_{\mathbf{g}}(B(t_{\mathbf{g}}))(F_0(t_{\mathbf{g}}) - F_1(t_{\mathbf{g}}))^2 - (1 - H_{\mathbf{g}}(B(t_{\mathbf{g}})))f_0(t_{\mathbf{g}}) - H_{\mathbf{g}}(B(t_{\mathbf{g}}))f_1(t_{\mathbf{g}}) < 0. \end{aligned}$$

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

Now consider nonintersectional policies which require $\delta_{\mathbf{g}}=\delta_{\mathbf{g}'}+\delta_{\mathbf{g}''}$ for all groups $\mathbf{g}$, $\mathbf{g}'$, and $\mathbf{g}''$ satisfying $\mathbf{g}=\mathbf{g}'+\mathbf{g}''$. 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 $\mathbf{g}=\mathbf{g}'+\mathbf{g}''$ implies

$$t^*_{\mathbf{g}}=x^*-\delta_{\mathbf{g}}=x^*-\delta_{\mathbf{g}'}-\delta_{\mathbf{g}''}=t^*_{\mathbf{g}'}+t^*_{\mathbf{g}''}-t^*_{\mathbf{0}}.$$

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

Formally, to connect with Proposition 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 $p(\cdot|\mathbf{g})\sim H_{\mathbf{g}}$ is continuous. Then the conclusion of Proposition 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: $\delta_{\mathbf{g}}=R^{-1}_{\mathbf{0}}(\alpha)-R^{-1}_{\mathbf{g}}(\alpha)$ for all $\mathbf{g}\in G$. 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 is largely focused on the case of unidimensional identities. 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.

In addition to the problems with non-intersectional policies set out in this paper, there are other fundamental challenges in designing affirmative action policies when identities are multidimensional. These include whether affirmative action policies strengthen existing identities and forms of division, how to define the set of relevant identities, and how to account for the manipulability of identity, either through some form of "passing" or misrepresentation.[8] For example, if the precise identity dimensions we care about and the values along each dimension are determined by a political process, it is not clear that this process will ever settle on a reasonable number of identities. It could be that there is an endless demand for introducing new identity dimensions, so that the design problem blows up. These issues are covered in some detail in Carvalho and Pradelski (2022, Section 5). One potentially fruitful way to proceed both theoretically and empirically is to examine how to design affirmative action policies that are robust to identity manipulation.

Appendix A: 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 $q_{\mathbf{g}} \circ q_{\mathbf{g}'} = q_{g \vee \mathbf{g}'} \circ q_{\mathbf{g} \wedge \mathbf{g}'}$.[9] First, we show how applying a normalization produces the simpler definition used in the text.

Proposition 4 (Normalize)**.** *Let $\mathbf{g}^* \in G$ be a group. If $\mathbf{q}$ is a nonintersectional* policy, then so is the policy r defined by $r_{\mathbf{g}} \equiv q_{\mathbf{g}^*}^{-1} \circ q_{\mathbf{g}}$ for all $\mathbf{g} \in G$.*

Proof. We have:

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

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

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

Proposition 5 (Relabel)**.** *Suppose the policy $\mathbf{q}$ is nonintersectional*. Let $\mathbf{g}^*$ be a group and relabel all groups according to the mapping $h(\mathbf{g}) = |\mathbf{g} - \mathbf{g}^*|$. Then the policy $\mathbf{r}$ defined by $r_{h(\mathbf{g})} = q_{\mathbf{g}}$ for all $\mathbf{g} \in G$ is likewise nonintersectional*.*

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

$$q_{\mathbf{g}} = q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ q_{g_1\mathbf{e}_1}) \circ \cdots \circ (q_{\mathbf{0}}^{-1} \circ q_{g_n\mathbf{e}_n}). \tag{3}$$

[8] 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).

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

It is enough to prove that the claim is true for $\mathbf{g}^* = \mathbf{e}_i$ for $1 \le i \le n$. For simplicity of notation but without loss of generality, consider $\mathbf{g}^* = \mathbf{e}_1$. As $h(\mathbf{0}) = e_1$ and $h(\mathbf{e}_1) = \mathbf{0}$, we have $r_{\mathbf{e}_1} = q_{\mathbf{0}}$ and $r_{\mathbf{0}} = q_{\mathbf{e}_1}$. For $i > 1$, $h(\mathbf{e}_1 + \mathbf{e}_i) = \mathbf{e}_i$ and so

$$\begin{aligned} r_{\mathbf{e}_i} = q_{\mathbf{e}_1+\mathbf{e}_i} = q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ q_{\mathbf{e}_1}) \circ (q_{\mathbf{0}}^{-1} \circ q_{\mathbf{e}_i}) &= q_{\mathbf{0}}^{-1} \circ q_{\mathbf{e}_1} \circ q_{\mathbf{e}_i} \\ &= q_{\mathbf{0}}^{-1} \circ r_{\mathbf{0}} \circ q_{\mathbf{e}_i} \end{aligned}$$

and so $q_{\mathbf{0}}^{-1} \circ q_{\mathbf{e}_i} = r_{\mathbf{0}}^{-1} \circ r_{\mathbf{e}_i}$. Denoting $h(\mathbf{g}) = (h_1, \ldots, h_n)$, for $i > 1$, $h_i = g_i$ and $h_1 = 1 - g_1$. Furthermore, $q_{g_1\mathbf{e}_1} = r_{h_1\mathbf{e}_1}$. Combining these observations, (3) can be written

$$\begin{aligned} r_{h(\mathbf{g})} = q_{\mathbf{g}} &= q_{\mathbf{0}} \circ (q_{\mathbf{0}}^{-1} \circ r_{h_1\mathbf{e}_1}) \circ (r_{\mathbf{0}}^{-1} \circ r_{h_2\mathbf{e}_2}) \circ \cdots \circ (r_{\mathbf{0}}^{-1} \circ r_{h_n\mathbf{e}_n}) \\ &= r_{\mathbf{0}} \circ (r_{\mathbf{0}}^{-1} \circ r_{h_1\mathbf{e}_1}) \circ (r_{\mathbf{0}}^{-1} \circ r_{h_2\mathbf{e}_2}) \circ \cdots \circ (r_{\mathbf{0}}^{-1} \circ r_{h_n\mathbf{e}_n}). \end{aligned} \tag{4}$$

Since $h: G \to G$ is a bijection, for all groups $h' \in G$ there is a group $\mathbf{g} \in G$ such that $h' = h(\mathbf{g})$; hence, $r_{h'}$ can be formulated as in (4). Using this formulation, it follows immediately that $\mathbf{r}$ is nonintersectional*. $\square$

Appendix B: Technical details for Theorem 1 and Corollary 3

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 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 2. *Assume Q to be rich and commutative. The subset of p for which $\beta_{\mathbf{1}} \neq (q_{\mathbf{e}_n}^{-1} \circ \cdots \circ q_{\mathbf{e}_1}^{-1})(\beta_{\mathbf{0}})$ (with $q_{\mathbf{e}_i}(\beta_{\mathbf{e}_i}) = \beta_{\mathbf{0}}$ for $1 \leq i \leq n$) and $p(\mathbf{g}) > 0$ for all $\mathbf{g} \in G$ is open and dense in P.*

Proof. First, let us show that the subset $A = \left\{p \in P : \beta_{\mathbf{1}} = (q_{\mathbf{e}_n}^{-1} \circ \cdots \circ q_{\mathbf{e}_1}^{-1})(\beta_{\mathbf{0}})\right\}$ is closed. Toward a contradiction, let $\{p_\gamma\}_{\gamma \in \Gamma}$ be a net in A converging to $p \in A^c$. Let $(\beta_{\mathbf{g}})_{\mathbf{g} \in G}$ be defined with respect to p and $(\beta_{\mathbf{g}}^\gamma)_{\mathbf{g} \in G}$ be defined with respect to p_γ for each $\gamma \in \Gamma$. By the definition of weak* convergence, for any $\epsilon > 0$,

$$F_{\mathbf{g}}^\gamma(\beta_{\mathbf{g}} - \epsilon) \to F_{\mathbf{g}}(\beta_{\mathbf{g}} - \epsilon) < 1 - \alpha, \; F_{\mathbf{g}}^\gamma(\beta_{\mathbf{g}} + \epsilon) \to F_{\mathbf{g}}(\beta_{\mathbf{g}} + \epsilon) > 1 - \alpha.$$

Hence, $\beta_{\mathbf{g}}^\gamma$ converges to $\beta_{\mathbf{g}}$ for all $\mathbf{g} \in G$. Thus, there exists γ_0 such that $\gamma \geq \gamma_0$ implies $\beta_{\mathbf{1}}^\gamma \neq \sum_{i=1}^n \beta_{\mathbf{e}_i}^\gamma - (n-1)\beta_{\mathbf{0}}^\gamma$ contradicting the assumption that $\{p_\gamma\}_{\gamma \in \Gamma}$ is a net in A. Furthermore, $B = \{p \in P : p(\mathbf{g}) = 0 \text{ for some } \mathbf{g} \in G\}$ 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 $F_{\mathbf{g}}^\gamma = F_{\mathbf{g}}$ for all $\gamma \in (0,1)$ and $\mathbf{g} \neq \mathbf{1}$, $p_\gamma(\mathbf{g}) = p(\mathbf{g})\gamma + p'(\mathbf{g})(1-\gamma)$ where $p'(\mathbf{g}) > 0$ for all $\mathbf{g} \in G$, $F_{\mathbf{1}}^\gamma = F_{\mathbf{1}}\gamma + F_{\mathbf{1}}'(1-\gamma)$ where $F_{\mathbf{1}}(\beta_{\mathbf{1}}) \neq F_{\mathbf{1}}'(\beta_{\mathbf{1}})$.[10] 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 3. 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 $\|\boldsymbol{\rho}(\mathbf{q}_m, x_m)\| < \epsilon$ if $m \geq m_\epsilon$. Continue to denote the $1-\alpha$th score quantile by $\beta_{\mathbf{g}} = F_{\mathbf{g}}^{-1}(1-\alpha)$ for each group $\mathbf{g} \in G$. It must be that $x_m \to x^* \equiv \beta_{\mathbf{0}}$ or else $\rho_{\mathbf{0}}$ is bounded away from zero. Similarly, denoting $\mathbf{q}_m = (q_g^m)_{\mathbf{g} \in G}$, it must also be that $q_{\mathbf{e}_i}^m(\beta_{\mathbf{e}_i}) \to x^*$ for each $i = 1, ..., n$ or else some $\rho_{\mathbf{e}_i}$ is bounded away from zero. But then, letting $\mathbf{q}$ be the unique nonintersectional policy satisfying $q_{\mathbf{e}_i}(\beta_{\mathbf{e}_i}) = x^*$ for all $i = 1, ..., n$, 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. $\square$

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

Appendix C: Proof of Proposition 3

To prove Proposition 3, 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. $\square$

Appendix D: Proof of Proposition 3

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

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

The total acceptance rate is

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

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

$$R_i(a_i, \boldsymbol{\theta}_{-i}, t(a_i, \boldsymbol{\theta}_{-i})) < \alpha < R_i(b_i, \boldsymbol{\theta}_{-i}, t(b_i, \boldsymbol{\theta}_{-i})) \text{ for all } \boldsymbol{\theta}_{-i} \in \times_{j \neq i}[a_j, b_j]. \tag{5}$$

For example, the following are two natural conditions that guarantee (5) 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 (5) 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 $q_{\mathbf{e}_i}(\cdot) = r(\cdot|\theta_i^*)$ for all $i = 1, \ldots, n$ and the threshold $x^* = t(\boldsymbol{\theta}^*)$ achieves a reductive representative outcome.[11]

To complete the proof, we show that for any distribution $\bar{p}$ satisfying (5), each p in a neighborhood of a $\bar{p}$ likewise satisfies (5) and then demonstrate the existence of a $\bar{p}$ satisfying (5).

Lemma 3. *Suppose that for $\bar{p} \in P$ there exist $\mathbf{a}, \mathbf{b} \in X^n$ such that (5) is satisfied. Then (5) 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

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

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 $p \in \cap_{i=1}^n U_i$, (5) is satisfied. $\square$

Lemma 4. *There exists a distribution $\bar{p} \in P$ satisfying (5).*

Proof. Consider a distribution $\bar{p}$ for which $\bar{p}(\mathbf{g} \in \{\mathbf{e}_i\}_{i=1}^n) + \bar{p}(\mathbf{g} = \mathbf{0}) = 1$, $\bar{p}(\mathbf{g} = \mathbf{0}) \in (0, 1)$, and the score distributions are the same for all groups $F_{\mathbf{g}} = F$ for all $\mathbf{g} \in G$. The acceptance rates simplify to $R_i(\boldsymbol{\theta}, t(\boldsymbol{\theta})) = 1 - F(r_{\mathbf{e}_i}^{-1}(t(\boldsymbol{\theta})|\boldsymbol{\theta}))$. 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

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

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

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 $\boldsymbol{\theta} \in X^n$. $\square$

Acknowledgments

to fill

References

Akerlof GA, Kranton RE (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 CD, Border KC (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide* (Springer Science & Business Media).

Altonji JG, Blank RM (1999) Race and gender in the labor market. *Handbook of Labor Economics* 3:3143–3259.

Arrow KJ (1973) The theory of discrimination. Ashenfelter O, Rees A, eds., *Discrimination in Labor Markets*, 3–33 (Princeton, NJ: Princeton University Press).

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

Aygun O, Bó I (2021) College admission with multidimensional privileges: The brazilian affirmative action case. *American Economic Journal: Microeconomics* 13(3):1–28.

Bagde S, Epple D, Taylor L (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 GS, Tomes N (1979) An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* 87(6):1153–1189.

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

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

Bordalo P, Coffman K, Gennaioli N, Shleifer A (2016) Stereotypes. *The Quarterly Journal of Economics* 131(4):1753–1794.

Borjas GJ (1992) Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics* 107(1):123–150.

Bowen WG, Bok D (1998) *The Shape of the River* (Princeton University Press).

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

Carvalho JP, Pradelski B (2022) Identity and underrepresentation: Interactions between race and gender. *Journal of Public Economics* (216):104764.

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

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

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

Chan J, Eyster E (2003) Does banning affirmative action lower college student quality? *American Economic Review* 93(3):858–872.

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

Chen Y, Li SX (2009) Group identity and social preferences. *American Economic Review* 99(1):431–57.

Chetty R, Hendren N, Kline P, Saez E (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, Roth A (2018) The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810* .

Chung KS (2000) Role models and arguments for affirmative action. *American Economic Review* 90(3):640–648.

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

Coffman KB, Exley CL, Niederle M (2021) The role of beliefs in driving gender discrimination. *Management Science* 67(6):3551–3569.

Collins PH, Bilge S (2020) *Intersectionality* (John Wiley & Sons).

Cooper B (2016) Intersectionality. Disch L, Hawkesworth M, eds., *The Oxford Handbook of Feminist Theory*, 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, Gneezy U (2009) Gender differences in preferences. *Journal of Economic Literature* 47(2):448–74.

Eguia JX (2017) Discrimination and assimilation at school. *Journal of Public Economics* 156:48–58.

Fang H, Moro A (2011) Theories of statistical discrimination and affirmative action: A survey. Jess Benhabib MOJ, Bisin A, eds., *Handbook of Social Economics*, 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, Pavan A (2021) "Soft" affirmative action and minority recruitment. *American Economic Review: Insights* 3(1):1–18.

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

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

Fryer RG, Loury GC (2005) Affirmative action and its mythology. *Journal of Economic Perspectives* 19(3):147–162.

Fryer RG, Loury GC (2013) Valuing diversity. *Journal of Political Economy* 121(4):747–774.

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

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

Hong F, Riyanto YE, Zhang R (2022) Multidimensional social identity and redistributive preferences: an experimental study. *Theory and Decision* 1–34.

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

Kleinberg J, Mullainathan S, Raghavan M (2017) Inherent trade-offs in the fair determination of risk scores. Papadimitriou CH, ed., *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 43:1–43:23.

Kleinberg J, Raghavan M (2018) Selection Problems in the Presence of Implicit Bias. *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS '18)*.

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

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

Loury GC (1981) Intergenerational transfers and the distribution of earnings. *Econometrica* 49(4):843–867.

Loury GC (2009) *The Anatomy of Racial Inequality* (Harvard University Press).

Mehrotra A, Pradelski BSR, Vishnoi NK (2022) Selection in the presence of implicit bias: The advantage of intersectional constraints. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 599–609, FAccT '22.

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

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

Pathak PA, Sönmez T, Ünver MU, Yenmez MB (2021) Fair allocation of vaccines, ventilators and antiviral treatments: leaving no ethical value behind in health care rationing. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 785–786.

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

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

Rambachan A, Roth J (2020) Bias In, Bias Out? Evaluating the Folk Wisdom. Roth A, ed., *1st Symposium on Foundations of Responsible Computing (FORC 2020)*, volume 156 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 6:1–6:15.

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

Saari DG (2018) *Mathematics motivated by the social and behavioral sciences* (SIAM).

Schroer M (2019) Sociology of attention: fundamental reflections on a theoretical program. Brekus WH, Ignatow G, 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).

Sgroi D, Yeo J, Zhuo S (2021) Ingroup bias with multiple identities: The case of religion and attitudes towards government size. *IZA Discussion Paper* .

Shayo M (2009) A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review* 103(2):147–174.

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

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

Sönmez T, Yenmez MB (2022) Affirmative action in India via vertical, horizontal, and overlapping reservations. *Econometrica* 90(3):1143–1176.

Supreme Court of the United States (2023) Students for Fair Admissions, Inc. v. President and fellows of Harvard College (20–1199).

Young HP (2015) The evolution of social norms. *Annual Review of Economics* 7(1):359–387.