Diversity Preferences and Affirmative Action

Oğuzhan Çelebi*

October 18, 2024

Abstract

In various contexts, institutions allocate resources using rules that determine selections given the set of candidates. Many of these rules incorporate affirmative action, accounting for both identity and (match) quality of individuals. This paper studies the relationship between these rules and the preferences underlying them. I map the standard setting of market design to the revealed preference framework, interpreting choice rules as observed choices made across different situations. I provide a condition that characterizes when a rule can be rationalized by preferences based on identities and qualities. I apply tests based on this condition to evaluate real-world mechanisms, including India's main affirmative action policy for allocating government jobs, and find that it cannot be rationalized. When identities are multidimensional, I show that non-intersectional views of diversity can be exploited by dominant groups to increase their representation and cause the choice rules to violate the substitutes condition, a key requirement for the use of stable matching mechanisms. I also characterize rules that can be rationalized by preferences separable in diversity and quality, demonstrating that they lead to a unique selection within the broader set of policies that reserve places based on individuals' identities.

*Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA 94305. Email: ocelebi@stanford.edu

I am grateful to Daron Acemoglu, Nick Arnosti, Yunus Aybas, Roberto Corrao, Federico Echenique, Glenn Ellison, Joel Flynn, Stephen Morris, Michael Ostrovsky, Parag Pathak, Alvin Roth, Tayfun Sönmez, Bumin Yenmez, Alexander Wolitzky and participants in the MIT Theory Lunch, Stanford Market Design Workshop, EC'24, and UC Davis for helpful discussions and comments. This paper was previously circulated under the name "Diversity Preferences, Affirmative Action and Choice Rules."

1. Introduction

Institutions in charge of allocating resources or hiring individuals make their decisions based on multiple criteria, such as the quality of the candidates, the benefits they receive from the allocated resource, and their socioeconomic characteristics. School districts in Chicago (Dur, Kominers, Pathak, and Sönmez, 2018) and Boston (Dur, Pathak, and Sönmez, 2020), and universities in Brazil (Aygun and Bó, 2021) prefer schools to have a diverse student body, medical authorities prefer the allocation of scarce treatments to consider equity and diversity (Pathak, Sönmez, Ünver, and Yenmez, 2021; Akbarpour, Budish, Dworczak, and Kominers, 2021; Grigoryan, 2021) and the Indian government uses protections for historically discriminated groups when allocating government positions (Aygün and Turhan, 2017; Sönmez and Yenmez, 2022). In these settings, individuals are heterogeneous in two domains. The first is their identity, which might include socioeconomic status for students, healthcare worker status for patients, or caste for government position applicants. The second is their score, such as exam scores in student assignment and government job allocation, or indices of clinical need in medical resource allocation. These scores may reflect match quality, allocating a medical resource to a sicker individual or a government job to a higher-achieving candidate could yield greater benefits; or represent individuals' property rights, for example, students with higher scores might deserve places in selective public schools more than others. The affirmative action programs implemented by these institutions, where an individual's selection depends not only on their score but also on their identity and the composition of the identities of other selected individuals, demonstrate a commitment to diversity and equity, alongside a preference for allocating resources to those with the highest scores.

This paper studies the relationship between (i) how institutions evaluate the composition of selected individuals, particularly with respect to their socioeconomic characteristics and diversity considerations; (ii) how they assess the trade-offs between diversity and scores; and (iii) how they establish their choice rules, which determine the set of chosen individuals from each pool of candidates. First, I adapt revealed preference analysis to market design by interpreting choice rules as an agent's choices from various feasible sets. I characterize the class of rules that can be designed by a decision maker with well-defined preferences, and further show when these preferences satisfy certain additional conditions, similar to the standard analysis of rationality of an agent based on their observed choice behavior. Applying these results, I identify shortcomings in various rules used in practice, including the main affirmative action rule used to allocate government jobs in India. Second, I explore the relationship between how diversity is evaluated when identities are multidimensional, and the properties of choice rules. When diversity is evaluated without considering intersectionality–

that is, when different dimensions, like race and gender, are assessed separately–the resulting choices exhibit complementarities among individuals belonging to different groups across dimensions.[1] I demonstrate that this results in the failure of a key property of choice rules, the substitutes condition, which is crucial for the existence of competitive equilibria and the use of stable matching mechanisms. Moreover, I show that non-intersectional views of diversity allows dominant groups to increase their representation without compromising the perception of diversity. Third, I study the class of choice rules rationalized by preferences that treat score and diversity domains in an additively separable way, a structure commonly assumed in applied theoretical work on affirmative action. I demonstrate that this class is defined by three well-known properties and encompasses many of the choice rules used in practice. However, additive separability imposes a unique processing order for quotas, providing evidence that the implementation details of such policies could have significant effects, contributing to the literature on quota policies.

Rationality of Choice Rules. I begin by establishing a connection between the standard model in market design, where a choice rule selects a subset from a set of applicants, and the framework of revealed preference/choice theory, where an agent makes choices from different feasible sets, known as a choice environment. The choice rules used in practice and studied in market design literature induce a specific choice environment, enabling the application of results from choice theory. I characterize the class of choice rules that can be rationalized by preferences based on the identities and scores of individuals in an unrestricted way, and when these preferences satisfy certain additional conditions. Rationality is characterized by an acyclicity condition based on the *congruence axiom* of Richter (1966) and its generalization which incorporates an exogenous preorder, as provided by Nishimura, Ok, and Quah (2016). On the theoretical side, these results complement the earlier literature by focusing on the preferences driving the choice rules instead of their axiomatic properties (*e.g.*, incentive compatibility). This analysis helps us understand the preferences behind the choice rules adopted by institutions and assess whether such preferences exist. Moreover, in many practical settings, including those studied in this paper, the choice rules are known to the analyst and can be directly evaluated, unlike the settings in revealed preference theory, where the evaluation often relies on the observation of potentially limited data from from agents. This means that my results can be used to assess the rationality of a rule without requiring any data, and this assessment can even be performed before the rule is implemented.

On the applied side, these results are valuable in two distinct ways. First, by constructing

[1]For example, an institution focused on racial (Black or white) and gender (male or female) diversity might evaluate gender diversity by the number of men and women, and evaluate racial diversity by the number of white and Black individuals, without accounting for the intersection of these identities.

cycles, I demonstrate that certain rules used in practice cannot be rationalized. Specifically, I evaluate the affirmative action mechanisms in India and Brazil, as studied in Sönmez and Yenmez (2022) and Aygun and Bó (2021). In both cases, affirmative action involves multiple overlapping dimensions: caste and gender in Indian government job allocations, and race and income in Brazilian college admissions. Both mechanisms display similar shortcomings. For instance, in both cases an individual may lose a position due to belonging to a target group for affirmative action (*e.g.*, underrepresented caste, or a low-income family), even though they would have secured the same position had they not belonged to those groups. As a result, declaring affirmative action status becomes a strategic choice, causing a failure of incentive compatibility.

I begin by examining the primary choice rule used in India from 1995 to 2020 for assigning candidates to government positions, a mechanism that resulted in hundreds of court cases and was eventually rescinded due to its flaws. This rule operated through an opaque process, initially assigning candidates to open positions (available to all applicants) and caste-reserve positions (restricted to underrepresented castes) based on caste membership and scores, followed by adjustments to ensure compliance with gender quotas. The rule encountered issues because it partially restricted caste-reserve-eligible women from being considered for open positions reserved for women during the adjustment phase. As a result, men's scores influenced the consideration of caste-reserve-eligible women at open positions reserved for women. Applying tests derived from my characterizations, I demonstrate that this rule causes to cycles in the induced preferences and therefore cannot be rationalized, further highlighting its deficiencies and underscoring the practical relevance of my characterization. Next, I examine the rule used in Brazilian college admissions. Although this rule is rationalizable, the preferences that rationalize it fail to meet an essential criterion: affirmative action monotonicity. Affirmative action monotonicity stipulates that individuals eligible for affirmative action in multiple dimensions should be (weakly) more preferred, reflecting the core principles of affirmative action. By identifying the shortcomings in both systems, these findings clarify the similarities and differences between the two mechanisms, providing insights for policymakers to design rules that better align with the preferences of their constituencies.

Second, my characterizations are useful for determining the rationalizability of choice rules, since one can prove that a rule does not admit any cycles, even when the preferences that rationalize it are not tractable. I study slot-specific priorities, a well-known class of rules that encompasses many practical mechanisms. Different members in this class are rationalized by different underlying preferences. I show that the entire class of slot-specific priorities is rationalizable, without needing to identify the specific preferences behind each distinct rule. Finally, I study the Multidimensional Privileges Choice Rule proposed by Aygun and Bó (2021), and later replaced the initial flawed rule, which ensures that a student is never disadvantaged by declaring affirmative action eligibility. I show that this rule is the unique choice rule rationalizable by preferences that are monotonic in affirmative action, offering further justification for its adoption.

Intersectionality and Substitutes. Next, I study the relationship between diversity preferences and the substitutes condition, an important theoretical property of choice rules necessary for the existence of competitive equilibria and stable matching (Kelso and Crawford, 1982; Roth, 1984; Hatfield and Milgrom, 2005). When identities are multidimensional, institutions can evaluate diversity in multiple ways. For instance, suppose that a company is focused on racial (Black or white) and gender (male or female) diversity. The company considers *intersectionality* if it evaluates the representation of all four cross-sectional groups when assessing the diversity of the workforce. In contrast, if the company evaluates the two dimensions separately, assessing gender diversity by the number of female workers and racial diversity by the number of Black workers, it fails to account for intersectionality, because the same level of "diversity" could be achieved with significantly different levels of representation of any given group.

Institutions and companies often highly value diversity, incorporating it explicitly into their allocation mechanisms and hiring practices, and even publish reports that evaluate the diversity of their workforce. However, reports from many institutions include only the marginal (and not cross-sectional) distribution of their workforce.[2] Similarly, many affirmative action programs in legislatures have quotas for women and minorities, but these policies have typically evolved separately and operate independently (Hughes, 2018). It is also noted that "new candidates who maximally complement incumbents can be preferred by the incumbent elites" (Celis, Erzeel, Mügge, and Damstra, 2014) and "provide a visible cue to voters that politics is diversifying, while minimising the disruption to white male incumbents" (Murray, 2016).[3]

Motivated by this, I study the relationship between how an institution evaluates multidimensional identities and its implications for its allocation and hiring decisions (*i.e.,* choice

[2] Apple (Apple Inclusion & Diversity) and Microsoft (Microsoft Global Diversity & Inclusion Report 2020) report the fraction of employees who belong to different races and genders, while MIT does the same for its student body (MIT Diversity Dashboard). Exceptions include Google (Google Diversity Annual Report 2020), where the cross-sectional distribution of identities is reported in the intersectional hiring section and Stanford University (Stanford Diversity Report). See Figure 1 for examples of different diversity reporting practices.

[3] This issue is the focus of the literature on intersectionality that studies how different identities combine to create various forms of discrimination and privilege, with a particular focus on the experience of Black women in the United States (Crenshaw (2013)).

rules). When diversity is evaluated considering intersectionality, there is a unique representative outcome that matches the population shares. However, many other outcomes can result in the same marginal distribution of characteristics (*e.g.*, the same number of women and minority individuals) while significantly differing in the representation of certain cross-sectional groups. I show that when diversity is evaluated without considering the intersectionality of identities, a dominant group (*e.g.*, incumbent members of a parliament or a company board that form a majority and share similar socioeconomic characteristics) can increase its representation significantly compared to their the population share while still appearing maximally diverse according to a non-intersectional view of diversity. Moreover, I formalize the connection between intersectionality and complementarities by showing that if an institution values diversity without considering intersectionality, the choice rule induced by their preferences fails to satisfy the substitutes condition, a widely studied condition in the market design literature which is crucial for the use of stable matching mechanisms and the existence of competitive equilibrium. These results demonstrate that intersectionality is not only important from an equity perspective but also crucial for ensuring that selection or allocation procedures satisfy the substitutes condition.

Separability in Match Quality and Diversity. Finally, I characterize the choice rules that treat diversity and score domains separately, as these preferences are both natural starting points and commonly used in the analysis of affirmative action policies (see *e.g.,* Chan and Eyster, 2003; Ellison and Pathak, 2021; Dessein, Frankel, and Kartik, 2023; Passaro, Kojima, and Pakzad-Hurson, 2023). Moreover, policies that subsidize firms and schools based on the identities of their workers or students result in additively separable preferences when these institutions' preferences are quasi-linear in money. Specifically, I analyze when an institution's preferences can be represented by a utility function that is additively separable in two components, one that depends on the scores of chosen individuals and the other on the number of chosen individuals from each socioeconomic group. Under this representation, the preference over two individuals can depend on their scores and the representation of their groups, but not on the scores of others or the representation of other groups. Perhaps surprisingly, these rules are characterized by adaptations of three well-known properties: a choice rule is rationalizable by a utility function that is additively separable in the score and diversity domains (with utility increasing in scores and concave in the representation of each group) if and only if it satisfies (within-group) responsiveness (Roth, 1985), substitutes (Roth, 1984), and the acyclicity condition of Tversky (1964). I then map existing choice rules, such as quotas and reserves, to this framework and show that additive separability in the score and diversity domains leads to a unique selection within the broader set of policies that reserve places based on individuals' identities.

Related Literature. A large body of literature on matching with affirmative action and diversity concerns was initiated by Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu (2005). Kojima (2012) studies quota policies, Hafalir, Yenmez, and Yildirim (2013) introduces alternative and more efficient minority reserves, and Ehlers, Hafalir, Yenmez, and Yildirim (2014) generalize reserves to accommodate policies with floors and ceilings. Dur et al. (2018) and Dur et al. (2020) study reserves in public schools in Boston and Chicago while Kamada and Kojima (2017), Kamada and Kojima (2018), and Goto, Kojima, Kurata, Tamura, and Yokoo (2017) study stability and efficiency in more general matching-with-constraints models. This paper contributes to this literature, which focuses on characterizing rules through desirable axioms, by employing the preference-based approach of revealed preference theory. This approach enables us to study the rationality of these rules (and their designers) in the same way as the revealed preference theory studies the rationality of an agent based on observed choices, a central question in economics, as well as various properties of preferences that rationalize the choice rules used in practice.

This paper is also related to the literature on revealed preference theory, building on the results of Richter (1966) and Nishimura et al. (2016). A similar preference-based approach is explored by Echenique and Yenmez (2015), who characterize the preferences that induce choice rules that maximize scores conditional on achieving (or minimizing the distance from) an ideal distribution of characteristics. This paper complements theirs by considering diversity preferences that can freely depend on the distribution of characteristics (without any restrictions or reference to an ideal point) and allowing for flexible trade-offs between scores of individuals and the distribution of characteristics, thereby encompassing a broader class of rules applied in practice.

This paper is also connected to the extensive literature on the substitutes condition in matching markets (Hatfield and Milgrom, 2005; Hatfield and Kojima, 2010; Aygün and Sönmez, 2013). Kojima, Sun, and Yu (2020a) characterize all feasibility constraints that preserve substitutability and Kojima, Sun, and Yu (2020b) complement the analysis of their earlier paper by characterizing when softer pecuniary transfer policies preserve substitutability, building on the theory of discrete convex analysis (Murota, 1998; Murota et al., 2016). Yokote, Hafalir, Kojima, and Yenmez (2024) studies path-independent rules, a stronger version of the substitutes property, and Alva (2018) studies path independence and rationalizability in a combinatorial choice setting.[4] My results complement theirs by explicitly

[4]My setting and results are not directly comparable to Alva (2018) as in his setting, substitutes, rationalizability, and weak axiom of revealed preference are equivalent, whereas I demonstrate that these three properties differ in a significant and practically relevant way within standard revealed preference and market

considering the multidimensional and overlapping structure of types of individuals, focusing on underlying preferences instead of constraints. This approach establishes a novel connection between the substitutes condition and the seemingly unrelated issue of intersectionality in the evaluation of diversity.

Relatedly, this paper builds on the literature addressing affirmative action with multidimensional and overlapping identities, by both extending the theoretical framework and examining the practical applications explored in previous studies. Kurata, Hamada, Iwasaki, and Yokoo (2017) propose a mechanism that is strategy-proof and implements student-optimal matching. Aygun and Bó (2021) study affirmative action policies in which students can qualify for affirmative action in two different dimensions and show that the treatment of overlapping identities in Brazilian university admissions can cause unfairness and incentive compatibility issues. Sönmez and Yenmez (2022) demonstrate the shortcomings of the main mechanism used to assign government positions in India, where protections for overlapping domains play an important role, and propose alternative mechanisms. By showing that the mechanism used in Brazilian college admissions can only be rationalized by preferences that violate the spirit of affirmative action, while the mechanism used in India is not rationalizable at all, my results shed light on the similarities and differences in the deficiencies of these two mechanisms. Finally, in a recent paper, Carvalho, Pradelski, and Williams (2024) study the representativeness of affirmative action policies that do not take into account. Although the models differ, the results in this paper complement theirs by focusing on the properties, rather than the outcomes, of non-intersectional policies.

Chan and Eyster (2003), Ellison and Pathak (2021), Çelebi and Flynn (2022), and Celebi and Flynn (2023) consider a designer with an additively separable utility function in the quality and diversity domains.[5] The analysis in Section 5 complements these papers by analyzing when the choice rules adopted by institutions are rationalizable by a utility function that is additively separable in the quality and diversity domains. Arnosti, Bonet, and Sethuraman (2024) define explainable affirmative action rules and show that they induce the same unique processing order that additive separability induces over reserve policies.

2. Model

Identities and Scores. There are N dimensions that represent the identities of individuals. For each $l \in \{1, \dots, N\}$, Θ_l denotes the finite set of possible groups to which an

design settings.

[5]In two recent papers, Passaro et al. (2023) and Dessein et al. (2023) analyze affirmative action motives in decentralized markets and study utility functions separable in these two domains.

individual might belong in dimension l. I assume that there are at least two groups in each dimension, $|\Theta_l| \geq 2$, and use $\Theta = \Theta_1 \times \ldots \times \Theta_N$ to denote the set of all possible identities.

Example 1. Θ_1 denotes race where $\Theta_1 = \{\text{Asian, Black, Hispanic, White}\}$ and Θ_2 denotes income where $\Theta_2 = \{\text{Rich, Middle class, Poor}\}$. △

Each individual has a score $s \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{R}$ is a finite set of possible scores. $T = \Theta \times \mathcal{S}$ denotes all possible *types* of individuals. For individual i, $\theta_l(i)$ denotes the group of i in dimension l, while $\theta(i) = (\theta_1(i), \ldots, \theta_N(i))$ denotes the *identity* of i. The function $s(i)$ denotes the *score* of i and $t(i) = (\theta(i), s(i))$ denotes the *type* of i. For a set of individuals I, $N_\theta(I)$ denotes the number of individuals with identity θ at I.

Example 1 (continued). To simplify notation, I use the first letter of each group to denote its name. Let $\mathcal{S} = \{0, 1\}$. Then, $\theta(i) = (\text{H, R})$ denotes the identity of i, $s(i) = 1$ denotes the score of i, while $t(i) = ((\text{H, R}), 1)$ denotes the type of i, a rich Hispanic individual with score 1. If $\theta(j) = (\text{W, M})$, then for $I = \{i, j\}$, we have $N_{(\text{H, R})}(I) = 1$ and $N_{(\text{H, P})}(I) = 0$ as the set I has one rich Hispanic individual and zero poor Hispanic individuals. △

Choice Rules. An institution chooses q individuals from a given set of individuals $I \subseteq \mathcal{I}$, where $\mathcal{I}$ denotes the set of all individuals.[6] Formally, a choice rule is a correspondence $C : 2^{\mathcal{I}} \to 2^{2^{\mathcal{I}}}$ such that if $I \in C(I')$, then

(i) $I \subseteq I'$, *i.e.*, I was available for selection at I'
(ii) $|I| \geq \min\{q, |I'|\}$, *i.e.*, the capacity is filled whenever there are enough individuals

Let $\mathcal{T}$ denote all possible type distributions for q or fewer individuals. Formally,

$$\mathcal{T} = \bigcup_{q' \in \{1,\ldots,q\}} \underbrace{T \times \ldots \times T}_{q' \text{ times}}$$

For I with $|I| \leq q$, let $\tau(I) \in \mathcal{T}$ denote the types of individuals in I. Formally,

$$\tau(I) = \{t(i_1), \ldots, t(i_{q'})\}, \text{ for } q' \leq q$$

As is common in all applications considered in this paper, I will assume that institution's choices are anonymous, meaning that they only depend on the scores and identities of individuals, and not on their names.

[6] As I assume in Assumption 1 that only identities and scores (and not the names) of individuals matter for choices, and the institution can choose at most q individuals, we can take $\mathcal{I}$ as any finite set that includes at least q individuals with each type.

Assumption 1. *Suppose that* $\tau(I) = \tau(J)$ *and* $\tau(\hat{I}) = \tau(\hat{J})$, *where* $I \subseteq \hat{I}$ *and* $J \subseteq \hat{J}$. *Then,*

$$I \in C(\hat{I}) \iff J \in C(\hat{J})$$

Preferences. The preferences of the institution are represented by complete preorder $\succeq$ (with the asymmetric part $\succ$ and symmetric part $\sim$) on $\mathcal{T}$.[7] I will slightly abuse the notation and write $I \succeq I'$ instead of $\tau(I) \succeq \tau(I')$. A preference relation $\succeq$ rationalizes C if C always chooses the $\succeq$-maximal sets of individuals, that is, $C(I) = \{I' : I' \succeq I'' \text{ for all } I'' \subseteq I\}$. Similarly, a choice rule C is induced by $\succeq$ if it returns the set of $\succeq$-maximal subsets of I.

This representation is very flexible and can incorporate various forms of diversity preferences. It can account for both the scores and identities of individuals, as well as the distribution of identities within the set. Most importantly, the preference relation $\succeq$ does not need to satisfy responsiveness (Roth, 1985).[8] For example, if I has more individuals with identity $\theta(i)$ the institution and I' has more individuals with identity $\theta(i')$, the institution may prefer i to i' when they are evaluated together with I', that is $i \cup I' \succ i' \cup I'$. This preference may be reversed when they are evaluated together with I, that is, $i' \cup I \succ i \cup I$. Thus, the identities of other chosen individuals can affect how the institution compares i and i', allowing the institutions' preferences to depend on the overall distribution of identities.

Before moving on to the results, I will explain how the standard model in market design, which involves selecting a subset of individuals from a given set, can be mapped to the framework of revealed preference theory.

Choice Environments. A choice environment is an ordered pair $((X, \trianglerighteq), \mathcal{A})$, where $(X, \trianglerighteq)$ is a preordered set and $\mathcal{A}$ is a collection of subsets of X. Here, X represents the consumption set, comprising all possible alternatives, while $\mathcal{A}$ denotes the feasible sets from which the decision maker is observed to make a choice. The observed choice correspondence maps $\mathcal{A}$ to X and encodes the information collected by the observer on the agent's choice behavior. Since a central question in economics is whether choices reflect an underlying preference relation, revealed preference theory studies the properties of choices that are consistent with this behavior. $\trianglerighteq$ is an exogenous dominance relation on X. A preference relation $\succeq$ extends $\trianglerighteq$ if dominance with respect to $\trianglerighteq$ implies the preference, that is, if (i) $I \succeq J$ whenever $I \trianglerighteq J$, and (ii) $I \succ J$ whenever $I \triangleright J$.

[7] Since $\mathcal{T}$ is finite, there is an equivalent utility representation.

[8] Responsive preferences require that the preference between any two individuals does not depend on the rest of the group, that is, for all I with $I \cap \{i, i'\} = \emptyset$, $i \succ i' \implies i \cup I \succ i' \cup I$. Here, $\succ$ restricted to singleton sets represents the primitive preference (*e.g.*, priority order of a school) and preferences over sets of individuals are derived from this order.

Definition 1. *A choice rule is $\trianglerighteq$-rationalizable if there exists a preference relation that rationalizes the choice rule and extends $\trianglerighteq$.*

The exogenous dominance relation is useful because we typically care not only about the existence of underlying preferences (which is recovered by setting $\trianglerighteq = \emptyset$) but also about their specific properties. A natural example is domination with respect to scores. We can define $\triangleright_{\mathcal{S}}$ as the score domination relation where $I \triangleright_{\mathcal{S}} J$ whenever I is obtained by increasing the scores of some individuals in J without changing their identities. If a choice rule is $\triangleright_{\mathcal{S}}$-rationalizable, then it can be rationalized by a preference relation that prefers higher scoring individuals to lower scoring ones, all else equal.

Mapping to Choice Environments. I will now map the standard setting in market design, where a choice rule encodes a complete plan for determining the selected applicants given the set of applicants, to a specific choice environment, $(X^*, \mathcal{A}^*)$. Let 2^I_x to denote all x-element subsets of I. Then

$$X^* = \cup_{k=1,\ldots,q} 2^{\mathcal{I}}_k \tag{1}$$

corresponds to the set of all sets of individuals that has fewer than q elements, denoting all possible alternatives the decision-maker can choose from. A special case of the model focuses on environments without ties, where the choice rule is defined over subsets of $\mathcal{I}$ in which all individuals have different scores. Specifically, instead of X^*, we consider $\tilde{X} = \{I : I \in X^* \text{ and } s(i) \neq s(j) \text{ for all } \{i, j\} \subseteq I\}$.[9]

As the choice rule specifies the set of selected individuals for any given set of individuals, it is reasonable to expect that $\mathcal{A}^*$ includes all possible alternatives and is equal to 2^{X^*}. However, even though the choice rule has enough information to determine the set of chosen individuals in all instances, it imposes a specific structure on feasible sets and induces a choice environment that is not complete. To illustrate, for any $A \in \mathcal{A}^*$, if $I \in A$ and $I' \in A$, then all sets with q or fewer elements that can be formed from $I \cup I'$ must also be in A. The following example illustrates these points.

Example 2. Suppose that $|\mathcal{S}| = 1$, $q = 3$, $\Theta = \{a, b, c\}$. The letters denote the groups of individuals, the subscripts denote different individuals from the same group, *e.g.*, $I = \{a_1, a_2, b_1\}$. Let $A = \{\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\}, \{c_1, c_2, c_3\}\}$. The choice rule does not encode any information about what the choice would be from the three possible outcomes in A, since whenever these sets are available, all three element subsets of $\{a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3\}$ are also available for selection. △

[9]This can be interpreted as using scores obtained after tie-breaking, in cases where ties in scores are possible.

The next Proposition characterizes the choice environment induced by choice rules. This allows us to use results from revealed preference theory to study the rationality of rules that allocates resources, and also shows that the standard model in market design can be mapped to a particular choice environment.

Proposition 1. *$\mathcal{A}^*$ is the largest subset of 2^{X^*} such that for any $A \in \mathcal{A}^*$, if $I \in A$ and $I' \in A$, then all subsets of $I \cup I'$ that have q or fewer elements are also in A.*

3. Rationality of Choice Rules

I now characterize choice rules that can be rationalized by preferences that satisfy certain conditions, as this provides a clearer understanding of the relationship between choice rules and preferences. To this end, I define a $\trianglerighteq$-cycle, the key property that determines rationality.

Definition 2. *$I_1, \ldots, I_n$ is a $\trianglerighteq$-cycle if*

- *for each $k < n$, either there exists an $\hat{I}_k$ such that $I_k \in C(\hat{I}_k)$ and $I_{k+1} \subset \hat{I}_k$, or $I_k \trianglerighteq I_{k+1}$;*
- *either there exists $\hat{I}_n$ such that $I_n \in C(\hat{I}_n)$, $I_1 \subset \hat{I}_n$ and $I_1 \notin C(\hat{I}_n)$, or $I_n \triangleright I_1$.*

An important special case of $\trianglerighteq$-cycle is obtained by setting $\trianglerighteq$ as the empty relation.

Definition 3. *$I_1, \ldots, I_n$ is a choice cycle if*

- *for each $k < n$, there exists an $\hat{I}_k$ such that $I_k \in C(\hat{I}_k)$ and $I_{k+1} \subset \hat{I}_k$,*
- *there exists $\hat{I}_n$ such that $I_n \in C(\hat{I}_n)$, $I_1 \subset \hat{I}_n$ and $I_1 \notin C(\hat{I}_n)$.*

The existence of a choice cycle corresponds to a violation of the *congruence* axiom of Richter (1966). If there is a choice cycle under C, then the institution has chosen I_1 when I_2 was available, ..., I_{n-1} when I_n is available. Therefore, I_1 is indirectly (weakly) revealed to be preferred to I_n. The fact that I_n is chosen when I_1 is available and is not chosen means that I_n is directly (strictly) revealed to be preferred to I_1.

A $\trianglerighteq$-cycle has the additional requirement that the choices do not violate the exogenous relation $\trianglerighteq$. The concept of $\trianglerighteq$-cycle is adapted from Nishimura et al. (2016), who generalize the characterization of Richter (1966) to settings with an exogenous order. Applying their result, we can characterize the choice rules that are rationalizable by a preference relation.

Theorem 1. *There exists a preference relation $\succeq$ that rationalizes C and extends $\trianglerighteq$ if and only if C does not admit a $\trianglerighteq$-cycle.*

Moreover, by setting $\trianglerighteq$ as the empty relation, we obtain the following corollary.

Corollary 1. *There exists a preference relation $\succeq$ that rationalizes C if and only if C does not admit a choice cycle.*

In the next sections, I will apply Theorem 1 and Corollary 1 in two different ways. First, by establishing cycles, we can deduce that certain choice rules used in practice cannot be rationalized by preferences. Consequently, Corollary 1 acts as a minimal requirement that a choice rule must satisfy. For rationalizable rules, Theorem 1 helps us determine whether the preferences behind those rules satisfy other desirable properties. Second, I use these results to prove which (classes of) choice rules can be rationalized by showing that they could never induce a choice cycle. This is useful because, for many choice rules, constructing the preferences that rationalize them is not straightforward.

Properties of the Choice Environment and the Weak Axiom. Before moving to the applications, I describe why weaker axioms, such as the weak axiom of revealed preference (WARP) do not characterize rationalizability for choice rules.[10] Example 2 demonstrated that even though the choice rule determines the chosen individuals from each set of candidates, the choice environment it induces is far from complete. In fact, if the induced choice environment was complete, then rationality would be characterized by the weak axiom of revealed preference.[11] The following example illustrates how choice rules can satisfy the weak axiom without being rationalizable.

Example 2 (continued). Consider following choice rule

- If there is at least one individual from each group, choose exactly one individual from each group.
- If choosing one individual from each group is not possible, then at least one group is not represented in the set of available individuals.
  - (i) If there are no c individuals, choose as many a individuals as possible.
  - (ii) If there are no a individuals, choose as many b individuals as possible.
  - (iii) If there are no b individuals, choose as many c individuals as possible.

Note that (i) implies $C(a_1, a_2, a_3, b_1, b_2, b_3) = \{a_1, a_2, a_3\}$, (ii) implies $C(b_1, b_2, b_3, c_1, c_2, c_3) = \{b_1, b_2, b_3\}$ and (iii) implies $C(a_1, a_2, a_3, c_1, c_2, c_3) = \{c_1, c_2, c_3\}$. In Appendix B.1, I prove that C does not fail the weak axiom. This is because the choice environment induced by the choice rule does not include $A = \{\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\}, \{c_1, c_2, c_3\}\}$, which would have created a

[10]Weak axiom requires that if $a \in C(A)$ where $a' \in A$, then $a \in C(A')$ for all A' such that $a' \in C(A')$.

[11]Caradonna (2020) introduces the property of well-coveredness, characterizes the choice environments for which the weak axiom does characterize rationalizability, and shows that this class includes not only the complete choice environment, but also other considerably smaller environments.

violation of weak axiom regardless of the choice made from it. In fact, we did not even need to specify what the choice rule would do in this case, since $A \notin \mathcal{A}^*$.

Moreover, the preference relation that "rationalizes" C must have $\{a_1, a_2, a_3\} \succ \{b_1, b_2, b_3\} \succ \{c_1, c_2, c_3\} \succ \{a_1, a_2, a_3\}$, which means that C is not rationalizable even though it does not fail the weak axiom. △

3.1. *Application: Rationalizability and The Supreme Court Mandated Choice Rule in India.*

Indian government operates one of the largest affirmative action systems in the world to promote the representation of historically discriminated groups in various areas, including government jobs and universities. Affirmative action is enshrined in the Indian constitution, and its implementation procedure in the allocation of government jobs was outlined in the important Supreme Court judgment, Anil Kumar Gupta v. State of U.P. (1995).[12] The two main classes of reservations are based on caste, to promote the representation of historically underrepresented groups such as Scheduled Castes, Scheduled Tribes, and Other Backward Classes, and on gender, to enhance the representation of women.

The procedure devised by the Supreme Court suffered from significant shortcomings, causing hundreds of court cases at various levels of the Indian judiciary and resulting in its demise in 2020 after over 25 years of use. Most notably, its poor design caused some individuals to lose their gender-based protections if they claimed caste-based protections, even if the caste-based protection did not secure them a position. The procedure involved a complex set of reserved positions for different groups. Affirmative action based on caste *sets aside* a certain number of positions for individuals belonging to specific castes. Affirmative action based on gender *requires* that within each caste reservation, as well as for the *open* positions that are not reserved for any caste, certain number of women to be chosen.

The problem with this rule stems from the fact that candidates eligible for caste reservations are only considered for the open positions if they rank high enough in the general population. In particular, if the number of open positions is n, then they must have one of the n highest scores to be considered for the open positions. This would not create a problem if caste and gender were not overlapping identities, since candidates outside the top n would not be selected for the open positions regardless. However, when caste-based set-asides are combined with gender-based minimum guarantees, the choice rule produces outcomes that conflict with the intent of affirmative action.

[12]See Sönmez and Yenmez (2022) for more details on this judgement and background on affirmative action in India, as well as detailed description and analysis of the setting.

I will now demonstrate that this issue prevents the choice rule from being rationalizable. To this end, I will define the setting and the choice rule mandated by the Supreme Court of India (C_S) in a context involving two groups and two dimensions. Let $\Theta_1 = \{g, r\}$ where g denotes the general population and r denotes the reserve eligible population (individuals belonging to underrepresented castes). Let $\Theta_2 = \{m, w\}$ denote the gender of the individuals, with m representing men and w representing women. For $x \in \{g, r, m, w\}$, I^x denotes the set of individuals in a given category. In this setting, C_S is characterized by 4 integers, r: the number of reserved positions, o: the number of positions open to everyone, $r_w \leq r$: the number of reserved positions protected for women, $o_w \leq o$: the number of open positions protected for women. The choice rule C_S proceeds as follows:

Supreme Court Mandated Choice Rule C_S

Step 1: Define $\mathcal{M}$ as the set of reserve-eligible candidates who are among the o highest scoring individuals in the population.

Step 2: Assign o_w positions to o_w highest scoring women in $I^g \cup \mathcal{M}$.[13]

Step 3: Assign remaining $o - o_w$ positions to $o - o_w$ highest scoring previously unassigned individuals in $I^g \cup \mathcal{M}$.

Step 4: Assign r_w positions to r_w highest scoring previously unassigned woman in I^r.

Step 5: Assign $r - r_w$ positions to highest scoring previously unassigned individuals in I^r.

$\mathcal{M}$ is referred to as the *meritorious reserve candidates.* As noted, C_S has some important shortcomings. It does not satisfy *no justified envy*; a reserve eligible individual can score higher than a general category individual and yet fail to receive a position, while the general category individual receives one. This outcome contradicts the philosophy of affirmative action. Moreover, this situation leads the reserve eligible individual to obtain the position by not disclosing their reserve eligibility, thereby violating *incentive compatibility.*

Example 3. There are three positions ($q = 3$), two of them are open ($o = 2$) with one protected for women ($o_w = 1$) and the remaining caste-reserve position is also protected for women ($r = r_w = 1$). The table below shows the applicants' identities and scores, where the first letter denotes gender and the second denotes reserve eligibility.

[13] If there are fewer than o_w or r_w women considered at Steps 2 or 4, then remaining positions are assigned to highest scoring men who are considered at that stage.

Applicants	Score	Meritorious Reserve	Admission		
			Open (Woman)	Open	Reserve (Woman)
mg	6			✓	
wr	4	✓	✓		
mr	3				
wr	2				✓
wg	1				

First, the only caste reserve eligible candidate considered for open positions is $(wr, 4)$, as she is among the top $o = 2$ in the score distribution. Then $(wr, 4)$ is chosen for the open position protected for women. The other open position is assigned to the highest scoring candidate that is considered at this point, $(mg, 6)$. Finally, $(wr, 2)$ is assigned the reserved position as the highest scoring reserve eligible woman. Therefore, the set $I_1 \equiv \{(mg, 6), (wr, 4), (wr, 2)\}$ is chosen, while $I_2 \equiv \{(mg, 6), (wr, 4), (wg, 1)\}$ was available and but not chosen. We will now consider an alternative case where the score of $(mr, 3)$ is increased to $(mr, 5)$.

Applicants	Score	Meritorious Reserve	Admission		
			Open (Woman)	Open	Reserve (Woman)
mg	6			✓	
wr	4				✓
mr	**5**	✓			
wr	2				
wg	1		✓		

After this change, $(mr, 5)$ becomes the meritorious reserve candidate and is the only reserve eligible candidate considered for open positions. Then $(wg, 1)$ is chosen for the open position protected for women, as she is the only woman considered at this stage. The other open position is assigned to the highest scoring candidate that is considered at this point, $(mg, 6)$. Finally, $(wr, 4)$ is assigned the reserve position as the highest scoring reserve eligible women. However, since $\{(mg, 6), (wr, 4), (wg, 1)\}$ is chosen while $\{(mg, 6), (wr, 4), (wr, 2)\}$ was available but not chosen, the sets I_1 and I_2 constitute a choice cycle. △

Example 3 suggests to the following proposition, which shows that C_S is not rationalizable regardless of the number of positions of different types.

Proposition 2. *The Supreme Court Mandated Choice Rule is not rationalizable.*

Therefore, the choice rule mandated by the Supreme Court not only suffers from the issues identified in prior literature, could not have been designed by a rational decision maker with well-defined preferences. The lack of rationality stems from the fact that the court's guidance outlines a complex allocation rule that combines multiple considerations in an ad-hoc fashion. In fact, the actual rule described in the court decision first allocates resources based on caste and then adjusts the allocation to meet gender quotas. This results in a more intricate rule than the one presented here, even though both are outcome-equivalent.[14]

3.2. Slot-Specific Priorities and Reserve Rules are Rationalizable

The previous section demonstrated that focusing on preferences rather than specific rules will increase transparency and result in more sensible and effective allocation mechanisms. In this section, I show that a very general and widely used class of rules, slot-specific priorities (Kominers and Sönmez, 2016), is rationalizable. Slot specific priority rules are defined by q slots $\sigma = \sigma_1, \ldots, \sigma_q$ such that each slot is assigned to one individual. For each slot σ_l, there is a priority order over individuals $\geq_{\sigma_l}$ such that higher priority individuals are chosen before lower priority ones, where $\geq_\sigma = \{\geq_{\sigma_1}, \ldots, \geq_{\sigma_q}\}$, with $>_{\sigma_l}$ denoting the strict part of this order. The priorities in each slot can depend on the scores of individuals as well as their identities. Following Kominers and Sönmez (2016), I focus on strict priorities. To this end, I assume that the choice environment is without ties, and $\geq_\sigma$ ranks any two individuals strictly whenever they do not have the same score.[15]

The slots are processed according to a *precedence order* which orders slots from the first to the last. At each step, the slot is assigned to the highest priority individual who has not already assigned a preceding slot. As the priority orders are unrestricted, it is without loss of generality to assume that the precedence order is given by the subscript.

Definition 4. *A choice rule C is generated by slot-specific priorities $(\sigma, \geq_\sigma)$ if given I, the chosen individuals are determined in q steps as follows. Set $I_1 = I$*

- *In Step k, choose the individual with the highest priority according to $>_{\sigma_k}$ from I_k.*
- *Let i_k denote the individual chosen in step k. Set $I_{k+1} = I_k \setminus i_k$ and move to step $k+1$.*

Different slot-specific priorities are rationalized by different preferences. Indeed, in Section 5, I will show that the some slot-specific rules could be rationalized by preferences that

[14] Sönmez and Yenmez (2022) defines the much simpler choice rule C_S studied here and shows the outcome equivalence between the two formulations.

[15] This can be interpreted as the scores being obtained after breaking any ties. In this setting, the highest score can be attained by individuals with different identities, but not simultaneously. An alternative approach to model strict priorities is to restrict attention to strict scores, but this requires adapting definitions and introducing further notation.

are additively separable in score and diversity domains, while some cannot be rationalized by such preferences. Applying Theorem 1, we can directly show that slot-specific priorities are rationalizable without tackling the task of recovering preferences underlying infinitely many members of this class.

Proposition 3. *If C is generated by slot-specific priorities, then it is rationalizable.*

To prove Proposition 3, I first assume that there is a choice cycle $I_1, \ldots, I_n$ and identify the individuals assigned to each slot at each step of the cycle. I then show that at each step of the cycle, for each slot, the slot specific priority of the individual assigned to that slot is (weakly) decreasing. Therefore, it is not possible for the set I_n to be chosen when I_1 was available but not chosen.

I now define a special case of slot specific priorities, *generalized reserves*, which are widely used in practice. A choice rule is generated by generalized reserves if each slot is either set aside for a particular set of groups, where individuals who belong to those groups have higher priority than those who do not, or is an open slot where priority is equal to the score.

Definition 5. *A generalized reserve rule is given by slots σ and a function $h : \sigma \to 2^{\Theta}$ that assigns each slot σ_l to a set of identities (including the $\emptyset$, which means that no identity is favored at that slot). Given σ, h, the at each slot, $i >_{\sigma_l} j$ if*

1. *$\theta(i) \in h(\sigma_s)$ and $\theta(j) \notin h(\sigma_s)$, that is, i qualifies for the reserve while j does not.*
2. *$\theta(i) \in h(\sigma_s)$, $\theta(j) \in h(\sigma_s)$ and $s(i) > s(j)$, that is, both qualify for the reserve and i has a higher score.*
3. *$\theta(i) \notin h(\sigma_s)$, $\theta(j) \notin h(\sigma_s)$ and $s(i) > s(j)$, that is, both do not qualify for the reserve and i has a higher score.*

A generalized reserve rule is a *reserve rule* if each $h(\sigma_s)$ is either a singleton or the empty set, meaning that, each slot is reserved for at most one identity. Since generalized reserves are special cases of slot specific priorities, we have the following corollary.

Corollary 2. *Choice rules generated by generalized reserve policies are rationalizable.*

3.3. Application: Affirmative Action Monotonicity and Brazilian College Admissions

Corollary 1 considers the identities of individuals without making affirmative action motives explicit. However, rationalizability is only a first step towards effective choice rules. I now illustrate how Theorem 1 allows for incorporating additional considerations to the preferences. For instance, the main goal of affirmative action is to provide advantages to certain underrepresented groups, thereby increasing their representation.

To study preferences that depend on affirmative action motives explicitly, I focus on a special case of my model where each dimension corresponds to a dimension of affirmative action, and has two groups, the individuals who are eligible for affirmative action in that dimension and those who are not eligible.[16] Formally, each Θ_l corresponds to a dimension of affirmative action and $\Theta_l = \{\theta_l^A, \theta_l^N\}$. θ_l^A denotes the target group of affirmative action in dimension l, while θ_l^N denotes the individuals who do not belong to the target group. Aygun and Bó (2021) define a key property (that they call privilege monotonicity) that require that if a student is not chosen, then she would not be chosen if she were eligible for affirmative action in more domains. If the choice rule does not satisfy this property, lower achieving students who do not qualify for affirmative action could be chosen instead of higher achieving students who do. Consequently, it may be optimal for some students to not declare their affirmative action status, making it a strategic choice.

Motivated by this, a reasonable criterion for preferences is affirmative action monotonicity: keeping everything constant, increasing affirmative action eligibility of a set of individuals makes the set more desirable.

Definition 6. *$I \trianglerighteq_A I'$ if there exists a bijection $\rho : I \to I'$ such that for all $i \in I$, $s(i) = s(\rho(i))$ and $\theta_l(\rho(i)) = \theta_l^A \implies \theta_l(i) = \theta_l^A$.*

A choice rule C is rationalizable by an *affirmative action monotonic* preference relation if there exists a preference relation $\succeq$ that rationalizes C and extends $\trianglerighteq_A$.

Multidimensional Identities in Brazilian College Admissions. Brazilian public universities are mandated to use affirmative action policies for candidates from racial and income minorities by a federal law enacted by the Brazilian Congress in 2012.[17] Mapping this setting to the model, the two dimensions $\Theta_R = \{R_A, R_N\}$ denote students who are and are not eligible for affirmative action in race, and $\Theta_I = \{I_A, I_N\}$ denote students who are and are not eligible for affirmative action in based on income. This affirmative action mandate was implemented by a reserve policy called Brazil Reserves, which reserves seats for three types of students, those who are eligible in both dimensions, (R_A, I_A), those eligile only in race dimension, (R_A, I_N), and those only eligible only in income dimension (R_N, I_A).[18]

[16] This framework is previously formulated by Aygun and Bó (2021) to study affirmative action in Brazilian College Admissions.

[17] To qualify for these positions, students also need to attend a public high school in Brazil. For simplicity, we abstract away from this issue, which does not affect any of our results and conclusions. See Aygun and Bó (2021) for a more detailed description of this setting.

[18] This is a slot-specific priority where $h(\sigma) \in \{(R_A, I_A), (R_A, I_N), (R_N, I_A), \emptyset\}$ for all σ and slots that do not map to $\emptyset$ have the same (strictly positive) number for the three groups.

Example 4. There is one reserve slot for each identity in $\{(R_A, I_A), (R_A, I_N), (R_N, I_A), \emptyset\}$ and $q = 4$. The applicants' identities and scores are given in the following table, where the first letter denotes eligibility in dimension one, while the second letter denotes eligibility in dimension two. The processing order of the slots is from left to right.

Applicants	Score	Admission			
		Open	A, A	A, N	N, A
N, N	5	✓			
A, N	4			✓	
A, A	3		✓		
N, A	2				✓
N, A	1				

This implies that $\{(N, N, 5), (A, N, 4), (A, A, 3), (N, A, 2)\} \equiv I_2$ is chosen when the alternative set $\{(N, N, 5), (A, N, 4), (A, A, 3), (N, A, 1)\} \equiv I_3$ was available. Now, consider a slightly modified version of this where $(N, A, 2)$ is also eligible in dimension one.

Applicants	Score	Admission			
		Open	A, A	A, N	N, A
N, N	5	✓			
A, N	4			✓	
A, A	3		✓		
A, A	2				
N, A	1				✓

First, let $\{(N, N, 5), (A, N, 4), (A, A, 3), (A, A, 2)\} \equiv I_1$ and note that $I_1 \trianglerighteq_A I_2$ as one individual is eligible in more dimension at I_1. Second, I_3 was chosen when I_1 was available and not chosen. Therefore, the Brazil Reserves choice rule induces the following $\trianglerighteq_A$-cycle: I_1, I_2, I_3. △

The cycle in Example 4 is caused by the fact that Brazil Reserves does not allow for $(A, A, 2)$, an individual who is eligible in both dimensions to obtain the position that only has protections for one dimension. The next proposition shows that Brazil reserves cannot be rationalized by an affirmative action monotonic preference relation regardless of the number of reserved seats and the processing order.

Proposition 4. *The Brazil Reserves choice rule cannot be rationalized by a affirmative action monotonic preference relation.*

Comparing Affirmative Action Mechanisms in India and Brazil. Proposition 2 and Proposition 4 shed light on an important difference between the problems in the affirmative action rules implemented in India and Brazil. While both settings appear to suffer from the same issues of failures of affirmative action monotonicity and incentive compatibility, the choice rule mandated by the Indian Supreme Court is not rationalizable, and therefore cannot be designed by a utility maximizing decision maker with well defined preferences. The choice rule in Brazil is a generalized reserve rule, and therefore is rationalizable. The problem with that rule is that the preferences that rationalize that rule are not affirmative action monotonic.

Multidimensional Privileges Choice Rule. To remedy the problems in the Brazil Reserves rule, Aygun and Bó (2021) formulates another class of choice rules, *multidimensional privileges choice rules.* An identity θ *dominates* θ' if $\theta'_l = \theta^A_l$ implies $\theta_l = \theta^A_l$. Multidimensional privileges choice rules are generalized reserve rules that satisfy the following property: whenever a slot is reserved for an identity θ, it is also reserved by any other identity θ' that dominates it.

Definition 7. *A generalized reserve rule is a multidimensional privileges choice rule if whenever $\theta \in h(\sigma)$ and θ' dominates θ, then $\theta' \in h(\sigma)$.*

The following result shows that if the designer values individuals who are eligible for affirmative action in more dimensions, the multidimensional privileges choice rule emerges as the unique rule within the class of generalized reserve rules.

Proposition 5. *Multidimensional privileges choice rule is the unique generalized reserve rule that can be rationalized by affirmative action monotonic preferences.*

4. Intersectionality, Substitutes and Representation

4.1. Background and Motivation

When identities are one dimensional, the number of individuals from each group in that single dimension determines diversity. However, when identities are multidimensional, evaluating diversity is a more complex issue. For instance, if an institution aims to consider diversity in terms of both gender (where the groups are men and women) and race (where the groups are majority and minority), it can assess its diversity by counting the number of men and women for gender diversity, and the number of majority and minority individuals

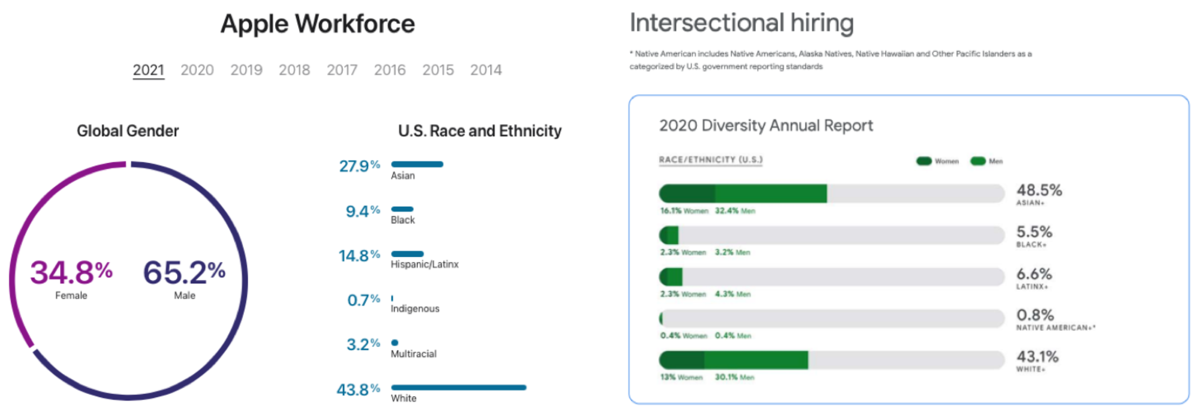

Fig. 1. **Diversity Reporting Practices of Apple and Google.** Apple (left pane) provides a breakdown of marginal distribution of workers in race and gender without any information on the cross-sectional distribution. Google (right pane) reports the percentages of men and women for each race.

for racial diversity. This approach, which focuses on the *marginal* distribution of characteristics, overlooks the intersectionality of identities, as it lacks information on the cross-sectional distribution of characteristics, such as the number of minority women.

Moreover, both marginal and intersectional evaluations of diversity are prevalent in practice. Figure 1 presents the diversity reporting practice of Apple (which reports only marginal distributions in gender and race) and Google (which reports the gender breakdown for each race) to illustrate the difference between preferences that depend on marginal and cross-sectional distributions of identities.[19] Policies designed to advance the representation of women and/or minority groups are also prevalent in politics. A large literature in political science studies the effect of these policies, emphasizing that these policies have typically evolved separately and operate independently Hughes (2018). Relatedly, Celis et al. (2014) discusses the *complementarity advantage*, where new candidates who "maximally" complement incumbents can be preferred by the incumbent elites. In particular, they discuss the example of younger minority women as complementing incumbent white male majority, allowing for maximal representativeness of the list while including a limited number of newcomers and maximizing the power of incumbent elites.[20]

[19] See Footnote 2 for more examples of intersectional and non-intersectional diversity reporting practices.

[20] Other works that emphasize this mechanism include Hughes (2011), who notes "…adding minority women to the national legislature helps to satisfy both gender and minority quotas; their election unseats fewer majority men," Murray (2016), who writes "…providing a visible cue to voters that politics is diversifying, while minimising the disruption to white male incumbents," and Celis and Erzeel (2017) who notes "the complementarity of newcomers does not only leave the position of incumbents unharmed, but even reinforces and re-establishes the latter's power."

Results. Motivated by these observations, the next section examines the relationship between diversity evaluation and the properties of choice rules. When intersectionality is not considered in diversity evaluation, a given level of "marginal" diversity can be achieved through various cross-sectional distributions of identities, and the choices display complementarities among individuals from different groups across dimensions.

To formally establish the impact of intersectionality, I first show that when preferences do not account for intersectionality, the resulting choice rules violate the substitutes condition—a widely studied property crucial for the existence of stable allocations and the application of stable matching mechanisms (Hatfield and Milgrom, 2005), which requires institutions to treat individuals as substitutes rather than complements. I then show that how these complementarities can be exploited by the dominant group to increase their representation without compromising the perception of (marginal) diversity, characterize the extent of this over-representation, and demonstrate that this arises from a particular, non-intersectional view of diversity.

4.2. Analysis with Homogeneous Scores

I start the analysis with individuals who are homogeneous in terms of quality, assuming $|\mathcal{S}| = 1$ and suppressing scores. Appendix B.2 allows $|\mathcal{S}| > 1$, and extends the results using the gross substitutes property of Kelso and Crawford (1982).[21]

Given a type profile $\tau \in \mathcal{T}$, $M_l(\tau)$ returns the number of individuals in each group in dimensions l and $M(\tau) = (M_1(\tau), \ldots, M_N(\tau))$. I will write $M(I)$ instead of $M(\tau(I))$ to simplify the notation. $M(I)$ is the *marginal distribution* of I, as it returns the number of individuals that belong to each group in each dimension, but does not have any information about the cross-sectional distribution of groups.[22] A preference relation $\succeq$ does not consider intersectionality if the marginal distribution of identities is a sufficient statistic to evaluate diversity.

Definition 8. *$\succeq$ does not consider intersectionality if for all I and I' with $M(I) = M(I')$, $I \sim I'$.*

Observe that if the identity is one dimensional ($|N| = 1$), the marginal distribution is sufficient to determine the composition of identities in a group. Therefore, intersectionality matters when $|N| \geq 2$.

[21] In that setting, scores can be interpreted as qualities of the individuals, as well as (inverse) salaries for individuals of homogeneous quality.

[22] For example, if the gender dimension consists of men and women, and the racial dimension consists of white and black individuals, the vector $M(I)$ includes the numbers of men, women, white and black individuals in I, but does not include, for example, white men in I.

Definition 9. *Let* $\tilde{I} \subseteq I' \subset I$. *C satisfies the substitutes condition if for all* $\tilde{I}$ *with* $\tilde{I} \subseteq \hat{I} \in C(I)$, *there exists* $\bar{I}$ *such that* $\bar{I} \in C(I')$ *and* $\tilde{I} \subseteq \bar{I}$.

This condition is the generalization of the substitutes condition of Roth (1984) to choice correspondences.[23] Substitutes condition states that whenever I^* is chosen from I, then I^* is also chosen from any set $I' \subset I$ that includes it. The following example illustrates the relationship between intersectionality and the substitutes condition.

Example 5. There are two dimensions, gender denoted by $\Theta_1 = \{M, W\}$ and race denoted by $\Theta_2 = \{u, o\}$, where u stands for individuals from underrepresented groups and o stands for individuals from overrepresented groups. Let $q = 4$ and suppose that the institution has preferences $\succeq$ that does not consider intersectionality and strictly prefers to have exactly two individuals from all four groups to any other distribution, attaining equal share of all "marginal" characteristics. First, suppose that $I = \{uM_1, uM_2, uW, oM, oW_1, oW_2\}$, where the first letter denotes the race and second letter denotes the gender. The most preferred distribution of characteristics can be achieved in two ways, as shown in the below table, where boxed individuals are chosen.

	M	**W**
u	[uM_1], [uM_2]	uW
o	oM	[oW_1], [oW_2]

	M	**W**
u	[uM_1], uM_2	[uW]
o	[oM]	[oW_1], oW_2

Now suppose that oM does not apply and consider $I' = \{uM_1, uM_2, uW, oW_1, oW_2\}$. In this case, The most preferred distribution of characteristics can be achieved in only one way, which does not include uW.

	M	**W**
u	[uM_1], [uM_2]	uW
o		[oW_1], [oW_2]

As uW is chosen from I but not from $I' \subset I$, the choice rule induced by these preferences does not satisfy the substitutes condition. △

[23] When C is a choice function (*i.e.*, $C(I)$ is singleton for all I), this condition is equivalent to the following: If $i \in C(I)$ and $I' \subseteq I$, then $i \in C(I')$. A similar generalization is employed by Kojima et al. (2020b) for a model with salaries.

In Example 5, the institution evaluates diversity by marginal distributions which cause uW and oM to become *complements*: when uW is chosen, individuals who belong to opposite groups in both dimensions become more desirable. Therefore, when oM is not available, choosing uW cannot be optimal, since no two individuals from I' can complement uW to achieve the preferred distribution.

Intersectionality and Substitutes Condition. First, I make a minimal assumption that makes diversity preferences nontrivial and assume that the most preferred distribution does not completely exclude a group from being chosen.[24] Formally, $M(I)$ is *at boundary* if I has no individual from some group, *i.e.*, there exists l and $\hat{\theta} \in \Theta_l$ such that $\theta_l(i) \neq \hat{\theta}$ for all $i \in I$. Conversely, $M(I)$ is *interior* if it is not at boundary, *i.e.*, if I has at least one individual from each group.

Definition 10. *$\succeq$ values diversity if there is no I such that $M(I)$ is at boundary and $I \succeq I'$ for all I'.*

This is a reasonable assumption for diversity preferences; it requires that the institution values diversity and prefers to choose at least one individual from each group, but puts no other restrictions on how it values different compositions of individuals. The following proposition shows that not considering intersectionality when evaluating diversity causes failure of the substitutes condition.

Theorem 2. *Suppose that $|N| \geq 2$, $\succeq$ values diversity, does not consider intersectionality, and induces $C_{\succeq}$. Then $C_{\succeq}$ does not satisfy the substitutes condition.*

Theorem 2 shows that the logic of Example 5 is indeed much more general. Evaluating diversity with the marginal distributions creates complementarities between certain types, and considering intersectionality is not only crucial from an equity standpoint but also necessary to satisfy the substitutes condition. The proof of this result starts with an arbitrary $\succeq$. It then proceeds to determine a particular distribution of groups, similar to the "most" diverse outcome in Example 5 and two individuals that complement each other the way uW and oM complement each other in Example 5 to construct a violation of the substitutes condition.

[24] If the preferences does not depend on the identities of individuals, then in this setting the institution is indifferent between all individuals, while in the setting with heterogeneous scores (see Appendix B.2) it prefers the individuals with highest scores regardless of their identity. Clearly in those cases, the induced choice rule does not exhibit complementarities based on identities and satisfy the substitutes condition.

Intersectionality and Representation of Incumbent Groups. Motivated by how non-intersectional views of diversity can contribute to over-representation of powerful incumbent groups, I study how much a group can increase its representation in a representative allocation when diversity is evaluated without taking intersectionality into account.

For this section, I will focus on identity/type distribution of sets of q individuals, denoted by $\mathcal{T}^q \subset \mathcal{T}$. I use $\tau \in \mathcal{T}^q$ to denote a generic element of this set and τ_θ to denote the number of individuals with identity θ at distribution τ. Let $\tau^* \in \mathcal{T}^q$ denote the *representative* identity distribution. For example, τ^* may reflect the allocation that matches the percentages of each cross-sectional group in the population, or any other target distribution.[25] Although τ^* uniquely determines the representative distribution of identities, there are many allocations that induce the same level of diversity in the marginal distribution of characteristics.

Example 5 (continued). Suppose that the institution still prefers to have equal representation, but this time considers intersectionality. Then the representative distribution has one individual from all 4 identities, $\{uM, uW, oM, oW\}$ and pins down the allocation. However, there are other distributions with different representation of different groups, such as $\{uM_1, uM_2, oW_1, oW_2\}$, that induce the same marginal distribution as the representative distribution. $\triangle$

A distrbution of identities τ is *marginally representative* if $M(\tau) = M(\tau^*)$. A marginally representative distribution will be as representative as the representative distribution when preferences does not take intersectionality into account. Let θ_l denote the group of identity θ at dimension l. Given the representative distribution τ^* and an identity θ, marginal representation of θ at dimension l is $\mu_l(\theta, \tau^*) = \sum_{\theta':\theta'_l=\theta_l} \tau_{\theta'}$. For instance, in the setting of Example 5, if the marginal representation of oM in gender dimension is $\mu_1(oM, \tau^*) = \tau^*_{oM} + \tau^*_{uM}$. Given τ^*, define the *minimal marginal representation* of θ at τ^* as $\min_l \mu_l(\theta, \tau^*)$.

Proposition 6. *Given the representative distribution τ^* and an identity θ, there exists an alternative, marginally representative distribution τ' that increases the representation of θ from τ^*_θ to its minimal marginal representation,* $\min_l \mu_l(\theta, \tau^*)$.

Proposition 6 characterizes the upper bound for the representation of a group in a distribution that looks as diverse as the representative population under a non-intersectional view of diversity.

Corollary 3. *Suppose that there are two dimensions with two groups each and the representative distribution have equal representation of all groups. The representation of a group in the representative distribution is* $1/4$, *whereas its minimal marginal representation is* $1/2$.

[25] For simplicity, I abstract from issues related to integer constraints.

Corollary 3 shows that when preferences does not take intersectionality account, then powerful incumbent group can increase its representation to half of the seats from a quarter of the seats while still matching the diversity in the representative distribution. Of course, in practice, other factors such as institutional constraints, political frictions, and external pressures prevent a perfect alignment with this theoretical result. Nevertheless, Proposition 6 and Corollary 3 formalizes the relationship between evaluation of diversity and complementarities across individuals, and characterizes the extent of over representation dominant groups can achieve when diversity is evaluated in a non-intersectional way.

5. Separable Utility Representation

The analysis in the previous sections did not impose any structure on the trade-offs between quality and diversity. This section characterizes the choice rules that can be rationalized by preferences additively separable in score and identity. These preferences are represented by a utility function where the utility from each set of individuals is equal to the sum of two terms. The first term depends only on the scores of the individuals, while the second term depends only on their identities. I study this class for three main reasons.

First, in many contexts, the domains of score (match quality) and identity (diversity) are not intrinsically connected. Institutions tend to value workforce diversity primarily for equity reasons, not because an individual's contribution to productivity is closely linked to their identity.[26] Moreover, government policies that subsidize firms and schools based on the identity of their workers or students lead to preferences that are additively separable in scores and diversity when their utility function is quasi linear in money and has no inherent preference for diversity.[27]

Second, it provides a highly tractable approach to considering trade-offs between diversity

[26]In some settings, these domains may as well be connected. For example, if the identities are front-end and back-end developers, a firm might want to have enough high quality employees in both sets. Thus, our model is a better fit when identities do not affect how the scores of individuals contribute to the preferences. Moreover, additive separability can still account for preferences that condition the choice of individuals from one group to the relative performance of the other group. For example, it could admit minority students only if they score high enough relative to majority students.

[27]Examples of such policies for firms include Work Opportunity Tax Credit for groups such as veterans and long-term unemployed individuals, and Empowerment Zone Employment Credit for individuals who live and work in Empowerment Zones, which are economically distressed areas identified by the government. For schools, the guidance of Magnet Schools Assistance Program (MSAP) emphasized the importance of the diversity and equity for awarding these subsidies *"These competitive preference priorities address a local educational agencies' (LEA) need for MSAP funding, the evidence base undergirding the LEA's program design for new or significantly revitalized magnet schools, the means of student selection for admission including use of lotteries and other non-academic means, and attention to socioeconomic factors in promoting diversity."* See Applications for New Awards; Magnet Schools Assistance Program, Office of Elementary and Secondary Education, Department of Education for the full guideline.

and quality, and studies on affirmative action rely on additively separable preferences when modeling preference for the representation of particular groups. For instance, in Chan and Eyster (2003), the utility of the school is the sum of the average score of admitted students and the number of minority students. Ellison and Pathak (2021) include a term that penalizes the utility as the distribution of characteristics deviates from an ideal point. Celebi and Flynn (2023) consider preferences additively separable in the average score and identity composition of admitted agents.[28] My characterization sheds light on the implications of additive separability on the resulting choice rules.

Third, many choice rules that are used in practice can be mapped to this setting and rationalized by additively separable preferences. This approach enables us to quantify the impact of different policies, as well as the effects of varying the strength of these policies within a given class. Additionally, for reserve rules, separability induces a specific processing order, ensuring that all open slots are filled only after the reserve slots have been processed.

In what follows, I will show that three well-known properties adapted to my setting characterize the choice rules that can be rationalized by the following utility function

$$U(I) = \sum_{i \in I} u(s(i)) + \sum_{\theta \in \Theta} h_\theta(N_\theta(I)) \tag{2}$$

where $u(s(i))$ is the benefit the institution receives from allocating the resource to an individual with score $s(i)$, while $h_\theta(N_\theta(I))$ is the benefit the institution receives from choosing $N_\theta(I)$ individuals with each identity θ. The utility from the score domain is obtained by summing the $u(s(i))$ over all chosen individuals, the diversity utility is obtained by summing $h_\theta(N_\theta(I))$ over all identities and the utility of the institution is the sum of these two terms.

5.1. *Preliminaries*

For this section, I assume that if $I' \in C(I)$ and $I'' \in C(I)$, then I' is equivalent to I'', that is, $\tau(I) = \tau(I')$. This means that although C is still a choice correspondence since there can be many individuals with the same type, it is actually a choice function if we restrict attention to equivalence classes $\mathcal{T}$.

[28]Other recent papers that consider additively separable preferences for affirmative action include Dessein et al. (2023), who studies an extension an extension where a college's utility function provides an additive bonus to students who belong to a particular group, and Passaro et al. (2023), who explores an extension where a firm incurs a constant per-worker disutility for hiring minority workers.

Given C, I construct the following binary relation $>_C$:

$$I \cup \{j\} \in C(I \cup \{j,k\}) \text{ and } I \cup \{k\} \notin C(I \cup \{j,k\}) \implies (s(j), \theta(j), N_{\theta(j)}(I \cup \{j\})) >_C (s(k), \theta(k), N_{\theta(k)}(I \cup \{k\}))$$

$>_C$ is the revealed preference relation over pairs of individuals induced by C. $(s, \theta, n) >_C (s', \theta', n')$ indicates that a θ individual with score s is chosen with $n-1$ other θ individuals instead of a θ' individual with score s' with $n'-1$ other θ' individuals.

Example 6. Suppose that $C(\{i,j,k\}) = \{i,j\}$, where all three individuals have different identities. Since i and j are chosen instead of k, we have

$$(s(i), \theta(i), 1) >_C (s(k), \theta(k), 1) \text{ and } (s(j), \theta(j), 1) >_C (s(k), \theta(k), 1)$$

△

Let $Q = \{1, \ldots, q\}$ and $D = \Theta \times Q$ denote the set of all (θ, n) with generic element $d \in D$.

Definition 11. *Given a binary relation $>$, a collection*

$$\begin{gathered}(s_1, d_1) > (s_1', d_1') \\ (s_2, d_2) > (s_2', d_2') \\ \vdots \\ (s_m, d_m) > (s_m', d_m')\end{gathered}$$

is a cycle if $(s_1', \ldots, s_m')$ is a permutation of $(s_1, \ldots, s_m)$ and $(d_1', \ldots, d_m')$ is a permutation of $(d_1, \ldots, d_m)$.

This definition is due to Tversky (1964) (see also Scott (1964); Adams (1965)) and is used to characterize preferences that admit an additively separable utility representation. The existence of a cycle under $>_C$ means that the evaluation of the diversity and quality domains are connected, since $\{(s_i, d_i)\}_{i \leq m}$ and $\{(s_i', d_i')\}_{i \leq m}$ are formed from the same scores and diversity levels, but $\{(s_i, d_i)\}_{i \leq m}$ are revealed strictly preferred to $\{(s_i', d_i')\}_{i \leq m}$ for all i.

Definition 12. *C satisfies acyclicity if there are no cycles under $>_C$.*

Acyclicity of $>_C$ rules out any connection between the diversity and score domains.

Definition 13. *C satisfies within-group responsiveness if for all i and j with $\theta(i) = \theta(j)$ and $s(i) > s(j)$, there does not exist $\hat{I} \subseteq I$ such that $\hat{I} \cap \{i,j\} = \emptyset$, $\hat{I} \cup \{j\} \in C(I)$ and $\hat{I} \cup \{i\} \notin C(I)$.*

Within-group responsiveness is the restriction of responsiveness to individuals with the same identity, as individuals with exactly same identity are comparable in isolation, while comparison between individuals with different identities may depend on the identities of other chosen individuals. It ensures that the choice rule is responsive to scores in the sense that higher scoring individuals are chosen before lower scoring individuals.

5.2. Analysis

The following result characterizes the class of choice rules that can be induced by a utility function that is separable in the diversity and quality domains.

Theorem 3. *C satisfies substitutes, within-group responsiveness and acyclity if and only if there exist increasing u and concave $\{h_\theta\}_{\theta\in\Theta}$ such that*

$$U(I) = \sum_{i\in I} u(s(i)) + \sum_{\theta\in\Theta} h_\theta(N_\theta(I)) \tag{3}$$

where U rationalizes C.

Theorem 3 demonstrates that a choice rule can be represented by a utility function given in Equation 3 if and only if it satisfies the substitutes condition (Roth, 1984), within-group responsiveness (Roth, 1985) and acyclicity (Tversky, 1964). In addition to additive separability, U incorporates a preference for diversity through the concavity of the functions h_θ. This concavity implies that the marginal benefit of selecting an individual with a given identity decreases (weakly) as the number of such individuals increases, reflecting a preference for avoiding an over representation of any single identity.

To prove Theorem 3, I first show that the incomplete binary relation $>_C$ can be represented by an additively separable utility function of the form $u(s) + h(\theta, n)$, where $u(s)$ is an increasing function of the score s, and $h(\theta, n)$ captures the benefit of adding the n'th individual with identity θ. This yields a representation, u and $\{h(\theta, n)\}_{\theta\in\Theta}$, that rationalizes C for decisions between pairs of individuals, but not for decisions over sets of individuals. Then I demonstrate that, when the substitutes condition is satisfied, we can construct concave functions h_θ from $h(\theta, n)$, where $h_\theta(n)$ represents the benefit of allocating the resource to n individuals with identity θ such that the utility function obtained by summing the score utilities across all individuals and diversity utility across all identities represents the preferences over sets of individuals, yielding representation in the theorem.

Many commonly used choice rules can be mapped to this framework, facilitating an understanding of the preferences underlying these rules.

Quota Policies. A quota policy restricts admission of individuals of each type θ by some $k_\theta \geq 0$ (Kojima, 2012). Given $\{k_\theta\}_{\theta\in\Theta}$, a quota policy can be rationalized by any $u(s) = s$ and $\{h_\theta\}_{\theta\in\Theta}$ given by

$$h_{\boldsymbol{\theta}}(N_\theta(I)) = \begin{cases} 0 & \text{if } N_\theta(I) \leq k_\theta \\ -q\bar{u} & \text{if } N_\theta(I) > k_\theta \end{cases}$$

where $\bar{u} > \hat{s} \equiv \max_{s,s'\in\mathcal{S}} s - s'$. Thus, a quota policy is rationalized by preferences where failing to meet the quota, which costs $q\bar{u}$, can never be remedied by improvements in score domain, which are capped by $q\hat{s}$.

Reserve Policies. Reserve policies (see Hafalir et al. (2013) and Dur et al. (2020)) is a special case of generalized reserves where each slot is assigned to at most a single identity.[29] $r_\theta \geq 0$ denotes the number of positions reserved for individuals with each identity θ. When the number of individuals with each identity is higher than the number of reserve positions for that identity, and open positions are processed after all reserve positions, reserve policies can be rationalized by $u(s) = s$ and $\{h_\theta\}_{\theta\in\Theta}$ given by

$$h_\theta(N_\theta(I)) = \begin{cases} N_\theta(I)\bar{u} & \text{if } N_\theta(I) \leq k_\theta \\ k_\theta\bar{u} & \text{if } N_\theta(I) > k_\theta \end{cases}$$

for $\bar{u} > \hat{s}$. This indicates that the diversity utility is increasing and more important than any gains in the score dimension until the reserve is met and is constant after the reserve requirements are satisfied. However, as the following example shows, if open positions are processed before reserves, the choice rule may fail acyclicity and cannot be represented by an additively separable utility function.

Example 7. There are two groups in one dimension, $\Theta = \{a, b\}$ and $q = 3$. There is one reserve position for each of the groups. The processing order is open positions, group a reserve and group b reserve. Let $I = \{a_3, b_2, a_1, b_0\}$ and $I' = \{a_2, b_3, a_1, b_0\}$, where letters denote groups and subscripts denote scores.

Applicants	Admission		
	Open	Reserve a	Reserve b
a_3	✓		
b_2			✓
a_1		✓	
b_0			

Applicants	Admission		
	Open	Reserve a	Reserve b
a_2		✓	
b_3	✓		
a_1			
b_0			✓

[29] This corresponds to $h(\sigma)$ being either $\emptyset$ or a singleton.

Under I, a_3 receives the open position, allowing a_1 to receive the position reserved for group a. Now consider $I' = \{a_2, b_3, a_1, b_0\}$. Under I', b_3 receives the open position, allowing b_0 to receive the position reserved for group b. Therefore, $C(I) = \{a_3, b_2, a_1\}$, implying $(1, a, 2) >_C (1, b, 2)$ and $C(I') = \{a_2, b_3, b_0\}$, implying $(1, b, 2) >_C (1, a, 2)$. This violates acyclicty and shows that this choice rule cannot be rationalized by a separable utility function. △

The intuition behind Example 7 can be explained as follows: In both cases, exactly one individual from each group is chosen before a_1 and b_0, so their contributions to the diversity domain should be identical. Additionally, since both individuals have the same score, their contributions to the score domain should also be equivalent. Nevertheless, a_1 is selected in the first case, while b_0 is selected in the second. This discrepancy indicates that the evaluation of these domains must be interconnected; under additive separability, the same individuals should have been chosen in both cases. The following proposition demonstrates that any choice rule that processes an open position before a reserve position suffers from the same issue.

Proposition 7. *C satisfies acyclicity if and only if it processes open positions after reserve positions.*

This result contributes to the discussion on the processing order of the positions, which is important for the distribution of positions among individuals from different groups. For example, Dur et al. (2020) shows that processing reserve positions of a group earlier is advantageous for that group and can serve as an additional lever in affirmative action programs. Arnosti et al. (2024) defines *outcome-based* affirmative action rules, which fix a set of feasible allocations and maximize scores conditional on selecting a feasible allocation. They show that reserve rules are outcome-based if and only if open slots are processed at the end. Proposition 7 shows that if an institution adopts a processing order that processes open positions before reserve positions, this indicates a fundamental difference in preferences compared to the case where open positions are processed last.

Ideal Point Policies. An ideal point is a distribution of characteristics $z^* = \{z^*_\theta\}_{\theta\in\Theta}$, where z^*_θ is the most preferred number of chosen individuals with identity θ. A choice rule is generated by ideal point z^* if it first chooses a distribution of students that is as close to z^* as possible in Euclidean distance, and then admits the highest scoring students of each identity (Echenique and Yenmez, 2015).[30] An ideal point policy can be rationalized by $u(s) = s$ and $\{h_\theta\}_{\theta\in\Theta}$ given by

$$h_\theta(N_\theta(I)) = \bar{u}(z_\theta - N_\theta(I))^2$$

[30] z^* must satisfy the following: z^*_θ is a positive an integer for all θ and $\sum_{\theta\in\Theta} \leq q$.

for $\bar{u} > \hat{s}$. Similarly to quota policies, the preferences that rationalize ideal point priorities first make sure that any step away from the ideal point cannot be compensated by gains in the score domain, and then there is a convex penalty for moving away from the ideal point, due to the Euclidean distance.

Priority Policies. The policies we have studied so far put the composition of characteristics first in the sense that they maximize scores only conditional on achieving certain distributional objectives and do not allow for flexible trade-off between diversity and score domains. An exception to this is the priority policies, which are defined by a vector of *bonus points* $\{b_\theta\}_{\theta \in \Theta}$. A priority policy transforms the scores by increasing the score of each individual with identity θ by b_θ, and chooses the individuals with the highest transformed scores (Celebi and Flynn, 2023). A priority policy with $\{b_\theta\}_{\theta \in \Theta}$ can be rationalized by $u(s) = s$ and $\{h_\theta\}_{\theta \in \Theta}$ given by

$$h_\theta(N_\theta(I)) = b_\theta N_\theta(I)$$

Priority policies are used in many real world markets. For example, in the centralized high school admission system of Taiwan, schools prefer higher scoring students, but deduct points from each student's scores based on the school's ranking in the student's preference list (Dur, Pathak, Song, and Sönmez, 2022).[31] In this setting, $\Theta = \{1, \ldots, n\}$ where n is the length of the preference lists and $\{b_\theta\}_\Theta$ determines how the school views the trade-off between admitting students with higher scores and students who prefer the school more.

6. Conclusion

This paper contributes to the study of affirmative action and diversity concerns in market design. On the theoretical side, I adapt revealed preference analysis to market design by interpreting choice rules as an agent's choices from various feasible sets. This allows me to characterize choice rules that can be rationalized (with preferences that satisfy certain desirable properties) and study how certain features of preferences affect the properties of choice rules.

Applying my results, I study two important settings, the affirmative action mechanisms in Indian government jobs and Brazilian college admissions. Although the earlier literature focused on similar shortcomings of these mechanisms, I show that the mechanism used in India is not rationalizable, and thus couldn't be designed by an authority with well defined preferences over outcomes. The mechanism used in Brazil is rationalizable, but not with

[31] For example, $b_\theta = 2(\theta - 1)$ deducts two points from student's exam score for school ranked above.

preferences that are increasing in eligibility of affirmative action, a desirable criterion for rules featuring affirmative action. Finally, I show that the multidimensional privileges choice rule proposed by Aygun and Bó (2021) and subsequently implemented by the Brazilian government is the unique mechanism that can be rationalizable with preferences that are increasing in affirmative action eligibility, providing some further foundation for its adoption.

Next, I study how different ways of evaluating diversity affect the properties of choice rules. When identities are multidimensional and overlapping, institutions can evaluate diversity in multiple ways. They can either evaluate each dimension separately, focusing on the marginal distribution of identities, or evaluate them jointly, focusing on their cross sectional distribution. When evaluation of diversity does not consider intersectionality, the choice rules induced by those preferences create complementarities between individuals who have belong to different groups across dimensions, and therefore does not satisfy the substitutes condition, which is necessary for the use of stable matching mechanisms. Moreover, I show that non-intersectional views of diversity allows incumbent groups to increase their representation above their population shares without compromising the perception of (marginal) diversity, formalizing observations made in the literature that studies diversity in legislative assemblies. These results show that intersectionality emerges not only as an important consideration from an equity standpoint but is also necessary to satisfy the substitutes condition, contributing to the literature that studies substitutes conditions in market design and the literature that studies representation of different identities in national politics.

Finally, I characterize the choice rules that can be rationalized by a class of preferences that treat scores and identities in an additively separable way. Perhaps surprisingly, three well-known axioms, substitutes, (within-group) responsiveness and acyclicity, fully characterize these preferences. I then identify the preferences that induce some well-known choice rules such as quotas and reserves within this framework, and show that separability induces a unique processing order for the reserve rules. My framework provides a systematic way of understanding and evaluating the trade-offs and relationship between diversity preferences and the affirmative action policies used in practice.

References

Abdulkadiroğlu, A., 2005. College admissions with affirmative action. International Journal of Game Theory 33, 535–549.

Abdulkadiroğlu, A., Sönmez, T., 2003. School choice: A mechanism design approach. American economic review 93, 729–747.

Adams, E. W., 1965. Elements of a theory of inexact measurement. Philosophy of science 32, 205–228.

Akbarpour, M., Budish, E. B., Dworczak, P., Kominers, S. D., 2021. An economic framework for vaccine prioritization. Available at SSRN 3846931 .

Alva, S., 2018. Warp and combinatorial choice. Journal of Economic Theory 173, 320–333.

Arnosti, N., Bonet, C., Sethuraman, J., 2024. Explainable Affirmative Action. Working Paper .

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

Aygün, O., Sönmez, T., 2013. Matching with contracts: Comment. American Economic Review 103, 2050–51.

Aygün, O., Turhan, B., 2017. Large-scale affirmative action in school choice: Admissions to iits in india. American Economic Review 107, 210–13.

Caradonna, P., 2020. How strong is the weak axiom. Unpublished paper, Department of Economics, Georgetown University.[856] .

Carvalho, J.-P., Pradelski, B. S., Williams, C., 2024. Intersectionality: Affirmative action with multidimensional identities. Management Science .

Celebi, O., Flynn, J., 2023. Adaptive priority mechanisms. arXiv preprint arXiv:2309.15997 .

Çelebi, O., Flynn, J. P., 2022. Priority design in centralized matching markets. The Review of Economic Studies 89, 1245–1277.

Celis, K., Erzeel, S., 2017. The complementarity advantage: Parties, representativeness and newcomers' access to power. Parliamentary Affairs 70, 43–61.

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

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

Crenshaw, K., 2013. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In: *Feminist Legal Theories*, Routledge, pp. 23–51.

Dessein, W., Frankel, A., Kartik, N., 2023. Test-optional admissions. arXiv preprint arXiv:2304.07551 .

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

Dur, U., Pathak, P. A., Song, F., Sönmez, T., 2022. Deduction dilemmas: The taiwan assignment mechanism. American Economic Journal: Microeconomics 14, 164–185.

Dur, U., Pathak, P. A., Sönmez, T., 2020. Explicit vs. statistical targeting in affirmative action: Theory and evidence from chicago's exam schools. Journal of Economic Theory 187.

Echenique, F., Yenmez, M. B., 2015. How to control controlled school choice. American Economic Review 105, 2679–94.

Ehlers, L., Hafalir, I. E., Yenmez, M. B., Yildirim, M. A., 2014. School choice with controlled choice constraints: Hard bounds versus soft bounds. Journal of Economic theory 153, 648–683.

Ellison, G., Pathak, P. A., 2021. The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from chicago's exam schools. American Economic Review 111, 943–75.

Fishburn, P. C., 1970. Utility theory for decision making. Tech. rep., Research analysis corp McLean VA.

Goto, M., Kojima, F., Kurata, R., Tamura, A., Yokoo, M., 2017. Designing matching mechanisms under general distributional constraints. American Economic Journal: Microeconomics 9, 226–62.

Grigoryan, A., 2021. Effective, fair and equitable pandemic rationing. Available at SSRN 3646539 .

Hafalir, I. E., Yenmez, M. B., Yildirim, M. A., 2013. Effective affirmative action in school choice. Theoretical Economics 8, 325–363.

Hatfield, J. W., Kojima, F., 2010. Substitutes and stability for matching with contracts. Journal of Economic theory 145, 1704–1723.

Hatfield, J. W., Milgrom, P. R., 2005. Matching with contracts. American Economic Review 95, 913–935.

Hughes, M. M., 2011. Intersectionality, quotas, and minority women's political representation worldwide. American Political Science Review 105, 604–620.

Hughes, M. M., 2018. Ethnic quotas in electoral politics. Gender Parity and Multicultural Feminism: Towards a New Synthesis p. 97.

Kamada, Y., Kojima, F., 2017. Stability concepts in matching under distributional constraints. Journal of Economic theory 168, 107–142.

Kamada, Y., Kojima, F., 2018. Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition. Theoretical Economics 13, 761–793.

Kelso, A. S., Crawford, V. P., 1982. Job matching, coalition formation, and gross substitutes. Econometrica: Journal of the Econometric Society pp. 1483–1504.

Kojima, F., 2012. School choice: Impossibilities for affirmative action. Games and Economic Behavior 75, 685–693.

Kojima, F., Sun, N., Yu, N. N., 2020a. Job matching under constraints. American Economic Review 110, 2935–47.

Kojima, F., Sun, N., Yu, N. N., 2020b. Job matching with subsidy and taxation. Available at SSRN 3624343 .

Kominers, S. D., Sönmez, T., 2016. Matching with slot-specific priorities: Theory. Theoretical Economics 11, 683–710.

Kurata, R., Hamada, N., Iwasaki, A., Yokoo, M., 2017. Controlled school choice with soft bounds and overlapping types. Journal of Artificial Intelligence Research 58, 153–184.

Murota, K., 1998. Discrete convex analysis. Mathematical Programming 83, 313–371.

Murota, K., et al., 2016. Discrete convex analysis: A tool for economics and game theory. Journal of Mechanism and Institution Design 1, 151–273.

Murray, R., 2016. The political representation of ethnic minority women in france. Parliamentary Affairs 69, 586–602.

Nishimura, H., Ok, E. A., Quah, J. K.-H., 2016. A Comprehensive Approach to Revealed Preference Theory. Working Papers 201614, University of California at Riverside, Department of Economics.

Passaro, D. G., Kojima, F., Pakzad-Hurson, B., 2023. Equal pay for similar work. arXiv preprint arXiv:2306.17111 .

Pathak, P. A., Sönmez, T., Ünver, M. U., Yenmez, M. B., 2021. Fair allocation of vaccines, ventilators and antiviral treatments: leaving no ethical value behind in health care rationing. In: *Working Paper*, pp. 785–786.

Richter, M. K., 1966. Revealed preference theory. Econometrica: Journal of the Econometric Society pp. 635–645.

Roth, A. E., 1984. Stability and polarization of interests in job matching. Econometrica: Journal of the Econometric Society pp. 47–57.

Roth, A. E., 1985. The college admissions problem is not equivalent to the marriage problem. Journal of economic Theory 36, 277–288.

Scott, D., 1964. Measurement structures and linear inequalities. Journal of mathematical psychology 1, 233–247.

Sönmez, T., Yenmez, M. B., 2022. Affirmative action in india via vertical, horizontal, and overlapping reservations. Econometrica 90, 1143–1176.

Tversky, A., 1964. Finite additive structures. Michigan Mathematical Psychology Program .

Yokote, K., Hafalir, I. E., Kojima, F., Yenmez, M. B., 2024. Rationalizing path-independent choice rules. arXiv preprint arXiv:2303.00892 .

Appendices

A. Proofs

A.1. Proof of Proposition 1

First, I will show that if for all $I \in A$ and $I' \in A$, all q-element subsets of $I \cup I'$ must also be in A, then $A \in \mathcal{A}^*$. Take an arbitrary A that satisfies this condition, which I will refer as condition $\mathcal{C}$. Define $I^* = \bigcup I \in A$. Let $\hat{A}$ denote all q-or fewer element subsets of I^*, that is, $\cup_{k=1,\dots,q} 2^{I^*}_k$. We observe the choice rule to make a choice from the feasible set $\hat{A}$ when the applicant set is I^*, thus $\hat{A} \in \mathcal{A}$. I will now show that $\hat{A} = A$.

First, as A satisfies condition $\mathcal{C}$ and $I^* = \bigcup I \in A$, we have $\hat{A} \subseteq A$. To see why, take any $I \in \hat{A}$. Enumerate the elements in I as $I = \{i_1, \dots, i_q\}$. As $I^* = \bigcup I \in A$, for any $i_j \in I$, there exists $I_j \in A$ such that $i_j \in I_j$. $K_2 = \{i_1, i_2\}$, where $K_2 \in A$ as it has (weakly) fewer than q elements, $K_2 \subset I_1 \cup I_2$, $I_1 \subseteq A$, $I_2 \subseteq A$ and A satisfies condition $\mathcal{C}$. Define the following sets inductively: $K_n = \{i_1, i_2, \dots, i_{n-1}, i_n\}$ where $K_n \in A$ as it has (weakly) fewer than q elements, $K_n \subset K_{n-1} \cup I_n$, $K_{n-1} \in A$, $I_n \in A$ and A satisfies condition $\mathcal{C}$. Then $K_q = I \in A$, proving the result. To show $A \subseteq \hat{A}$, suppose that there exists $J \in A$ with $J \notin \hat{A}$. As the institution can choose at most q elements, $|J| \leq q$. As $J \in A$, $J \subseteq I^*$. As $|J| \leq q$, $J \in \hat{A}$ which proves $A \subseteq \hat{A}$, and therefore $\hat{A} = A$.

Second, take an arbitrary A that fails condition $\mathcal{C}$. Then there exists I and I' with $I'' \subseteq I \cup I'$, $|I''| \geq q$ and $I'' \notin A$. Let $\hat{I}$ denote the set of applicants that induced A. Then $I \cup I' \subseteq \hat{I}$. Then the choice rule can also select I'' as $I'' \subseteq \hat{I}$ and $|I''| \geq q$. As $I'' \notin A$, this implies that $A \notin \mathcal{A}$.

A.2. Proof of Theorem 1

Define the relation $\geq_C$ as follows: $I \geq_C I'$ if $I \in C(\hat{I})$ and $I' \subset \hat{I}$. Let $\geq_C \cup \trianglerighteq$ denote the union of $\geq_C$ and $\trianglerighteq$ and tran$(\geq_C \cup \trianglerighteq)$ denote the transitive closure of this relation.

C satisfies $\trianglerighteq$-congruence if (i) I tran$(\geq_C \cup \trianglerighteq) I'$ and $I' \in C(\hat{I})$ imply $I \in C(\hat{I})$ for every $\hat{I}$ that contains I and (ii) I tran$(>_C \cup \trianglerighteq) I'$ imply that we do not have $I' \triangleright I$. The following lemma follows from the finiteness of $\mathcal{I}$.

Lemma 1. *C satisfies $\trianglerighteq$-congruence if and only if C doesn't admit a $\trianglerighteq$-cycle.*

Proof. Suppose that C admits $\trianglerighteq$-cycle $I_1, \dots, I_n$. Then for each $i \leq n-1$, either $I_i \geq_C I_{i+1}$ or $I_i \trianglerighteq I_{i+1}$. Thus, I_1 tran$(\geq_C \cup \trianglerighteq) I_n$. Moreover, as either $I_n \triangleright I_1$ or $I_n \in C(\hat{I})$, $I_n \notin C(\hat{I})$ and $I_1 \subset \hat{I}$ for some $\hat{I}$, $\trianglerighteq$-congruence fails.

Conversely, suppose that C does not satisfy $\trianglerighteq$-congruence and let $I_1, \dots, I_n$ denote the sets that cause the violation. Then I_1 tran$(\geq_C \cup \trianglerighteq) I_n$. Then for each $i < n$, either (i) there exists an $\hat{I}_i$ such that $I_i \in C(\hat{I}_i)$ and $I_{i+1} \subset \hat{I}_i$ or (ii) $I_i \trianglerighteq I_{i+1}$. Moreover, we also have either (i) there exists $\hat{I}_n$ such that $I_n \in C(\hat{I}_n)$, $I_1 \subset \hat{I}_n$ and $I_1 \notin C(\hat{I}_n)$ or (ii) $I_n \trianglerighteq I_1$, which completes the proof. ☐

The result then follows from Theorem 7 in Nishimura et al. (2016), who shows that a choice rule C is rationalizable by a preference relation $\succeq$ that extends $\trianglerighteq$ if and only if it satisfies $\trianglerighteq$-congruence.

A.3. Proof of Corollary 1

Define $\trianglerighteq$ as the empty relation. The result follows then from Theorem 1.

A.4. Proof of Proposition 2

Fix r,r_w,o and o_w such that $r_w > 0$ and $o_w > 0$. I will now define 4 sets of individuals of each group with increasing integer scores. $I_{wr} = \{(wr, 1), (wr, 2), \ldots, (wr, r_w)\}$ are r_w reserve eligible women. If $r - r_w = 0$ then $I_{mr} = \emptyset$, while if $r - r_w > 0$ then $I_{mr} = \{(mr, r_w + 1), (mr, r_w + 2), \ldots, (wr, r)\}$ are $r - r_w$ reserve eligible men. $I_{wg} = \{(wg, r + 1), (wr, r + 2), \ldots, (wr, r + o_w)\}$ are o_w general category women. If $o - o_w = 0$ then $I_{mg} = \emptyset$, while if $o - p_w > 0$ then $I_{mg} = \{(mg, r + o_w + 4), \ldots, (mg, r + o + 3)\}$ are $o - o_w$ general category men. Let $\hat{I} = I_{wr} \cup I_{mr} \cup I_{wg} \cup I_{mg}$. I will define two more individuals. $i_w = (wr, r + o_w + 2)$, $i_m = (mg, r + o_w + 1)$ and $i'_m = (mg, r + o_w + 3)$.

First, consider $C(I \cup \{i_w, i_m\})$. Note that $\mathcal{M} = \{i_w\}$ The first o_w slots are assigned to $I_{wg} \setminus \{(wg, r + 1)\} \cup \{i_w\}$. Next $o - o_w$ slots are assigned to I_{mg}. The next r_w slots are assigned to I_{wr} and the remaining $r - r_w$ slots are assigned to I_{mr}. This means that $I \cup \{i_w\} \setminus \{(wg, r + 1)\}$ is chosen when $I \cup \{i_w\} \setminus \{(wr, 1)\}$ was available.

Now consider $C(I \cup \{i_w, i'_m\})$, Note that $\mathcal{M} = \emptyset$. Then the first o_w slots are assigned to I_{wg}. Next $o - o_w$ slots are assigned to I_{mg}. The next r_w slots are assigned to $I_{wr} \cup \{i_w\} \setminus \{(wr, 1)\}$ and the remaining $r - r_w$ slots are assigned to I_{mr}. This means that $I \cup \{i_w\} \setminus \{(wr, 1)\}$ when $I \cup \{i_w\} \setminus \{(wg, r + 1)\}$ was available and creates the choice cycle.

A.5. Proof of Proposition 3

For a contradiction, suppose that there is a choice cycle $I_1, \ldots, I_n$. Note that I_n has at least q elements as I_1 was not chosen at the last step of the cycle. Thus, all sets in the choice cycle have at least q elements, as I_{n-1} was chosen when I_n was available, I_{n-2} was chosen when I_{n-1} was available, and so on. Take any two consecutive sets in the choice cycle, I_k and I_{k+1}. Without loss of generality, rename individuals in I_l as i^l_j if i^l_j was assigned to slot σ_j at the cycle step I_l was chosen.

Lemma 2. *For all j, either $i^k_j = i^{k+1}_j$ or $i^k_j >_{\sigma_l} i^{k+1}_j$*

Proof. The proof is by induction. First, note that since when I_k was chosen, for σ_1 all individuals in $I_k \cup I_{k+1}$ were considered, and i^k_1 was assigned to slot σ_1 as the individual with the highest priority at that slot. Therefore, $i^k_1 >_{\sigma_l} j$ for all $j \in I_k \cup I_k$ such that $j \neq i^k_1$, implying either $i^k_1 = i^{k+1}_1$, or $i^k_1 >_{\sigma_1} i^{k+1}_1$.

Now take any slot m, and suppose that the induction hypothesis holds, in other words, i_j^k has a weakly higher priority at slot σ_j than i_j^{k+1} for all $j < m$. If $i_m^k = i_m^{k+1}$, then we are done. Otherwise, I will first show that i_m^{k+1} is not assigned to any σ_l with $l < m$ at the cycle step I_k was chosen. Suppose for a contradiction, that is the case for slot l. Then by the induction hypothesis, $i_m^{k+1} = i_l^k \neq i_l^{k+1}$ and therefore $i_m^{k+1} >_{\sigma_l} i_l^{k+1}$, which contradicts that i_l is chosen for the slot σ_l at the cycle step I_{k+1} was chosen. Therefore, i_m^{k+1} was available when i_m^k was chosen, which implies that $i_m^k >_{\sigma_m} i_m^{k+1}$. □

Repeatedly applying Lemma 2, starting with I_1 and iterating until I_n, we obtain that for all slots j, either $i_j^1 >_{\sigma_j} i_j^n$ or $i_j^1 = i_j^n$. However, as $I_1 \neq I_n$, there exists i_j^n such that $i_j^n \neq i_j^1$. Moreover, as I_n was selected when I_1 was available, it must be that $i_j^n >_{\sigma_l} i_j^1$, which is a contradiction.

A.6. Proof of Proposition 4

In Brazil Reserves, any slot reserved for (R_A, I_N) and $((R_N, I_A))$ satisfies the following criterion: there exists $\hat{\theta}$ dominates θ', $\theta' \in h(\sigma)$ and $\hat{\theta} \notin h(\sigma)$. Let σ_k denote the final slot that satisfies this criterion. Consider the following set of individuals $I = \{i_1, \ldots, i_q\}$ such that

- For each $l < k$, $\theta(i_l) \in h(\sigma_l)$.
- For each $l \in \{k, \ldots, q\}$, $\theta_j(i_l) = \hat{\theta}$ for all j.
- For each $l \in \{1, \ldots, q\}$, $s(i_l) = q - l$.

Moreover, define i' where $\theta(i') = \theta'$ and $s(i') = 1$, and $\hat{i}$ where $\theta(\hat{i}) = \theta'$ and $s(\hat{i}) = 0$. First, note that $I \trianglerighteq_A I \setminus \{i_q\} \cup \{i'\}$. This is the first step of they cycle.

Second, $C(I \setminus \{i_q\} \cup \{i', \hat{i}\}) = I \setminus \{i_q\} \cup \{i'\}$. This is because, first $k - 1$ slots are assigned to individuals $\{i_1, \ldots, i_{k-1}\}$, σ_k is assigned to i' as the only remaining individual with an identity in $h(\sigma_k)$ and the remaining slots, which are reserved for $\hat{\theta}$ whenever they are reserved for θ', are assigned to individuals in $\{i_k, \ldots, i_{q-1}\}$ as the highest scoring remaining individuals. Thus, $I \setminus \{i_q\} \cup \{i'\}$ is chosen when $I \setminus \{i_q\} \cup \{\hat{i}\}$ is available. This is the second step in the cycle.

Third, $C(I \setminus \{i_q\} \cup \{\hat{i}\}) = I \setminus \{i_q\} \cup \{\hat{i}\}$. This is because, first $k - 1$ slots are assigned to individuals $\{i_1, \ldots, i_{k-1}\}$, σ_k is assigned to $\hat{i}$ as the only remaining individual with an identity in $h(\sigma_k)$ and the remaining slots are assigned to individuals in $\{i_k, \ldots, i_{q-1}\}$ as the highest scoring remaining individuals with the same identity. Thus, $I \setminus \{i_q\} \cup \{\hat{i}\}$ is chosen when I is available. This is the third step in the cycle, and completes the cycle and proves the result.

A.7. Proof of Proposition 5

I first show that the multidimensional privileges choice rule can be rationalized by preferences that extend $\trianglerighteq_A$. For a contradiction, suppose this is not the case. Then there is a $\trianglerighteq_A$-cycle $I_1, \ldots, I_n$. Moreover, Define $I_1 \equiv I_{n+1}$. As $\triangleright_A$ is defined to be an empty relation, at least one step of the cycle includes a choice step (that is, I_k was chosen when I_{k+1} is available). Without loss of generality, let the first step I_1, I_2 denote one of those steps.

Now suppose that, I_k and I_{k+1} are two consecutive sets in the cycle. Without loss of generality, rename individuals in I_l as i^l_j if i^l_j was assigned to slot σ_j when the choice rule is used on I_l only. As multidimensional privileges choice rule is a slot specific rule, we will consider the induced priorities at each slot. I use $\sigma_j(i)$ to denote the ranking of individual i at slot j.

Lemma 3. *Suppose that k'th step corresponds to a choice step. Let σ_j denote the first (lowest indice) slot such that $\sigma_j(i^k_j) \neq \sigma_j(i^{k+1}_j)$. Such a slot exists and $i^k_j >_{\sigma_j} i^{k+1}_j$.*

Proof. I first show the existence of such a slot. Suppose that such a slot does not exist and take $i \in I_k \setminus I_{k+1}$ and σ_i denote the slot i was assigned. As $i \notin I_{k+1}$, there does not exist $j \in I_{k+1}$ with $\sigma_i(i) \neq \sigma_i(j)$, as there does not exist any individual at I_{k+1} with the same score as i.

Moreover, $\sigma_j(i^k_j) \neq \sigma_j(i^{k+1}_j)$ implies that $t(i^k_j) \neq t(i^{k+1}_j)$. As the k'th step in the cycle corresponds to choice, in other words, I_k was chosen when I_{k+1} was available, the result follows from the fact that i^k_j was chosen at slot σ_j while $i^{k+1}_j \neq i^k_j$ was available to choose. □

Lemma 4. *Suppose that k'th step corresponds to $\trianglerighteq_A$. Let σ_j denote the first slot such that $\sigma_j(i^k_j) \neq \sigma_j(i^{k+1}_j)$, if such a slot exists. Then $i^k_j >_{\sigma_j} i^{k+1}_j$.*

Proof. As all individuals were same until slot j, and $I^k \trianglerighteq_A I^{k+1}$, then there is an individual, say i^*, remaining at slot j in I^k who has the same score and is eligible in more dimensions than i^{k+1}_j. As i^k_j was chosen in σ_j when i^* was available, $i^k_j \geq_{\sigma_l} i^{k+1}_j$. As $\sigma_j(i^k_j) \neq \sigma_j(i^{k+1}_j)$, $i^k_j >_{\sigma_j} i^{k+1}_j$ □

Let m denote the first slot such that $i^n_m \neq i^{n+1}_m$. Lemmas 3 and 4 imply that $i^n_m \geq_{\sigma_m} i^1_m$. The following lemma creates a contradiction and proves the if part of the result.

Lemma 5. $i^1_m >_{\sigma_m} i^n_m$

Proof. I first introduce a piece of notation. Given $r \in \{1, \ldots, n\}$, let σ_{p_r} denote the first slot such that $\sigma_{p_r}(i^1_{p_r}) \neq \sigma_{p_r}(i^r_{p_r})$. As the first step in the cycle corresponds to choice, by Lemma 3, $i^1_{p_2} >_{\sigma_p} i^2_{p_2}$. Now suppose that for all $l < k+1$, $i^1_{p_l} >_{\sigma_{p_l}} i^l_{p_l}$.

Consider the first slot such that $\sigma_{p_j}(i_j^k) \neq \sigma_{p_j}(i_j^{k+1})$. If such a slot does not exist, then $i_{p_k}^1 >_{\sigma_{p_k}} i_{p_k}^k$, as I_k and I_{k+1} are identical in slot priorities. If such a slot exists, then by Lemmas 3 and 4, $i_j^k \geq_{\sigma_l} i_j^{k+1}$. We will consider two cases. If $j \geq p_l$, then as all slots with indices lower than p_l the types of individuals are same, $i_{p_k}^1 >_{\sigma_{p_k}} i_{p_k}^k$. If $j < p_l$, then $j = p_k$ and again, $i_{p_k}^1 >_{\sigma_{p_k}} i_{p_k}^k$. The result then follows by induction. □

To prove the converse, take any generalized reserve rule that is not in the multidimensional privileges choice rules class. Then there exists at least one slot σ such that $\hat{\theta}$ dominates θ', $\theta' \in h(\sigma)$ and $\hat{\theta} \notin h(\sigma)$. Let σ_k denote the final slot that satisfies this criterion. Consider the following set of individuals $I = \{i_1, \ldots, i_q\}$ such that

- For each $l < k$, $\theta(i_l) \in h(\sigma_l)$.
- For each $l \in \{k, \ldots, q\}$, $\theta_j(i_l) = \hat{\theta}$ for all j.
- For each $l \in \{1, \ldots, q\}$, $s(i_l) = q - l$.

Moreover, define i' where $\theta(i') = \theta'$ and $s(i') = 0$. Note that $C(I \cup \{i'\}) = I \setminus \{i_q\} \cup \{i'\}$, as first $k-1$ slots are assigned to individuals $\{i_1, \ldots, i_{k-1}\}$, σ_k is assigned to i' as the only remaining individual with an identity in $h(\sigma_k)$ and the remaining slots are assigned to individuals in $\{i_k, \ldots, i_{q-1}\}$ as the highest scoring individuals with the same identity. Therefore, $I \setminus \{i_q\} \cup \{i'\}$ was chosen when I was available.

Let i_q^1 denote an individual with identity θ' and score 1 (which is same as i_q). Note that $C(I \setminus \{i_q\} \cup \{i', i_q^1\}) = I \setminus \{i_q\} \cup i_q^1$, as i_q^1 is assigned to slot k and all slots after that are assigned according to score, according to which i' ranks last. Then we have the following relations

- $I \trianglerighteq_A I \setminus \{i_q\} \cup i_q^1$
- $I \setminus \{i_q\} \cup i_q^1$ is chosen when $I \setminus \{i_q\} \cup \{i'\}$ is available.
- $I \setminus \{i_q\} \cup \{i'\}$ was chosen when I and not chosen

These form a $\trianglerighteq_A$-cycle, proving the only if part of Proposition 5.

A.8. Proof of Theorem 2

Suppose that $\mathcal{J}$ includes q individuals from each $\theta \in \Theta$ and $\succeq$ doesn't consider intersectionality. I say that i is a (j, k) individual if $\theta_j(i) = k$.

Let D^* denote the set of all optimal marginal distributions. Formally, $d \in D^*$ if there exists I' such that $I' \in C(\mathcal{J})$ and $M(I') = d$. Let d_1 denote an element of D^* with the highest number of $(1,1)$ individuals. m_{11} denotes the number of $(1,1)$ individuals at d_1. D_1^* denotes the set of all optimal group distributions where the number of $(1,1)$ individuals is

m_{11}. Let d^*_{11} be a group distribution in D^*_1 with the highest number of $(2,1)$ individuals. m_{21} denotes the number of $(2,1)$ individuals at d^*_{11}.

A set of individuals I is *compatible with* marginal distributions d^* if there exists I' such that $M(I \cup I') = d^*$ and I' is a *complement* of I for d^*. Let $M_{ij}(I)$ ($M_{ij}(d)$) denote the number of group i individuals in dimension j in I (d).

Lemma 6. *If $M_{ij}(I) \leq M_{ij}(d)$ for all i and j, then I is compatible with d.*

Proof. If $M_{ij}(I) < M_{ij}(d)$ for some ij, then for each dimension i', there exists a group j' such that $M_{i'j'}(I) < M_{i'j'}(d)$. Let t denote an individual who belongs to group j' at each dimension i'. Then the set $I \cup \{t\}$ still satisfies $M_{ij}(I) \leq M_{ij}(d)$ and repeating this procedure yields a $\tilde{I}$ such that $M_{ij}(\tilde{I}) = M_{ij}(d)$ and I is compatible with d. □

Case 1: $m_{11} \leq m_{21}$.

Claim 1. *There exists $I_{11} = \{i^{11}_1, \ldots, i^{11}_{m_{11}}\}$ compatible with d^*_{11} where all $i \in I_{11}$ are $(1,1)$ and $(2,1)$ individuals.*

Proof. Since groups $(1,1)$ and $(2,1)$ have (weakly) more individuals at d^*_{11}, one can choose the groups of individuals in I_{11} in other dimensions to satisfy $M_{ij}(I_{11}) \leq M_{ij}(d^*_{11})$ for all i and j. Then the result follows from Lemma 6. □

Let I' denote a complement of I_{11} at d^*_{11}. Take $j \in I_{11}$ and $k \in I'$, where k isn't a $(1,1)$ or $(2,1)$ individual.[32] Define $\tilde{j}$ and $\tilde{k}$ as

$$\theta_1(\tilde{j}) = \theta_1(j), \theta_\ell(\tilde{i}) = \theta_\ell(k) \text{ for all } \ell \neq 1$$

$$\theta_1(\tilde{k}) = \theta_1(k), \theta_\ell(\tilde{k}) = \theta_\ell(j) \text{ for all } \ell \neq 1$$

Let $\tilde{I}_{11} = I_{11} \setminus \{j\} \cup \tilde{j}$ and $I'' = I' \setminus \{k\} \cup \tilde{k}$. Note that I'' is a complement of $\tilde{I}_{11}$ at d^*_{11}. The following claim holds by construction.

Claim 2. *I' and I'' doesn't have any $(1,1)$ individuals. Moreover, I'' has $m_{21} - m_{11} + 1$ group $(2,1)$ individuals.*

Let $\bar{I} = I_{11} \cup I' \cup \{\tilde{j}, \tilde{k}\}$. As $M(I_{11} \cup I') = d^*_{11}$, $I_{11} \cup I' \in C(\bar{I})$.

Lemma 7. *There doesn't exist an $I^* \in C(\bar{I} \setminus \{k\})$ such that $I_{11} \subset I^*$*

[32] This is possible since the preferences value diversity and all individuals in I_{11} are both $(1,1)$ and $(2,1)$ indviduals.

Proof. Suppose that there is such an I^*. Then it must be that, $\tilde{j} \notin I^*$, since otherwise there will be more than m_{11} $(1,1)$ individuals at I^*, which is a contradiction. However, this means that $I^* = I_{11} \cup I''$. But then I^* has $m_{21}+1$ $(2,1)$ individuals and m_{11} $(1,1)$ individuals, which contradicts the optimality of I^* as d^*_{11} is a group distribution in D^*_1 with the highest number of $(2,1)$ individuals and has m_{21} such individuals. Since $\tilde{I}_{11} \cup I''$ is available and optimal, this is a contradiction. □

The result then follows from the fact that I_{11} is chosen from $\bar{I}$, but not from $\bar{I} \setminus \{k\}$.

Case 2: $m_{11} > m_{21}$. Let $n = m_{11} - m_{21}$.

Claim 3. *There exists $I_{12} = \{i^{11}_1, \dots, i^{11}_{m_{21}}, i^1_1, \dots, i^1_n\}$ where the first m_{21} elements are $(1,1)$ and $(2,1)$ individuals, rest are $(1,1)$ individuals and I_{12} is compatible with d^*_{11}.*

Proof. Since groups $(1,1)$ and $(2,1)$ have (weakly) more individuals at d^*_{11}, one can choose the groups of individuals in I_{12} to satisfy $M_{ij}(I_{12}) \leq M_{ij}(d^*_{11})$ for all i and j. Then the result follows from Lemma 6. □

Let I' denote a complement of I_{12} at d^*_{11}. Take $j \in I_{12}$ and $k \in I'$, where k isn't a $(1,1)$ or $(2,1)$ individual.[33] Define $\tilde{j}$ and $\tilde{k}$ as

$$\theta_2(\tilde{j}) = \theta_2(j), \theta_\ell(\tilde{j}) = \theta_\ell(k) \text{ for all } \ell \neq 1$$

$$\theta_2(\tilde{k}) = \theta_2(k), \theta_\ell(\tilde{k}) = \theta_\ell(j) \text{ for all } \ell \neq 1$$

Let $\tilde{I}_{12} = I_{12} \setminus \{j\} \cup \tilde{j}$ and $I'' = I' \setminus \{k\} \cup \tilde{k}$. Note that I'' is a complement of $\tilde{I}_{12}$ at d^*_{11}. The following claim holds by construction.

Claim 4. *I' and I'' doesn't have any $(2,1)$ individuals. Moreover, I'' has 1 group $(1,1)$ individual.*

Let $\bar{I} = I_{12} \cup I' \cup \{\tilde{j}, \tilde{k}\}$. First, note that $I_{12} \cup I' \in C(\bar{I})$ and $\tilde{I}_{12} \cup I'' \in C(\bar{I})$, since $M(I_{12} \cup I') = M(\tilde{I}_{12} \cup I'') = d^*_{11}$.

Lemma 8. *There doesn't exist an $I^* \in C(\bar{I} \setminus \{k\})$ such that $I_{12} \subset I^*$.*

Proof. Since $\tilde{I}_{12} \cup I''$ is available and optimal, I^* must also be optimal. For a contradiction, suppose that such an I^* exists. Then it must be that, $\tilde{j} \notin I^*$, as otherwise I^* would have m_{11} $(1,1)$ individuals and $m_{21}+1$ $(2,1)$ individuals. However, this means that $I^* = I_{11} \cup I''$. But then I^* has $m_{11}+1$ $(1,1)$ individuals, which contradicts the optimality of I^*. □

As I_{11} is chosen from $\bar{I}$, but not from $\bar{I} \setminus \{k\}$, the result follows.

[33] This is possible since the preferences value diversity and all individuals in I_{12} are both $(1,1)$ and $(2,1)$ individuals.

A.9. *Proof of Proposition 6*

Fix τ^* and let I denote a set of individuals with $\min_l(\theta, \tau)$ individuals of identity θ. As in the proof of Theorem 2, let M_{ij} denote the number of group i individuals in dimension j in I.

Note that $M(\tau^*)$ is the unique optimal marginal distribution. Moreover, $M_{ij}(I) \leq M_{ij}(M(\tau^*))$ for all i and j by construction. By Lemma 6, I is compatible with $M(\tau^*)$, proving that there exists I' such that $|I \cup I'| = q$, $M(I \cup I') = M(\tau^*)$. This proves the result as $I \cup I'$ has $\min_l \mu_l(\theta, \tau^*)$ individuals of identity θ.

A.10. *Proof of Theorem 3*

Suppose that C satisfies substitutes, within-group responsiveness and acyclicty. Observe that $(s, \theta, n) >_C (s', \theta', n')$ implies that $(s, \theta, n) \neq (s', \theta', n')$. Let $\mathcal{S} = \{s_0, \ldots, s_K\}$ denote the ordered set of scores. Let $H(s, \theta, n)$ denote a set formed by n individuals of type (s, θ).

Lemma 9. *For each $i > 0$, there exist θ and n such that $(s_i, \theta, n) >_C (s_{i-1}, \theta, n)$.*

Proof. Take arbitrary $\theta \neq \theta'$. Consider $I = H(s_i, \theta, q) \cup H(s_i, \theta', q) \cup \{k_\theta, k_{\theta'}\}$, where $t(k_\theta) = (s_{i-1}, \theta)$ and $t(k_{\theta'}) = (s_{i-1}, \theta')$. Suppose that $\hat{I} \in C(I)$. By within-group responsiveness, either $k_\theta \notin \hat{I}$ or $k_{\theta'} \notin \hat{I}$. Wlog, let $k_\theta \notin \hat{I}$ and $n = N_\theta(I)$. Then $(s_i, \theta, n) >_C (s_{i-1}, \theta, n)$. $\square$

For each θ, we compute the number of θ individuals who are guaranteed to be chosen.

$$I_\theta = \left(\bigcup_{\theta' \neq \theta, s \in \mathcal{S}} H(s, \theta', q) \right) \cup H(s_0, \theta, q) \tag{4}$$

Take a $J \in C(I_\theta)$ and let $n_\theta = N_\theta(J)$.

Lemma 10. *For each integer $n \in [n_\theta, q-1]$, $(s_0, \theta, n) >_C (s_0, \theta, n+1)$.*

Proof. Let $\hat{I} = H(s_0, \theta, q)$. As $J \in C(I_\theta)$, by substitutes, $J \in C(J \cup \hat{I})$. Remove $n - n_\theta$ non θ individuals from J to define $\tilde{J}$. By substitutes, there exists $I' \in C(\tilde{J} \cup \hat{I})$ such that all $i \in \tilde{J}$ and $\theta(i) \neq \theta$ are in I'. Thus, $N_\theta(I') = n$, proving $(s_0, \theta, n) >_C (s_0, \theta, n+1)$. $\square$

Lemma 11. *Suppose that $I^* \in C(I)$ and $N_\theta(I^*) < n_\theta$. Then all group θ agents in I are in I^*.*

Proof. Suppose that this doesn't hold. Then there exists I and $i \in I$ such that $I^* \in C(I)$, $\hat{n} = N_\theta(I^*) < n_\theta$, $\theta(i) = \theta$ and $i \notin I^*$. By substites, $I^* \in C(I^* \cup \{i\})$. Let $\hat{I}$ denote the set of non-θ individuals in I^*. Let $\tilde{I} = H(s_0, \theta, q)$.

Claim 5. $I^* \in C(I^* \cup \{i\} \cup \tilde{I})$

Proof. If $I^* \notin C(I^* \cup \{i\} \cup \tilde{I})$, then there is $I' \in C(I^* \cup \{i\} \cup \tilde{I})$ such that $N_\theta(I) > \hat{n}$ and by within-group responsiveness I includes all identity θ individuals in $I^* \cup \{i\}$. Then by substitutes, there exists $I_1 \in C(I^* \cup \{i\})$ such that $N_\theta(I_1) = \hat{n} + 1$. Then $\tau(I^*) \neq \tau(I_1)$, which is a contradiction as both are chosen from $I^* \cup \{i\}$. □

As $I^* \in C(I^* \cup \tilde{I} \cup \{i\})$, by substitutes, $I^* \in C(I^* \cup \tilde{I})$. As $\hat{I} \subseteq I^*$, there is $I_2 \in C(\hat{I} \cup \tilde{I})$ such that $\hat{I} \subseteq I_2$, which implies that $N_\theta(I_2) = \hat{n} < n_\theta$. By construction, there exists $I'_\theta \subset I_\theta$ such that $\tau(I'_\theta) = \tau(\hat{I} \cup \tilde{I})$. Moreover, by substitutes, there is $J' \in C(I'_\theta)$ such that $N_\theta(J') = n_\theta$. However, this is a contradiction as $\tau(J') \neq \tau(I_2)$, $J' \in C(I'_\theta)$ and $I_2 \in C(\hat{I} \cup \tilde{I})$. □

Lemma 12. *There exist u and h such that $s, d >_C s', d'$ implies $u(s) + h(d) > u(s') + h(d')$.*

Proof. Follows from Theorem 4.1 in Fishburn (1970). □

By Lemma 9 u is strictly increasing. To ensure concavity, I will modify h without changing the sets of chosen individuals. Let $\bar{u} = \max_{\theta,n} h(\theta, n) + u(s_k)$ as the largest utility contribution an individual can have under u and h. Define $\tilde{h}$ as follows:

$$\tilde{h}(\theta, n) = \begin{cases} h(\theta, n) & \text{if } n > n_\theta \\ \bar{u} & \text{if } n \leq n_\theta \end{cases} \tag{5}$$

Let $\tilde{h}_\theta(n) = \sum_{i=1}^{n} \tilde{h}(\theta, n)$. By Lemma 10, $\tilde{h}_\theta$ is concave.

Lemma 13. *$U(I)$ where*

$$U(I) = \sum_{i \in I} u(s(i)) + \sum_{\theta \in \Theta} \tilde{h}_\theta(N_\theta(I))$$

rationalizes C.

Proof. For a contradiction, assume it doesn't rationalize C. Then there exists $q-$element subsets I and I' such that $U(I) > U(I')$, $I' \in C(\hat{I})$ for some $\hat{I}$ that includes I. Moreover, we can take I to be a maximizer of $U(\tau(I)) = \max_{\tilde{I} \in 2^{\hat{I}}_q} U(\tau(\tilde{I}))$, which exists by the finiteness of $\hat{I}$.

First, if there exists $i \in I \setminus I'$ and $j \in I' \setminus I$ such that $t(i) = t(j)$, let $\tilde{I} = I' \setminus \{j\} \cup \{i\}$. Note that the statement $\tau(I) \neq \tau(\tilde{I})$, $U(I) > U(\tilde{I})$, $\tilde{I} \in C(\hat{I})$ for some $\hat{I}$ that includes I still holds. We can repeat this until there doesn't exist any $i \in I \setminus \tilde{I}$ and $j \in \tilde{I} \setminus I$ such that $t(i) = t(j)$.

Choose an arbitrary $i \in \tilde{I} \setminus I$. Since C satisfies substitutes, there exists I_C such that $I_C \in C(I \cup \{i\})$ and $i \in I_C$. Thus, there exists $j \in I$ such that $j \notin I_C$. As $t(i) \neq t(j)$, if $\theta(i) = \theta(j)$, within-group responsiveness of C implies $s(i) > s(j)$, which implies $U(I \setminus \{i\} \cup \{j\}) > U(I)$, which contradicts that I is a maximizer of U. Thus, $\theta(i) \neq \theta(j)$, which implies $I_C \setminus \{i\} \cup \{j\} \notin C(I \cup \{i\})$, which implies $(s(i), \theta(i), N_{\theta(i)}(I_C)) >_C (s(j), \theta(j), N_{\theta(j)}(I_C) + 1)$. Then

$$u(s(i)) + h(\theta(i), N_{\theta(i)}(I_C)) > u(s(j)) + h(\theta(j), N_{\theta(j)}(I_C) + 1)$$

Moreover, as $j \notin I_C$, $N_{\theta(j)}(I_C) + 1 > n_{\theta(j)}$, and therefore

$$u(s(i)) + \tilde{h}(\theta(i), N_{\theta(i)}(I_C)) > u(s(j)) + \tilde{h}(\theta(j), N_{\theta(j)}(I_C) + 1)$$

However, above equation indicates $U(I \cup \{i\} \setminus \{j\}) > U(I)$, which is a contradiction as I maximizes utility in $\hat{I}$, which includes $I \cup \{i\} \setminus \{j\}$. □

To prove the second part let C denote the indued choice function. Given $h_\theta(n)$, define $h(\theta, n) = h_\theta(n) - h_\theta(n-1)$, which are concave in n as $h_\theta(n)$ are concave. Assume for a contradiction there exists a cycle at $>_C$. This means that for each (s_i, d_i) and (s'_i, d'_i), $u(s_i) + h(d_i) > u(s'_i) + h(d'_i)$, which implies $\sum_i (s'_i, d'_i)$, $u(s_i) + h(d_i) > \sum_i u(s'_i) + h(d'_i)$, which is a contradiction (s'_i, d'_i) is a permutation of (s_i, d_i). Within-group responsiveness of C is immediate as u is strictly increasing.

To show that C_U satisfies substitutes, suppose that $I'_1 \subseteq I_1 \in C(\hat{I}_1)$ and $\hat{I}_2 \subseteq \hat{I}_1$. Take any $I'_2 \subseteq I'_1 \cap \hat{I}_2$. I will show that there exists $\tilde{I} \in C(\hat{I}_2)$ such that $I'_2 \subseteq \tilde{I}$. Take $I_2 \in C(\hat{I}_2)$ and suppose that $i \in I'_2$ but $i \notin I_2$.

Claim 6. *There exists $j \in I_2$, $j \notin I'_2$ and $I_2 \setminus \{j\} \cup \{i\} \in C(\hat{I}_2)$.*

Proof. Let $\theta(i) = \hat{\theta}$. First, if $N_{\hat{\theta}}(I_1) \leq N_{\hat{\theta}}(I_2)$, then there exists $j \in I_2$, $j \notin I_1$ $\theta(j) = \hat{\theta}$. Moreover, as $i \in I'_2 \subseteq I_1 \in C(\hat{I}_1)$ and $j \notin I_1$, $s(i) \geq s(j)$, as otherwise this would be a contradiction that u is increasing in s. Then $U(I_2 \setminus \{j\} \cup \{i\}) \geq U(I_2)$, which proves the result.

Second, if $N_{\hat{\theta}}(I_1) > N_{\hat{\theta}}(I_2)$, then there exists $j \in I_2$, $j \notin I_1$, $\theta(j) = \theta' \neq \hat{\theta}$ such that $N_{\theta'}(I_2) > N_{\theta'}(I_1)$. As $j \notin I_1$

$$s(i) + h(\hat{\theta}, N_{\hat{\theta}}(I_1)) \geq s(j) + h(\theta', N_{\theta'}(I_1)) \tag{6}$$

As h_θ are concave for all θ, we have

$$s(i) + h(\hat{\theta}, N_{\hat{\theta}}(I_2)) \geq s(j) + h(\theta', N_{\theta'}(I_2)) \tag{7}$$

which implies that $U(I_2 \setminus \{j\} \cup \{i\}) \geq U(I_2)$ and proves the result. $\square$

Repeatedly applying Claim 6, starting with any $I_2 \in C(\hat{I}_2)$, we arrive at a $\tilde{I} \in C(\hat{I}_2)$ such that $I'_2 \subseteq \tilde{I}$, which shows that C satisfies substitutes.

A.11. Proof of Proposition 7

Let θ_1 denote a group with a reserve position that is processed after the final open position that preceeds reserve positions. Let I denote a set of individuals that have q individuals from θ_1 and θ_2, where all θ_1 individuals have scores s_0 and all θ_2 individuals have scores s_{K-1} and for all $j \geq 3$, r_j individuals from θ_j with scores s_1. Let r_1 and r_2 denote the number of reserve positions for θ_1 and θ_2 and o denote the number of open positions. Under I all r_j reserve positions are assigned to θ_j individuals, while open positions are assigned to θ_2 individuals, giving $(s_{K_1}, \theta_2, r_2 + o) >_C (s_0, \theta_1, r_1 + 1)$.

Let n denote the number of θ_1 reserve positions before the first open position. Define $\hat{I}$ by increasing the scores of $n+1$ θ_1 individuals to s_K. At $\hat{I}$, one open position is assigned to a θ_1 individual, while all other open positions are assigned to θ_2 individuals. Thus there are $r_1 + 1$ θ_1 such individuals and $r_2 + o - 1$ θ_2 individuals in $C(\hat{I})$. Thus, $(s_0, \theta_1, r_1 + 1) >_C (s_{K_1}, \theta_2, r_2 + o)$, violating acyclicity.

B. Extensions and Additional Results

B.1. Example 2 Satisfies WARP

Suppose that $A \in C(\hat{A})$ and $A' \in \hat{A}$. I will show there does not exist $\tilde{A}$ such that $A' \in C(\tilde{A})$, $A \subseteq \tilde{A}$ and $A \notin C(\tilde{A})$. We will consider two cases.

First, if $\tau(A) = \tau(A')$, then $A' \in C(\tilde{A})$ implies $A \in C(\tilde{A})$ for all $\tilde{A}$ with $A \subseteq \tilde{A}$ and the result follows.

Second, if $\tau(A) \neq \tau(A')$, there are two subcases. A has one individual from each group, then whenever $A \in \tilde{A}$, then A will be one of the sets that is chosen, and we are done. If $A \in C(\hat{A})$ and A does not have one individual from each group, there are either no a, b or c individuals at $\hat{A}$. I will prove the result for the case where there are no c individuals at $\hat{A}$, the other cases are symmetric. If there are no c individuals at $\hat{A}$, then that there are no c individuals at A and A'. Then A has more a individuals than A'. Moreover, as $\tau(A) \neq \tau(A')$, A' has at least one b individual, thus $\hat{A}$ includes both a and b individuals. Now suppose that $A' \in C(\tilde{A})$ such that $A \subseteq \tilde{A}$. As A' is chosen at $\tilde{A}$, this implies that $\tilde{A}$ has both a and b individuals. Moreover, it also implies that $\tilde{A}$ does not have any c individual, as otherwise a set that includes one individual from each group would be selected. Which is a contradiction as A has more a individuals than A' and there are no c individuals at $\tilde{A}$, so A should have been selected instead of A'.

B.2. Heterogeneous Qualities and Gross Substitutes

This section extends the analysis to the setting where $|\mathcal{S}| > 1$. When $\succeq$ is increasing in scores (or equivalently, satisfies within-group responsiveness), the scores in this model are analogous to (inverse) salaries in Kelso and Crawford (1982), where a higher salary is worse for the institution. Therefore, I adopt the following gross substitutes definition given in Kelso and Crawford (1982). I use $s(I)$ to denote the vector of scores of individuals in I.

Definition 14. *Let $\tilde{I} \subseteq \hat{I} \in C(I)$. Define I' by (weakly) decreasing the scores of all $I \setminus \tilde{I}$. If C satisfies gross substitutes, then there exists $\bar{I}$ such that $\tilde{I} \subset \bar{I}$ and $\bar{I} \in C(I')$.*

Gross substitutes condition requires that if a set of individuals are chosen, and the scores of other individuals decrease, then that set of individuals must still be chosen. I also extend the definition of preferences that don't consider intersectionality to settings with heterogeneous qualities.

Definition 15. *$\succeq$ does not consider intersectionality if $\{s(I), M(I)\} = \{s(I'), M(\tilde{I}')\}$ implies $I \sim I'$.*

With heterogeneous qualities, an institution does not consider intersectionality is indifferent between two sets of individuals whenever they have the same cross-sectional distribution of groups and the same scores. The following proposition shows that the the relationship between intersectionality and the substitutes condition generalizes to this setting.

Proposition 8. *Suppose that $C_{\succeq}$ is induced by $\succeq$ that does not consider intersectionality, satisfies within-group responsiveness and values diversity. Then $C_{\succeq}$ doesn't satisfy gross substitutes.*

Proof. The proof closely follows the proof of Theorem 2 with minor modifications, and included for completeness. Let $\mathcal{S} = \{s_0, \ldots, s_K\}$ denote the ordered set of scores.

Suppose that $\mathcal{J}$ includes q individuals from each $\theta \in \Theta$ with scores s_K and $\succeq$ doesn't consider intersectionality. I say that i is a (j,k) individual if $\theta_j(i) = k$.

Let D^* denote the set of all optimal marginal distributions when all individuals have maximum score s_K. Formally, $d \in D^*$ if there exists I' such that $I' \in C(\mathcal{J})$ and $M(I') = d$. Let d_1 denote an element of D^* with the highest number of $(1,1)$ individuals. m_{11} denotes the number of $(1,1)$ individuals at d_1. D_1^* denotes the set of all optimal group distributions where the number of $(1,1)$ individuals is m_{11}. Let d^*_{11} be a group distribution in D^*_1 with the highest number of $(2,1)$ individuals. m_{21} denotes the number of $(2,1)$ individuals at d^*_{11}.

A set of individuals I is *compatible with* marginal distributions d^* if there exists I' such that $M(I \cup I') = d^*$ and I' is a *complement* of I for d^*. Let $M_{ij}(I)$ ($M_{ij}(d)$) denote the number of group i individuals in dimension j in I (d).

Lemma 14. *If $M_{ij}(I) \leq M_{ij}(d)$ for all i and j, then I is compatible with d.*

Proof. If $M_{ij}(I) < M_{ij}(d)$ for some ij, then for each dimension i', there exists a group j' such that $M_{i'j'}(I) < M_{i'j'}(d)$. Let t denote an individual who belongs to group j' at each dimension i'. Then the set $I \cup \{t\}$ still satisfies $M_{ij}(I) \leq M_{ij}(d)$ and repeating this procedure yields a $\tilde{I}$ such that $M_{ij}(\tilde{I}) = M_{ij}(d)$ and I is compatible with d. □

Case 1: $m_{11} \leq m_{21}$.

Claim 7. *There exists $I_{11} = \{i^{11}_1, \ldots, i^{11}_{m_{11}}\}$ compatible with d^*_{11} where all $i \in I_{11}$ are $(1,1)$ and $(2,1)$ individuals.*

Proof. Since groups $(1,1)$ and $(2,1)$ have (weakly) more individuals at d^*_{11}, one can choose the groups of individuals in I_{11} in other dimensions to satisfy $M_{ij}(I_{11}) \leq M_{ij}(d^*_{11})$ for all i and j. Then the result follows from Lemma 14. □

Let I' denote a complement of I_{11} at d^*_{11}. Take $j \in I_{11}$ and $k \in I'$, where k isn't a $(1,1)$ or $(2,1)$ individual.[34] Define $\tilde{j}$ and $\tilde{k}$ as

$$\theta_1(\tilde{j}) = \theta_1(j), \theta_\ell(\tilde{i}) = \theta_\ell(k) \text{ for all } \ell \neq 1$$

$$\theta_1(\tilde{k}) = \theta_1(k), \theta_\ell(\tilde{k}) = \theta_\ell(j) \text{ for all } \ell \neq 1$$

Let $\tilde{I}_{11} = I_{11} \setminus \{j\} \cup \tilde{j}$ and $I'' = I' \setminus \{k\} \cup \tilde{k}$. Note that I'' is a complement of $\tilde{I}_{11}$ at d^*_{11}. The following claim holds by construction.

Claim 8. *I' and I'' doesn't have any $(1,1)$ individuals. Moreover, I'' has $m_{21} - m_{11} + 1$ group $(2,1)$ individuals.*

Let $\bar{I} = I_{11} \cup I' \cup \{\tilde{j}, \tilde{k}\}$. As $M(I_{11} \cup I') = d^*_{11}$, $I_{11} \cup I' \in C(\bar{I})$. Moreover, let $\hat{k}$ denote an individual with $\theta(\hat{k}) = \theta(k)$ and $s(\hat{k}) < s_K$. Let $\bar{I}_{\hat{k}} = \bar{I} \cup \{\hat{k}\} \setminus \{k\}$.

Lemma 15. *There doesn't exist an $I^* \in C(\bar{I}_{\hat{k}})$ such that $I_{11} \subset I^*$*

[34] This is possible since the preferences value diversity and all individuals in I_{11} are both $(1,1)$ and $(2,1)$ indviduals.

Proof. Suppose that there is such an I^*. First, note that $\hat{k} \notin I^*$, as $\tilde{I}_{11} \cup I'' \succ J$ for all J that includes $\hat{k}$ by within-group responsiveness.

Moreover, it must be that, $\tilde{j} \notin I^*$, since otherwise there will be more than m_{11} $(1,1)$ individuals at I^*, which is a contradiction. However, this means that either $I^* = I_{11} \cup I''$. But then, I^* has $m_{21} + 1$ $(2,1)$ individuals and m_{11} $(1,1)$ individuals, which contradicts the optimality of I^* as d^*_{11} is a group distribution in D^*_1 with the highest number of $(2,1)$ individuals and has m_{21} such individuals. Since $\tilde{I}_{11} \cup I''$ is available and optimal, this is a contradiction. □

The result then follows from the fact that I_{11} is chosen from $\bar{I}$, but not from $\bar{I}_{\hat{k}}$.

Case 2: $m_{11} > m_{21}$. Let $n = m_{11} - m_{21}$.

Claim 9. *There exists* $I_{12} = \{i^{11}_1, \ldots, i^{11}_{m_{21}}, i^1_1, \ldots, i^1_n\}$ *where the first* m_{21} *elements are* $(1,1)$ *and* $(2,1)$ *individuals, rest are* $(1,1)$ *individuals and* I_{12} *is compatible with* d^*_{11}.

Proof. Since groups $(1,1)$ and $(2,1)$ have (weakly) more individuals at d^*_{11}, one can choose the groups of individuals in I_{12} to satisfy $M_{ij}(I_{12}) \leq M_{ij}(d^*_{11})$ for all i and j. Then the result follows from Lemma 14. □

Let I' denote a complement of I_{12} at d^*_{11}. Take $j \in I_{12}$ and $k \in I'$, where k isn't a $(1,1)$ or $(2,1)$ individual.[35] Define $\tilde{j}$ and $\tilde{k}$ as

$$\theta_2(\tilde{j}) = \theta_2(j), \theta_\ell(\tilde{j}) = \theta_\ell(k) \text{ for all } \ell \neq 1$$

$$\theta_2(\tilde{k}) = \theta_2(k), \theta_\ell(\tilde{k}) = \theta_\ell(j) \text{ for all } \ell \neq 1$$

Let $\tilde{I}_{12} = I_{12} \setminus \{j\} \cup \tilde{j}$ and $I'' = I' \setminus \{k\} \cup \tilde{k}$. Note that I'' is a complement of $\tilde{I}_{12}$ at d^*_{11}. The following claim holds by construction.

Claim 10. I' *and* I'' *doesn't have any* $(2,1)$ *individuals. Moreover,* I'' *has* 1 *group* $(1,1)$ *individual.*

Let $\bar{I} = I_{12} \cup I' \cup \{\tilde{j}, \tilde{k}\}$. First, note that $I_{12} \cup I' \in C(\bar{I})$ and $\tilde{I}_{12} \cup I'' \in C(\bar{I})$, since $M(I_{12} \cup I') = M(\tilde{I}_{12} \cup I'') = d^*_{11}$. Moreover, let $\hat{k}$ denote an individual with $\theta(\hat{k}) = \theta(k)$ and $s(\hat{k}) < s_K$. Let $\bar{I}_{\hat{k}} = \bar{I} \cup \{\hat{k}\} \setminus \{k\}$

Lemma 16. *There doesn't exist an* $I^* \in C(\bar{I}_{\hat{k}})$ *such that* $I_{12} \subset I^*$.

[35] This is possible since the preferences value diversity and all individuals in I_{12} are both $(1,1)$ and $(2,1)$ individuals.

Proof. Since $\tilde{I}_{12} \cup I''$ is available and optimal, I^* must also be optimal. For a contradiction, suppose that such an I^* exists. First, note that $\hat{k} \notin I^*$, as $\tilde{I}_{11} \cup I'' \succ J$ for all J that includes $\hat{k}$ by within-group responsiveness.

Moreover, $\tilde{j} \notin I^*$, as otherwise I^* would have m_{11} $(1,1)$ individuals and $m_{21}+1$ $(2,1)$ individuals. However, this means that $I^* = I_{11} \cup I''$. But then I^* has $m_{11}+1$ $(1,1)$ individuals, which contradicts the optimality of I^*. □

As I_{11} is chosen from $\bar{I}$, but not from $\bar{I}_{\hat{k}}$, the result follows. □

Proposition 8 is proved by making the appropriate adjustments to the proof of Theorem 2, where decreasing the scores of a set of individuals mirrors the effect of removing those individuals. It shows that whenever the institutions values higher scoring individuals and has a non-trivial preference for diversity, not considering intersectionality when evaluating diversity will cause failure of the gross substitutes condition.