On Meritocracy in Optimal Set Selection

Thomas Kleine Buening
thomkl@ifi.uio.no
University of Oslo
Oslo, Norway

Meirav Segal
meiravs@ifi.uio.no
University of Oslo
Oslo, Norway

Debabrota Basu
debabrota.basu@inria.fr
Scool, Inria, Univ. Lille, Centrale Lille
UMR 9189 – CRIStAL
Villeneuve-d'Ascq, France

Anne-Marie George
annemage@ifi.uio.no
University of Oslo
Oslo, Norway

Christos Dimitrakakis
christos.dimitrakakis@gmail.com
University of Neuchatel
Neuchatel, Switzerland

ABSTRACT

Typically, merit is defined with respect to some intrinsic measure of worth. We instead consider a setting where an individual's worth is *relative*: when a decision maker (DM) selects a set of individuals from a population to maximise expected utility, it is natural to consider the *expected marginal contribution* (EMC) of each person to the utility. We show that this notion satisfies an axiomatic definition of fairness for this setting. We also show that for certain policy structures, this notion of fairness is aligned with maximising expected utility, while for linear utility functions it is identical to the Shapley value. However, for certain natural policies, such as those that select individuals with a specific set of attributes (e.g. high enough test scores for college admissions), there is a trade-off between meritocracy and utility maximisation. We analyse the effect of constraints on the policy on both utility and fairness in an extensive experiments based on college admissions and outcomes in Norwegian universities.

1 INTRODUCTION

Meritocracy [1] is the idea that individuals should be allocated opportunities, resources and power in proportion to their talent, abilities or achievements. However, it is hard to envisage an intrinsic measure of individual worth: for example, merit for a particular position depends on the required qualifications. When considering filling multiple positions simultaneously, merit also depends on who else has been selected. In this paper, we examine a definition of meritocracy for such *set selection* problems, where the decision maker (DM) aims to maximise expected utility. Intuitively,

meritocracy can be defined in terms of how much each individual contributes to the utility. As we discuss extensively in this paper, this intuition holds true for particular utility functions, as long as the DM's selection policy is not constrained by other fairness considerations.

Set selection problems appear in many settings where the DM must select a subset from a candidate population, such as college admissions or hiring decisions. The DM wishes to find a selection policy that maximises utility in expectation. On the other hand, meritocracy demands that individuals with higher merit have a higher probability of being selected. Our first question is how to define meritocracy in this setting.

In particular, as long as there is an inherent, static measure of worth, what is or is not a merit-based (or meritocratic fair) decision is well-defined. This is typically the case when there is an undisputed or dictated scoring system that induces a ranking over individuals [26]. In the absence of such a scoring system, the DM may base the merit of an individual on the utility function instead. However, while a utility function over sets supplies us with a definite best set (or collection thereof), it does not generate individual-level judgements.

One possibility is to use the Shapley value [22], a solution concept from cooperative game theory [8], founded on the idea that the potential contributions of an individual to the utility correspond to their merit. However, contributions to the utility can be relative and dependent on circumstances. More specifically, in this paper we focus on inherently non-linear utility functions defined over sets so that individual contributions to utility depend on who else is being selected. The following example illustrates this.

EXAMPLE 1. *In another scenario, Alice (A), Bob (B), Carlos (C) and David (D) may be candidates for an engineering team. Then, the overall utility we can extract from the team is not necessarily a simple linear function as the qualities of the individuals interact. For example, suppose that we wish to select a team of two and D performs adequately only when paired with C, otherwise he is disruptive. We can model this by setting $U(\{A, D\}) = U(\{B, D\}) = 0$ and $U(\{C, D\}) = 1$, where the function U denotes expected utility. Now, suppose that A works well together with B and adequately with C, i.e. $U(\{A, B\}) = 2$ and $U(\{A, C\}) = 1$. Let in all other cases the utility be zero.*

The Shapley value for the utility function defined in Example 1 can be seen to be Shapley$(U) = (^1/_6, 0, 0, -^1/_6)$. This would suggest that a fair, merit-based selection should choose A with highest

probability, B and C equally likely, and D with lowest probability. However, can we consider this fair? Clearly, in Example 1, a utility maximising DM will *always* prefer to select A rather than D, as selecting A is guaranteed to yield at least as high utility as selecting D. This raises an important question: Should the contribution to utility of an individual when paired with D be as strongly weighted as the contribution of said individual when paired with A?

We argue that since the actual contribution of an individual to utility depends on who else has been selected, *the contribution of individuals should depend on the DM's selection policy.* If the DM is maximising expected utility, then more useful sets have a higher probability of selection and individual contributions should be weighted more heavily in relation to those sets. Hence, unlike static measures of merit such as the Shapley value, we will view the utility-based merit of individuals, i.e. the potential contribution to utility, as a policy-dependent quantity.

Contributions. This paper investigates the problem of selecting a set of individuals out of a candidate pool as an utility maximisation problem. In this setting, we define an individual's contribution to the utility as a dynamic measure of merit. We do this by introducing the notion of *Expected Marginal Contribution* (EMC), modelling the potential contribution of an individual to the utility under a given policy. We then propose a definition of *meritocracy* derived from and suited for arbitrary utility functions over sets and analyse its links to the EMC of individuals. We show that when the DM's policy is egalitarian, i.e. confers equal selection probability to every individual, the EMC is identical to the Shapley value. A natural way to move an egalitarian policy towards meritocracy is to increase the probability of selecting an individual according to their EMC. We show that this idea corresponds to a *policy gradient algorithm* for a specific class of policies, which are separably parameterised over the population. For this class of policies, we show that the policy gradient is a linear transformation of the EMC, and thus maximising utility also achieves meritocratic outcomes. While meritocracy per say does not guarantee other notions of fairness, the DMs utility function could include some fairness measure. This, however, decouples the EMC from the true merit of an individual. Another option is to add constraints to enforce fairness notions. For *constrained policies*, e.g. those that must select candidates through a parameterised function, or which have group fairness requirements, maximising utility does not yield meritocratic policies. Finally, we also perform experiments in a simulated college admission setting with a dataset derived from Norwegian university data. In this setup, we measure utility and meritocracy of the selection policies with and without group fairness and structural constraints. We show that while unconstrained policies maximise both, the constraints result in either reduced utility or meritocracy, which supports our theoretical results.

2 RELATED WORK

Most work on fairness in set selection has focused on policies that rank individuals according to some fixed criterion [4, 7, 15, 20, 27]. This approach satisfies meritocracy, since "better" and higher ranked individuals are preferred. In particular, Kearns et al. [15] consider a probabilistic ordering, generalising Dwork et al. [10]'s notion of similar treatment to selection over multiple groups. More precisely, a person *i* in group *A* is preferred to a person *j* in group *B* only if their relative percentile ranking is higher. They extend this basic definition to different amounts of information available to the decision maker ranging from *ex ante* to *ex post* fairness. Singh and Joachims [23] propose a fair ranking approach for Plackett-Luce models. Kleinberg and Raghavan [16] analyse a stylised parametric model of individual *potential* and Celis et al. [6] consider interventions for ranking, where each individual has a latent utility they would generate if hired. Emelianov et al. [13] also examine latent worth with variance depending on group (e.g. gender) membership. Instead, we use a utility maximisation perspective, where meritocracy rewards individuals according to their contribution to utility, which depends on who else is selected. This is in contrast to the above ranking methods, which implicitly assume a fundamental worth for individuals. In our setting, the contribution of each individual to the utility depends on who else is selected so that the worth of an individual depends directly on the DM's policy.

Fair set selection problems can also naturally be found in social choice, e.g. in participatory budgeting [2] and committee voting [12, 18], as well as matching problems [14, 19]. However, in our setting the selected set is evaluated by a general utility function of the DM rather than by some specific aggregation of voters' preferences like in committee voting or participatory budgeting (for which adding another candidate is usually always beneficial). Our work is more closely related to [17], which considered linear utilities, with the individual performance and group fairness depending on who else is selected. Dwork and Ilvento [11] and Bairaktari et al. [3] considered fairness-as-smoothness in *cohort selection* for linear utilities. In contrast, we focus on a quite different question: how to define and ensure meritocratic fairness under non-linear utilities in set selection.

The EMC, to which our definition of meritocracy is strongly related, can be seen as a generalisation of the Shapley value [22]. The Shapley value measures the weighted average of individual (marginal) contributions over all possible sets. This measure, possibly most known from cooperative game theory, is often used in order to share compensations, costs or other utilities among agents according to their contribution to the entity in a game [24]. An experimental study of reward allocations given by humans [9] found that the Shapley distribution is often not a natural choice as humans tend to weigh single-player coalitions more heavily and often violate some of the defining axioms of the Shapley value (while satisfying others). The EMC weights individuals' contributions to sets/coalitions based on a given policy's probability of selecting the set. For some policy spaces, iteratively adapting the policy in the direction of EMCs leads to a meritocratic policy. Whether a meritocratic distribution corresponds more closely to social norms than the Shapley value would be an interesting real-life study for future work.

3 SETTING AND NOTATION

We formulate the problem of selecting a set of individuals from a population from a general decision theoretic perspective, where the DM aims to select a subset of individuals maximising expected utility.

We consider a population of N candidates $\mathcal{N} = \{1, \ldots, N\}$. The DM observes the features of the population $\mathbf{x} \in \mathcal{X}$ with $\mathbf{x} \triangleq (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, makes a decision $\mathbf{a} \in \mathcal{A}$ about the population using a (stochastic) policy $\pi(\mathbf{a} \mid \mathbf{x})$, observes an outcome $\mathbf{y} \in \mathcal{Y}$, and obtains utility $u(\mathbf{a}, \mathbf{y})$. Since the outcomes are uncertain, the DM's goal is to maximise the *expected* utility $U(\pi, \mathbf{x})$ given features $\mathbf{x}$ and policy π.

To make our results concrete, we focus on the case where $\mathcal{A} = \{0, 1\}^N$ and interpret a decision $\mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_N)$ with $\mathbf{a}_i = 1$ and $\mathbf{a}_j = 0$ as selecting individual i and rejecting individual j. We slightly overload notation and let $\pi(\mathbf{a}_i = 1 \mid \mathbf{x})$ denote the (marginal) probability of i being selected under π given $\mathbf{x}$, i.e. $\pi(\mathbf{a}_i = 1 \mid \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 1} \pi(\mathbf{a} \mid \mathbf{x})$. Similarly, $\pi(\mathbf{a}_i = 0 \mid \mathbf{x})$ denotes the probability of i being rejected under policy π.

In our experiments (Section 6), $\mathcal{Y}$ is a product space encoding outcomes for every individual in the population. The utility function is defined over sets and outcomes $u : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, where $u(\mathbf{a}, \mathbf{y})$ denotes the utility of the selection $\mathbf{a}$ w.r.t. the outcomes $\mathbf{y}$. In this setting, the expected utility $U(\mathbf{a}, \mathbf{x})$ of taking action $\mathbf{a}$ given $\mathbf{x}$ can be calculated by marginalising over outcomes:

$$U(\mathbf{a}, \mathbf{x}) \triangleq \mathbb{E}[u \mid \mathbf{a}, \mathbf{x}] = \sum_{\mathbf{y} \in \mathcal{Y}} \mathbb{P}(\mathbf{y} \mid \mathbf{a}, \mathbf{x}) \, u(\mathbf{a}, \mathbf{y}). \tag{1}$$

Here, $\mathbb{P}(\mathbf{y} \mid \mathbf{a}, \mathbf{x})$ is assumed to be a given predictive model used by the DM for outcome probabilities. We want to emphasise that the focus of this paper is not the fairness or bias of the predictive model $\mathbb{P}(\mathbf{y} \mid \mathbf{a}, \mathbf{x})$, but instead notions of meritocratic fairness based on the DM's utility function and selection policy. Now, the expected utility $U(\pi, \mathbf{x})$ of a policy π given population $\mathbf{x}$ takes the form:

$$U(\pi, \mathbf{x}) \triangleq \mathbb{E}_\pi[u \mid \mathbf{x}] = \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a} \mid \mathbf{x}) U(\mathbf{a}, \mathbf{x}). \tag{2}$$

While $U(\mathbf{a}, \mathbf{x})$ naturally induces a ranking over sets, it does not necessarily provide a ranking over individuals as each individual's contribution to utility may depend on the group selected alongside it.[1]

Problem formulation. The goal of the DM is to find a parameterised policy in the policy space $\Pi = \{\pi_\theta \mid \theta \in \Theta\}$ after observing a given population with features $\mathbf{x}$ that maximises expected utility. That is, we seek

$$\theta^*(\mathbf{x}) = \underset{\theta \in \Theta}{\arg\max}\, U(\pi_\theta, \mathbf{x}), \tag{3}$$

so that the chosen policy takes into account all the information $\mathbf{x}$ we have about the current population.[2]

4 EXPECTED MARGINAL CONTRIBUTIONS AND MERITOCRACY

Here we consider the notion of expected marginal contribution of individuals for a specific policy and expected utility U. In the following, we will omit $\mathbf{x}$ for brevity, since the expected utility is always conditional on $\mathbf{x}$. In addition, the outcomes $\mathbf{y}$ and their distribution play no role in the following development.

Our definition of meritocracy is derived from the DM's utility and is based on two stability axioms. We show that these stability notions have meaningful links to the expected marginal contribution of individuals. For formal proofs of the statements in this section, we refer to Appendix A.1.

4.1 The Expected Marginal Contribution

In the following, we will use $U(\mathbf{a}) \equiv U(\mathbf{a}, \mathbf{x})$ to denote the expected utility of selecting $\mathbf{a}$, and $U(\mathbf{a} + i)$ of adding the individual i to the selection, i.e. $\mathbf{a} + i = (\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, 1, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_N)$. Their difference, $U(\mathbf{a} + i) - U(\mathbf{a})$ can be seen as the marginal contribution of the individual to the set $\mathbf{a}$.

We generalise this to the *Expected Marginal Contribution* (EMC) of an individual i under a policy π:

$$\mathrm{EMC}_i(U, \pi) \triangleq \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a}) \big[U(\mathbf{a} + i) - U(\mathbf{a}) \big], \tag{4}$$

where the implicit dependence on $\mathbf{x}$ has been dropped for brevity.

Informally, the EMC of individual i corresponds to the gain (or loss) in expected utility when a policy is modified so as to always pick individual i. The concept of individual contributions to utility has been studied in cooperative game theory, where the celebrated Shapley value [22] constitutes a fair resource allocation based on marginal contributions. The Shapley value of a utility function U (also called a characteristic function) is defined as

$$\mathrm{Shapley}_i(U) = \frac{1}{N} \sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 0} \binom{N-1}{\|\mathbf{a}\|_1}^{-1} \big[U(\mathbf{a} + i) - U(\mathbf{a}) \big].$$

In particular, we see that the Shapley value corresponds to the EMC for an egalitarian policy under which every individual has the same probability of being selected.

REMARK 1. *Under the egalitarian selection policy* $\pi_{egal}(\mathbf{a}) = \frac{1}{N\binom{N-1}{\|\mathbf{a}\|_1}}$, *we have* $\mathrm{EMC}(U, \pi_{egal}) = \mathrm{Shapley}(U)$.

To emphasise the difference between the Shapley value and the EMC, let us revisit the set selection problem of Example 1.

EXAMPLE 2 (EXAMPLE 1 CONTINUED). *Recall the scenario from Example 1 with applicants Alice (A), Bob (B), Carlos (C) and David (D) as well as expected utilities $U(\{A, B\}) = 2$, $U(\{A, C\}) = 1$, $U(\{C, D\}) = 1$, and $U(S) = 0$ for all other sets S. In this situation, A has (arguably) the highest merit and D the lowest merit. In particular, a utility maximising DM would always prefer to select A over D. This is also expressed in the Shapley value and the EMC under the egalitarian policy π_{egal}:* $\mathrm{Shapley}(U) = \mathrm{EMC}(U, \pi_{egal}) = (^1/_6, 0, 0, -^1/_6)$. *Now, let us account for this and consider the uniform policy that would never select D without also selecting A:*

$$\pi(\mathbf{a}) = \begin{cases} 0, & \text{if } \mathbf{a}_A = 0 \wedge \mathbf{a}_D = 1 \\ \frac{1}{12}, & \text{otherwise.} \end{cases}$$

Under policy π, we see that the EMCs of (A, B, C, D) are given by $\mathrm{EMC}(U, \pi) = (^3/_{12}, ^1/_{12}, -^1/_{12}, -^2/_{12})$. *Thus, in contrast to the Shapley value (which is oblivious to π), the EMC endows B with higher merit than C; accounting for the fact that A should always be preferred over D. We see that as the DM changes their policy towards sets with high utility, the contributions of individuals shift as well and we may have to rethink an individual's utility-based merit.*

---

[1]For the majority of this paper, we state definitions and results with respect to U, without loss of generality. To see this, consider deterministic outcomes.

[2]The related problem of choosing a policy before seeing the current population is not treated in this paper.

The Shapley value [22] is characterised by four desirable axioms of fair division: symmetry, linearity, the treatment of null players, and efficiency. In fact, we see that the EMC satisfies analogous axioms, except for efficiency, which does not apply to the EMC as it lacks a corresponding normalisation.

Lemma 1 (Axioms of Fair Division). *The EMC satisfies*

1) **Symmetry:** *If utility function U and individuals $i, j \in \mathcal{N}$ are such that $U(\mathbf{a}+i) = U(\mathbf{a}+j)$ for all $\mathbf{a} \in \mathcal{A}$, then $\mathrm{EMC}_i(U, \pi) = \mathrm{EMC}_j(U, \pi)$ for all policies $\pi \in \Pi$.*

2) **Linearity:** *If U_1 and U_2 are two utility functions, then $\mathrm{EMC}(\alpha U_1 + \beta U_2, \pi) = \alpha \, \mathrm{EMC}(U_1, \pi) + \beta \, \mathrm{EMC}(U_2, \pi)$ for all policies $\pi \in \Pi$ and $\alpha, \beta \in \mathbb{R}$.*

3) **Null Players:** *If $i \in \mathcal{N}$ has zero contribution to every set, i.e. $U(\mathbf{a} + i) = U(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}$, then $\mathrm{EMC}_i(U, \pi) = 0$ for all policies $\pi \in \Pi$.*

4.2 Meritocracy and Stability Criteria

We now define a notion of meritocracy for general utilities over candidate sets. This definition of meritocratic fairness is based on two properties of the selection policy, namely, swap stability and local stability. We will show that these properties have natural links to the EMC, which suggests that the EMC can be understood as the additional reward individuals should obtain under a policy.

Definition 1 (Swap Stability). A policy π is called *swap stable* if for any two individuals $i, j \in \mathcal{N}$ with $\pi(\mathbf{a}_i = 1) > \pi(\mathbf{a}_j = 1)$, we have $U(\pi + i - j) \geq U(\pi - i + j)$, where $U(\pi + i - j) \triangleq \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a}) U(\mathbf{a} + i - j)$.

In other words, a policy π is swap stable if for any two individuals $i, j \in \mathcal{N}$, from which i is more likely to be selected, the utility of selecting i but not j is higher than the utility of selecting j but not i. Note that instead of writing $U(\pi + i - j)$, we can also write $U(\pi')$, where policy π' shifts all the mass to sets that include i but not j, i.e. $\pi'(\mathbf{a}) = \pi(\mathbf{a}) + \pi(\mathbf{a} - i) + \pi(\mathbf{a} + j) + \pi(\mathbf{a} - i + j)$ if $a_i = 1$ and $a_j = 0$, and $\pi'(\mathbf{a}) = 0$ otherwise. We establish the following link to the EMC, where we use that $\mathrm{EMC}_i(U, \pi - i - j) = \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a})[U(\mathbf{a} + i - j) - U(\mathbf{a} - i - j)]$.

Lemma 2. *For $i, j \in \mathcal{N}$ and policy π, $U(\pi + i - j) \geq U(\pi - i + j)$ if and only if $\mathrm{EMC}_i(U, \pi - i - j) \geq \mathrm{EMC}_j(U, \pi - i - j)$.*

Hence, for a swap stable policy, the merit of i, measured by the EMC of i when selecting neither i nor j, is higher than that of j. Next, we define locally stable policies. Intuitively, these have the property that the DM rewards individuals as much as their self-interest allows them to. In other words, if an individual could contribute positively to the utility, the individual ought to be selected with higher probability.

Definition 2 (Local Stability). A policy π is *locally stable* if for any $i \in \mathcal{N}$, we have $U(\pi) \geq U(\pi + i)$, where $U(\pi + i) \triangleq \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a}) U(\mathbf{a} + i)$.

Instead of $U(\pi + i)$, we can once again write $U(\pi'')$, where policy π'' shifts all the mass to sets that include i, i.e. $\pi'(\mathbf{a}) = \pi(\mathbf{a}) + \pi(\mathbf{a} - i)$ if $a_i = 1$, and $\pi''(\mathbf{a}) = 0$ otherwise. Thus, stability under local changes guarantees that the expected utility of selecting i is lower than the expected utility of π. In fact, we see that this is equivalent to $\mathrm{EMC}_i(U, \pi) \leq 0$.

Lemma 3. *For $i \in \mathcal{N}$ and policy $\pi \in \Pi$ we have $U(\pi) \geq U(\pi + i)$ if and only if $\mathrm{EMC}_i(U, \pi) \leq 0$.*

We now define a meritocratic policy as the one satisfying both stability properties.

Definition 3 (Utility-Based Meritocracy). A policy π is called *meritocratic* if it is swap stable and locally stable.

This notion of meritocracy ensures that our policy is fair to individuals in the sense that individuals with lower chances of being selected have no justification of increasing their chances or swapping their chances of being selected with another individual under a meritocratic policy.

Example 3 (Example 1 continued). *Once again, let us recall Example 1 with applicants Alice (A), Bob (B), Carlos (C) and David (D) as well as expected utilities $U(\{A, B\}) = 2$, $U(\{A, C\}) = 1$, $U(\{C, D\}) = 1$, and $U(S) = 0$ for all other sets S. The uniform policy $\pi(\mathbf{a}) = 1/16$ for all $\mathbf{a} \in \mathcal{A}$ is vacuously swap stable as all individuals have equal probability of selection. However, π is not locally stable as $U(\pi) = 4/16$, whereas $U(\pi + A) = 6/16$. In contrast, the policy that selects the set $\{C, D\}$ with probability one is meritocratic. While $\{C, D\}$ is not utility maximising, the selection is locally optimal in the sense that changing only C or D does not yield strictly larger utility. In particular, we see that selecting the "globally" optimal set (in this case $\{A, B\}$) with probability one is always meritocratic fair according to our definition.*

Lemma 4. *Any deterministic policy π that maximises expected utility is meritocratic.*

Whereas local stability is satisfied by any optimal policy (cf. Proof of Lemma 4), we find that swap stability generally does not hold for *stochastic* optimal policies.

Lemma 5. *A stochastic policy π that maximises expected utility may not satisfy swap stability.*

The proof of Lemma 5, which can be found in Appendix A.1, relies on a counterexample. The counterexample suggests that members of utility maximising sets (or optimal sets) can have different merit by being more (or less) compatible across these optimal sets. This suggests that for the swap stability to hold the selection policy has to satisfy one of the following conditions. First, the DM needs to select optimal sets in a manner that gives each optimal set member the same selection probability, thereby avoiding comparison all together. Alternatively, the DM needs to select individuals with similar contributions across *all* the optimal sets similarly often.

The latter condition leads to the concept of individually *smooth* policies in the sense that similar individuals (in terms of compatibility across optimal sets) are being selected similarly often. This might be another desirable property of meritocratic fair decisions as it would favour individuals that do well in a variety of optimal groups. Note that in general stochastic optimal policies could be preferred over deterministic ones, since from a group-level perspective, it appears unfair to select any single set with probability one in the presence of other sets with equal utility. More generally, the concept of smoothness could also be applied to the selection of suboptimal individuals, which would result in smooth decisions over the whole population (and not only optimal individuals).

5 POLICY STRUCTURES AND OPTIMISATION

Here we consider how different constraints on the policy structure affect the utility and meritocracy of optimal policies. Section 5.1 shows that when the DM is can decide to accept or reject each individual separately, maximising utility is meritocratic. However, frequently the DM is constrained to apply a uniform criterion for acceptance or rejection, such as a threshold, as we discuss in Section 5.2. However, such policies

In particular, Section 5.1 shows that the policy gradient $\nabla_\theta U(\pi_\theta, \mathbf{x})$ is a linear transformation of the EMC for specific separably parameterised policies. This demonstrates a connection between meritocracy and utility maximisation via the EMC of individuals by the results of the previous section. Threshold policies, in contrast, may not always enjoy this property. In either case, optimal policies can be found through a policy gradient algorithm (Section 5.3). For brevity, we again omit the feature vector $\mathbf{x}$ here and remind ourselves that all policies $\pi(\mathbf{a} \mid \mathbf{x})$ as well as expected utilities $U(\pi, \mathbf{x})$ and $U(\mathbf{a}, \mathbf{x})$ are conditional on $\mathbf{x}$.

5.1 Separable Policies

We say that a parameterised policy π_θ is *separable over the population* $\mathcal{N}$ if $\theta = (\theta_1, \ldots, \theta_N)$ and $\pi_\theta(\mathbf{a}) = \prod_{i=1}^{N} \pi_\theta(\mathbf{a}_i)$, where the probability of selecting individual i takes form

$$\pi_\theta(\mathbf{a}_i) = \frac{g(\mathbf{a}_i, \theta_i)}{Z(\theta)} \tag{5}$$

for some function g and normalisation $Z(\theta)$. This means that the probability of a selection $\mathbf{a}$ can be factorised in terms of individual parameters θ_i and a common denominator depending on θ.

Softmax policies. A natural choice for policies that take form as in (5) are softmax policies of the following kind. For $\beta \geq 0$ and $\theta = (\theta_1, \ldots, \theta_N)$, we define

$$\pi_\theta(\mathbf{a}) = \frac{e^{\beta \theta^\top \mathbf{a}}}{\sum_{\mathbf{a}' \in \mathcal{A}} e^{\beta \theta^\top \mathbf{a}'}}. \tag{6}$$

Here, $\beta \geq 0$ is called the inverse temperature of the distribution. While such policies have the advantage that the probability of selecting a set is naturally constrained, they are clearly impractical for large scale experiments as calculating the denominator is computationally heavy. However, we get a first glance at the intimate relationship between the policy gradient and the EMC of individuals.

LEMMA 6. *The gradient of the softmax policy π_θ as defined in* (6) *is a linear transformation of the EMC. More precisely, for every $i \in \mathcal{N}$ we have $\nabla_{\theta_i} U(\pi_\theta) = \beta \pi_\theta(\mathbf{a}_i = 1) \, \mathrm{EMC}_i(U, \pi_\theta)$.*

Note that the EMC of individual i under policy π constitutes the expected gain (or loss) in expected utility of always adding individual i under π. Consequently, if individual i is being selected with probability 1, then $\mathrm{EMC}_i(U, \pi_\theta)$ is zero. More generally, the EMC of an individual i is decreasing as the probability of selecting individual i is increasing. For this reason, we can understand the factor $\pi_\theta(\mathbf{a}_i = 1)$ in Lemma 6 as a normalisation of the EMC opposing this effect, as we would expect the policy gradient to be agnostic about the current share of individual i.

Linear policies. From a computational point of view, it is appealing to consider separable policies so that the probability of selecting individual i truly depends on the parameter θ_i only, i.e. $Z(\theta) = const$ in equation (5). Then, the probability of selecting or rejecting individual i is given by $\pi_\theta(\mathbf{a}_i) = \pi_{\theta_i}(\mathbf{a}_i)$. We want to emphasise that such policy structures do not render decisions about individual i independent from those about individual j. For instance, the expected gain (or loss) $U(\pi + i)$ from adding individual i might be small or large depending on whether individual j is likely to be included under π. We now introduce separable linear policies as these have a particularly intuitive structure. Separable linear policies select individual i with probability θ_i, i.e.

$$\pi_{\theta_i}(\mathbf{a}_i) = \theta_i \mathbb{I}\{\mathbf{a}_i = 1\} + (1 - \theta_i)\mathbb{I}\{\mathbf{a}_i = 0\}.$$

A decision $\mathbf{a} \in \mathcal{A}$ is then assigned the probability $\pi_\theta(\mathbf{a}) = \prod_{i=1}^{N} \pi_{\theta_i}(\mathbf{a}_i)$. We again find that there is a natural link between the policy gradient and the EMC of individuals.

LEMMA 7. *For separable linear policies, if $\pi_{\theta_i}(\mathbf{a}_i = 0) > 0$, then $\nabla_{\theta_i} U(\pi_\theta) = \frac{\mathrm{EMC}_i(U, \pi_\theta)}{\pi_{\theta_i}(\mathbf{a}_i=0)}$ for every $i \in \mathcal{N}$.*

We observe that the gradient has a similar form as in Lemma 6, where in this case the "normalising" factor is the reciprocal of $\pi_{\theta_i}(\mathbf{a}_i = 0)$. In particular, we see that for separable linear policies as well as separable softmax policies the policy gradient is equal to the (uniformly scaled) EMC when evaluated at any egalitarian policy that gives the same selection probability to all individuals. The relation of the Shapley value, EMC and policy gradient is also exemplarily illustrated in Figure 1.

5.2 Threshold Policies

While separable policies can make arbitrary decisions, which may be undesirable from the point of view of fairness, we here consider a class of policies that apply the same decision rule to every individual. This uniformity of treatment should lead to similar outcomes for similar individuals. While this may not make sense when hiring a team of experts, it is eminently suitable for college admission settings where transparent and easily interpretable decision rules are preferable. We define these policies by parameterising over the feature space $\mathcal{X}$. One natural example is a policy of logistic type, where $\theta = (\theta_1, \ldots, \theta_{|\mathcal{X}|})$ and the probability of selecting individual i is given by

$$\pi_\theta(\mathbf{a}_i = 1 \mid \mathbf{x}_i) = \frac{e^{\theta^\top \mathbf{x}_i}}{1 + e^{\theta^\top \mathbf{x}_i}}.$$

Such policies can be understood as threshold policies, since an individual i with feature vector $\mathbf{x}_i$ such that $\theta^\top \mathbf{x}_i > 0$ is being selected at least 50% of the time. In practice, these policies can be projected to the next closest vertex in the simplex so as to attain a deterministic threshold, namely, individual i is being selected if only if $\theta^\top \mathbf{x}_i > 0$.

5.3 A Policy Gradient Algorithm

A natural way to find an optimal policy parameter θ^* is to use policy gradient (Algorithm 1). For policies that are separably parameterised over the population, this naturally shifts the DM's

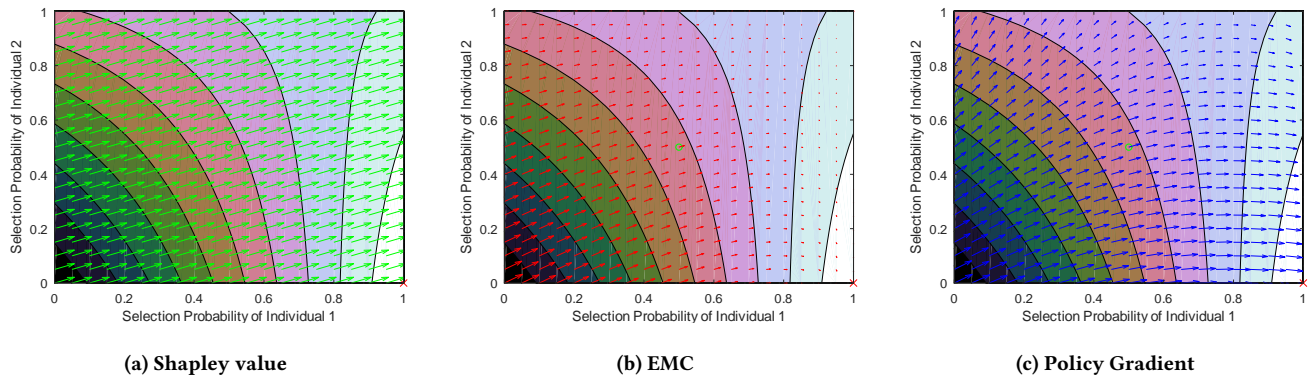**(a) Shapley value** **(b) EMC** **(c) Policy Gradient**

Figure 1: Shapley value, EMC and policy gradient for two individuals, a separable linear policy, a log-linear utility, and a selection cost of $c = 0.3$ (cf. Section 6). Light colours denote higher utility, dark colours lower utility. The maximum is indicated by a red cross and the egalitarian policy $(0.5, 0.5)$ by a green circle. We see that at $(0.5, 0.5)$, the Shapley value, EMC and policy gradient point in the same direction. While the Shapley value points to the same direction everywhere, the EMC changes directions depending on the current policy and so does the policy gradient. In particular, the EMC tends to zero as the probability of selecting both individuals tends to one. As we will see in Lemma 6 and Lemma 7 the policy gradient is closely related to the EMC. In particular, across the diagonal $y = x$ both point in the same direction.

policy towards meritocratic fair decisions as the policy gradient is a linear transformation of the EMC.

Algorithm 1 Policy gradient algorithm

1: **Input**: a model $\mathbb{P}(\mathbf{y}|\mathbf{a}, \mathbf{x})$, a population $\mathcal{N}$ with features $\mathbf{x}$ and a utility function u.
2: **Initialise**: θ^0, $\delta > 0$, learning rate $\eta > 0$
3: **while** $\| \theta^{i+1} - \theta^i \| > \delta$ **do**
4:      Evaluate $\nabla_{\theta} U(\pi_{\theta}, \mathbf{x})$ from u, $\mathbf{x}$ and $\mathbb{P}$.
5:      $\theta^{i+1} \leftarrow \theta^i + \eta [\nabla_{\theta} U(\pi_{\theta}, \mathbf{x})]_{\theta=\theta^i}$
6:      $i ++$
7: **end while**
8: **return** $\pi_{\theta^{i+1}}$

We visualise the policy gradients and associated EMCs in Figure 1. They have similar directions except for the edges where the probability of selecting an individual is almost 1. This illustrates the effect of the "normalising" factors emerging in Lemma 6 and Lemma 7. We also observe that the Shapley value, the EMC, and the policy gradient are similar near the point $(0.5, 0.5)$, i.e. when the policy is almost egalitarian. This validates our arguments of the policy gradient reducing to the EMC and in turn the EMC reducing to the Shapley value under an egalitarian policy.

6 CASE STUDY: COLLEGE ADMISSIONS IN NORWAY

In this section, we perform an empirical case-study on the selection of applicants from a candidate pool in a simulated college admission system based on real-world student data from Norway. Recall that deterministic selection of a utility maximising set always satisfies our notion of meritocracy (Lemma 4). For this reason, we also consider the setting where we impose an ε-statistical parity constraint

on the optimisation and analyse its effect on the meritocracy and utility of the decisions.

6.1 Data and Experimental Setup

Data. We use two datasets from the Norwegian Database for Statistics on Higher Education:

(1) Application data[3] of all applicants to all Norwegian university programs including the following details: birth date, semester of application, gender, citizenship, country of educational background, high school grades in form of GPA and summarised language/science points, other points, admission decision, and each applicant's preference for the program.

(2) Exam data[4] of all students at Norwegian universities for all their taken exams including: courses, study program, and achieved grades.

We use data from the time period of $2017 - 2020$ and limit ourselves to one 5-year master's program with a large number of admitted students at one Norwegian university. In terms of the outcomes, we consider exam data for three mandatory courses in the same program.

Simulator. As our setting is interactive[5], we cannot use a static dataset for evaluation. Instead, we use the data to construct a simulator, which generates populations and outcomes for our experiments (for more details see Appendix B). We simulate applications for the considered 5-year master's program using ctgan [25] with the following features: age (at application), gender, citizenship, country of educational background, high school GPA, science points and language points (from the high school degree), other points and priority (applicant's preference for the specific program). The part

---

[3]https://dbh.nsd.uib.no/dokumentasjon/tabell.action?tabellId=379
[4]https://dbh.nsd.uib.no/dokumentasjon/tabell.action?tabellId=472
[5]Specifically, the exam outcomes depend on our actions, i.e. which students have been admitted, which are fixed in the historical data.

of the simulator producing exam outcomes uses linear regression trained on the exam outcomes of admitted students in the original data.

DM's motivation. We assume that the DM (the faculty) is interested in having successful students across all three courses/disciplines. This could, for instance, be motivated by the wish to produce capable graduates for all necessary fields of practice, so as to uphold the department's reputation.

Data generation. We first generate records for 5,000 students, which represent the admitted students from previous years that have graduated, using the simulator described above and an admission policy described in the supplementary material. Next, we sample again 200 students from the same simulator. This time, the DM does not observe the course results of the students. These students represent this year's applicants from which the DM has to choose.

The DM's regression modelling. Since the DM's utility is tied to individual outcomes, they must estimate a model $\mathbb{P}(\mathbf{y} \mid \mathbf{x}, \mathbf{a})$ for the course results $\mathbf{y}$ of this year's applicants $\mathbf{x}$ in order to make their selection. The DM's model (which is of course different from the one the simulator uses) is estimated using regression on the data from the 5,000 simulated students admitted in prior years.

Utility function and constraint. We assume that the DM is interested in good course results of admitted students across all three considered disciplines/courses. In addition, the DM has a cost c associated with admitting a student. Then, the DM would only select a student if the gain in utility from adding the student to the selection exceeds the cost c of selecting them. For any given set of *outcomes* of the admitted applicants, we design a log-linear utility that values balanced course results across all areas[6]:

$$u(\mathbf{a}, \mathbf{y}) = \sum_{j=1}^{3} \log \left(\sum_{i \in \mathcal{N}} \mathbf{a}_i \cdot \mathbf{y}_{i,j} \right) - c \cdot \| \mathbf{a} \|_1. \tag{7}$$

Since future results $\mathbf{y}$ are unknown to the DM, they instead maximise expected utility by marginalising the above utility function u over the outcomes as described in equation (1).

In addition to the unconstrained optimisation problem, we are interested in the effect that additional constraints have on the meritocracy of the selection policy. Constraints could be needed in order to enable educational opportunities for a minority group. Consider the case where a minority and majority group perform well on their own but have very limited utility when members are mixed. Here we would only select members of the majority group. We apply a typical notion of group fairness, namely, ε-statistical parity with respect to the feature gender. Then, a selection policy π must satisfy

$$|\pi(\mathbf{a}_i = 1 \mid i \text{ is male}) - \pi(\mathbf{a}_j = 1 \mid j \text{ is female})| \leq \varepsilon. \tag{8}$$

Algorithmic comparisons. We test the policy gradient approach for separable linear and threshold policies as introduced in Section 5. For the threshold policies, we use all high-school grades available (grade points, language points, science points and other

---

[6]The potential outcomes of non-admitted applicants do not contribute to the utility

points) as a numerical four dimensional feature. This policy simulates the case in which the DM provides an interpretable policy with predetermined admission criteria. For both policy structures, we allow the policy gradient algorithm 250 updates to converge. In addition, we use the uniformly random selection of sets as a trivial lower baseline. We also run a stochastic greedy algorithm (see e.g. [21]), which is a robust algorithm for finding the utility maximising set in *unconstrained* set selection problems. For the constrained optimisation problem, we use adaptive penalty terms to ensure that the policy gradient algorithm satisfies the constraint (see Appendix B). To obtain a selection satisfying the constraints from the stochastic greedy algorithm, we set the utility of all sets violating the constraint to zero. In particular, note that the stochastic greedy algorithm is usually deployed for unconstrained utility maximisation problems only so that one can expect it to achieve comparably low utility in the constrained case.

6.2 Measuring Deviation from Meritocracy

We are interested in the tensions that arise between utility maximisation and meritocratic fair decisions. To this end, we aim to quantify the *deviation from meritocracy* in our experiments. To measure violations of swap stability as introduced in Definition 1, we suggest to use

$$\text{Dev}_{\text{swap}}(\pi, \mathbf{x}) = \sum_{i,j \in \mathcal{N}} (\pi(\mathbf{a}_i = 1 \mid \mathbf{x}) - \pi(\mathbf{a}_j = 1 \mid \mathbf{x}))^+ \left(U(\pi - i + j, \mathbf{x}) \right.$$
$$\left. - U(\pi + i - j, \mathbf{x}) \right)^+,$$

where $(X)^+ \triangleq \max\{0, X\}$. Here, large values of $\text{Dev}_{\text{swap}}(\pi, \mathbf{x})$ indicate large deviations from swap stable decisions. Note that our choice of Dev_{swap} not only accounts for the number of infringements, but also the magnitude of the deviation from swap stability. For instance, if $U(\pi + i - j) \ll U(\pi - i + j)$ while $\pi(\mathbf{a}_i = 1) \gg \pi(\mathbf{a}_j = 1)$, the measured deviation from swap stability is accordingly large. In particular, if $\text{Dev}_{\text{swap}}(\pi, \mathbf{x}) = 0$, the policy π is swap stable. To measure the deviation from local stability (Definition 2), we use the cumulative positive EMCs under policy π:

$$\text{Dev}_{\text{local}}(\pi, \mathbf{x}) = \sum_{i \in \mathcal{N}} \left(\text{EMC}_i(U, \pi, \mathbf{x}) \right)^+.$$

Again, larger values of $\text{Dev}_{\text{local}}(\pi, \mathbf{x})$ indicate more severe deviations from locally stable decisions. Recall that Lemma 3 states that a policy π is stable under local changes if and only if $\text{EMC}_i(U, \pi, \mathbf{x}) \leq 0$ for all $i \in \mathcal{N}$. Thus, if $\text{Dev}_{\text{local}}(\pi, \mathbf{x}) = 0$, the policy π is stable under local changes.[7]

6.3 Experimental Analysis

Unconstrained optimisation. When solving the pure utility maximisation problem without additional constraints, we observe that the policy gradient algorithm with separable linear policies and the stochastic greedy algorithm achieve the highest expected utility (Figure 2a). Both algorithms yield almost meritocratic fair policies as Dev_{swap} and $\text{Dev}_{\text{local}}$ show small values. In fact, in the unconstrained setting, a deterministic policy that maximises utility is

---

[7]Clearly, one could define deviation from meritocracy differently, e.g. by counting the number of violations. However, we choose to use Dev_{swap} and $\text{Dev}_{\text{local}}$ as they also give insights into the severity of the deviation.

(a) Unconstrained Optimisation

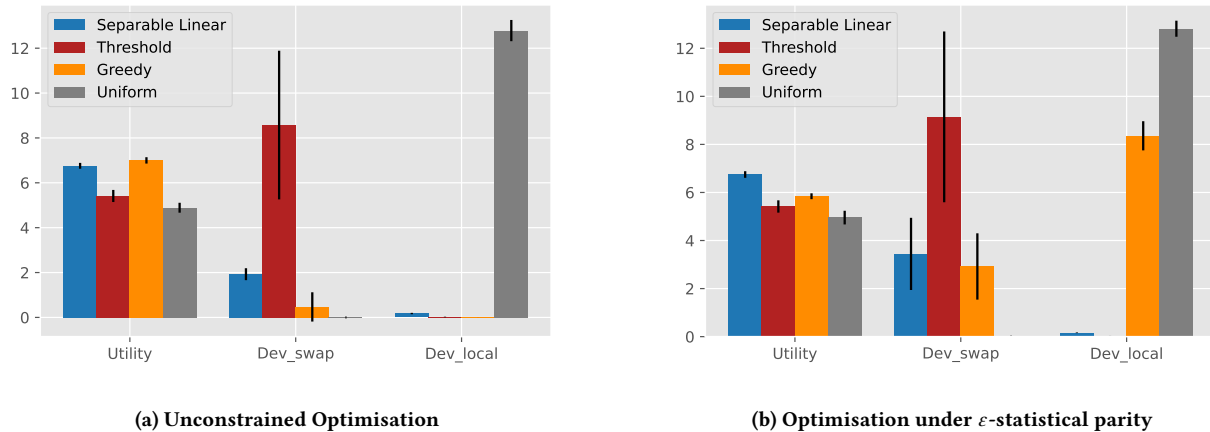(b) Optimisation under ε-statistical parity

Figure 2: Expected utility, Dev$_{swap}$ and Dev$_{local}$ with respect to the *true outcomes* for log-linear utility with cost $c = 0.05$ and 200 applicants. (a): unconstrained utility maximisation for specific policy structures. (b): utility maximisation under ε-statistical parity constraint ($\varepsilon = 0.1$). The results are averaged over 5 repeats (each with different simulated data). The black lines show the standard deviation.

always meritocratic as shown in Lemma 4. The non-zero values of Dev$_{swap}$ and Dev$_{local}$ can be explained by the imperfect predictions of the underlying predictive model. Moreover, the policy gradient algorithm might not have converged to a fully deterministic solution yet, resulting in slightly larger deviations from meritocracy than for the stochastic greedy algorithm, which always yields a deterministic policy. The policy gradient algorithm with threshold policies achieves much lower expected utility than with separably linear policies, which is expected since the policy space is highly constrained. Whereas the threshold policy satisfies local stability, it deviates heavily from swap stability. The uniform set selection policy (which is egalitarian) unsurprisingly achieves the lowest expected utility and is swap stable, i.e. Dev$_{swap}$ = 0, as all individuals have equal probability of selection. However, the uniform policy is not locally stable due to Dev$_{local}$ being non-zero.

Optimisation under statistical parity. The additional ε-statistical parity constraint causes a drop in expected utility for all algorithmic approaches (Figure 2b). Most notably, we see a rise in deviation from swap stability for separable linear policies, threshold policies, and the stochastic greedy algorithm. This suggests that the statistical parity constraint is enforcing the selection of worse individuals (in terms of merit) over better ones on the basis of their group affiliations. Moreover, we observe high deviation from local stability for the stochastic greedy algorithm. This is in accordance with its notable decrease in utility in the constrained setting.[8] The uniform selection policy satisfies the ε-statistical parity constraint and we thus observe no changes.

In general, we observe that unconstrained utility maximisation using separable linear policies or the stochastic greedy selection,

yields approximately meritocratic fair decisions. However, once additional constraints on the policy structure, e.g. threshold policies, or the selection probabilities, e.g. statistical parity, are being imposed, utility maximisation appears to clash with our notion of meritocracy.

7 DISCUSSION AND FUTURE WORK

We have provided a first look into how one can define meritocracy in a general set selection scenario. We define expected marginal contribution (EMC) as the quantifier of an individual's contribution to DM's expected utility. We propose to measure meritocratic fairness with EMC as a dynamic representative of individual merits, in contrast to the static scoring or ranking systems. We show that the Shapley value is a special case of the EMC, when the DM applies an egalitarian policy, and that it satisfies analogues of the Shapley 'fairness' axioms, as well as two notions of selection stability.

We also observe that any utility maximising deterministic policy satisfies meritocracy (Lemma 4) but it might not be the case for stochastic policies (Lemma 5). Specially, the counterexample developed to prove this fact indicates that this might require *individually smooth* policies in the sense that similar individuals (in terms of compatibility across optimal sets) are being selected similarly often. Similar types of observations are found in the fairness literature, where smoothness induces individual fairness. It would be interesting in future to explore this link between smoothness and fairness to design meritocratic fair policies.

In particular, our experiments support the connection between meritocracy, EMC and policy gradient for linearly separable policies. A similar connection for threshold policies seems unlikely given our experimental results and the fact that the policy parameter space has a different dimension from the space of individual contributions. This is perhaps an important consideration for policy makers that create formulaic decision rules. In future work, it will

---

[8]We want to emphasise again that the stochastic greedy algorithm is typically deployed for unconstrained utility maximisation settings. As the stochastic greedy algorithm performs successive selection of individuals it may stop selection prematurely in the presence of additional constraints such as statistical parity, which can prevent it from adding an individual to its selection.

be interesting to analyse the case where individual decisions must be made independently of the remaining applicants.

Our experiments also show that the meritocratic fairness deteriorates as we impose group fairness constraints, such as statistical parity, on the policy. This indicates the philosophical and quantitative tension between meritocratic notions of fairness and demographic notions of fairness [5]. Meritocratic fairness cares about the individuals and tries to influence the selection policy in terms of the worth (or merit) of individuals. In contrast, demographic fairness cares about different subpopulations in the candidate pool and tries to influence the policy in order to select from these subpopulations equally or proportionally. Thus, under some circumstances meritocratic and demographic notions of fairness may be conflicting. Our experimental observations resonate this and our approach of formulating statistical parity as a constraint on the policy and EMC as a measure of merit allows us to quantify this tension. In future, it would be interesting to examine general conditions on the policy constraints under which demographic notions of fairness align with meritocracy.

Finally, another interesting line of work would be to extend this framework to matching problems.

REFERENCES

[1] 2021. Meritocracy, noun. In *Cambridge English Dictionary Online.* Cambridge University Press. https://dictionary.cambridge.org/dictionary/english/meritocracy Retrieved 15 December 2021.

[2] Haris Aziz and Nisarg Shah. 2021. Participatory budgeting: Models and approaches. In *Pathways Between Social Science and Computational Social Science.* Springer, 215–236.

[3] Konstantina Bairaktari, Huy Le Nguyen, and Jonathan Ullman. 2021. Fair and Optimal Cohort Selection for Linear Utilities. *arXiv preprint:2102.07684* (2021). arXiv:2102.07684 [cs.DS]

[4] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval.* 405–414.

[5] Reuben Binns. 2020. On the Apparent Conflict between Individual and Group Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20).* Association for Computing Machinery, New York, NY, USA, 514–524. https://doi.org/10.1145/3351095.3372864

[6] L Elisa Celis, Anay Mehrotra, and Nisheeth K Vishnoi. 2020. Interventions for ranking in the presence of implicit bias. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 369–380.

[7] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with fairness constraints. *arXiv preprint:1704.06840* (2017).

[8] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. 2011. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5, 6 (2011), 1–168.

[9] Greg d'Eon and Kate Larson. 2020. Testing Axioms Against Human Reward Divisions in Cooperative Games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (Auckland, New Zealand) *(AAMAS '20).* International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 312–320.

[10] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference.* 214–226.

[11] Cynthia Dwork and Christina Ilvento. 2018. Group fairness under composition. In *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT* 2018).*

[12] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare* 48, 3 (2017), 599–632.

[13] Vitalii Emelianov, Nicolas Gast, Krishna P Gummadi, and Patrick Loiseau. 2020. On Fair Selection in the Presence of Implicit Variance. In *Proceedings of the 21st ACM Conference on Economics and Computation.* 649–675.

[14] Rustamdjan Hakimov and Dorothea Kübler. 2020. Experiments on centralized school choice and college admissions: a survey. *Experimental Economics* (2020), 1–55.

[15] Michael Kearns, Aaron Roth, and Zhiwei Steven Wu. 2017. Meritocratic fairness for cross-population selection. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70.* JMLR. org, 1828–1836.

[16] Jon M. Kleinberg and Manish Raghavan. 2018. Selection Problems in the Presence of Implicit Bias. *arXiv preprint:1801.03533* (2018).

[17] Matt J. Kusner, Chris Russell, Joshua R. Loftus, and Ricardo Silva. 2018. Causal Interventions for Fairness. *arXiv preprint:1806.02380* (2018). arXiv:1806.02380 [stat.ML]

[18] Martin Lackner and Piotr Skowron. 2020. Approval-Based Committee Voting: Axioms, Algorithms, and Applications. *arXiv preprint:2007.01795* (2020). arXiv:2007.01795 [cs.GT]

[19] David Manlove. 2013. *Algorithmics of matching under preferences.* Vol. 2. World Scientific.

[20] Michael Mathioudakis, Carlos Castillo, Giorgio Barnabo, and Sergio Celis. 2019. Affirmative Action Policies for Top-k Candidates Selection, With an Application to the Design of Policies for University Admissions. *arXiv preprint arXiv:1905.09947* (2019).

[21] Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi, Jan Vondrák, and Andreas Krause. 2015. Lazier than lazy greedy. In *Proceedings of the AAAI Conference on Artificial Intelligence,* Vol. 29.

[22] Lloyd S. Shapley. 1951. *Notes on the n-Person Game II: The Value of an n-Person Game.* Technical Report. Rand Corporation.

[23] Ashudeep Singh and Thorsten Joachims. 2019. Policy learning for fairness in ranking. In *Advances in Neural Information Processing Systems.* 5427–5437.

[24] Eyal Winter. 2002. Chapter 53 The Shapley value. Handbook of Game Theory with Economic Applications, Vol. 3. Elsevier, 2025–2054. https://doi.org/10.1016/S1574-0005(02)03016-3

[25] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. 2019. Modeling Tabular data using Conditional GAN. In *Advances in Neural Information Processing Systems.*

[26] Michael Dunlop Young. 1994. *The rise of the meritocracy.* Transaction Publishers.

[27] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* 1569–1578.

A PROOFS

A.1 Proofs from Section 4

PROOF OF LEMMA 1. Symmetry: If $U(\mathbf{a} + i) = U(\mathbf{a} + j)$ for all $\mathbf{a} \in \mathcal{A}$, then the marginal contributions of i and j for all $\mathbf{a}$ must be equal. This implies that $\text{EMC}_i(U, \pi) = \text{EMC}_j(U, \pi)$ for any policy $\pi \in \Pi$.

Linearity: Using that $(\alpha U_1 + \beta U_2)(\mathbf{a}) = \alpha U_1(\mathbf{a}) + \beta U_2(\mathbf{a})$, we get

$$\text{EMC}_i(\alpha U_1 + \beta U_2, \pi) = \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a})[(\alpha U_1 + \beta U_2)(\mathbf{a} + i)$$
$$- (\alpha U_1 + \beta U_2)(\mathbf{a})]$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a})[\alpha U_1(\mathbf{a} + i) - \alpha U_1(\mathbf{a}) + \beta U_2(\mathbf{a} + i) - \beta U_2(\mathbf{a})]$$
$$= \alpha \text{EMC}_i(U_1, \pi) + \beta \text{EMC}_i(U_2, \pi).$$

Null Player Property: If $U(\mathbf{a} + i) = U(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}$, then the marginal contributions of individual i are zero for all $\mathbf{a} \in \mathcal{A}$. Thus, the expected contributions under any policy $\pi \in \Pi$ must be zero as well, i.e. $\text{EMC}_i(U, \pi) = 0$. □

PROOF OF LEMMA 2. First, recall that $U(\pi - i - j) = U(\pi')$, where $\pi'(\mathbf{a}) = \pi(\mathbf{a}) + \pi(\mathbf{a} + i) + \pi(\mathbf{a} + j) + \pi(\mathbf{a} + i + j)$ for $a_i = a_j = 0$, otherwise $\pi(\mathbf{a}) = 0$. Thus,

$$\text{EMC}_i(U, \pi - i - j)$$
$$= \sum_{\substack{a_i = 0 \\ a_j = 0}} \pi'(\mathbf{a})[U(\mathbf{a} + i) - U(\mathbf{a})]$$
$$= \sum_{\substack{a_i = 0 \\ a_j = 0}} (\pi(\mathbf{a}) + \pi(\mathbf{a} + i) + \pi(\mathbf{a} + j)$$
$$+ \pi(\mathbf{a} + i + j))U(\mathbf{a} + i) - U(\pi - i - j)$$
$$= \sum_{\substack{a_i = 1 \\ a_j = 0}} (\pi(\mathbf{a} - i) + \pi(\mathbf{a}) + \pi(\mathbf{a} - i + j)$$
$$+ \pi(\mathbf{a} + j))U(\mathbf{a}) - U(\pi - i - j)$$
$$= U(\pi + i - j) - U(\pi - i - j)$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a})[U(\mathbf{a} + i - j) - U(\mathbf{a} - i - j)]$$

Now, it follows that

$$U(\pi + i - j) - U(\pi - i + j)$$
$$= U(\pi + i - j) - U(\pi - i - j) + U(\pi - i - j) - U(\pi - i + j)$$
$$= \text{EMC}_i(U, \pi - i - j) - \text{EMC}_j(U, \pi - i - j).$$

This shows that $U(\pi + i - j) \geq U(\pi - i + j)$ is equivalent to $\text{EMC}_i(U, \pi - i - j) \geq \text{EMC}_j(U, \pi - i - j)$. □

PROOF OF LEMMA 3. This readily follows from the definition of the EMC of individual i under π as $\text{EMC}_i(U, \pi) = \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a})[U(\mathbf{a} + i) - U(\mathbf{a})] = U(\pi + i) - U(\pi)$. □

PROOF OF LEMMA 4. Let $\mathcal{A}^* \triangleq \{\mathbf{a} \in \mathcal{A} : U(\mathbf{a}) = \max_{\mathbf{a}' \in \mathcal{A}} U(\mathbf{a}')\}$ denote the set of utility maximising selections. Note that for any optimal (utility maximising) policy π with $\pi(\mathbf{a}) > 0$, it must hold that $\mathbf{a} \in \mathcal{A}^*$. In particular, if π is a deterministic optimal policy, then $\pi(\mathbf{a}^*) = 1$ for some $\mathbf{a}^* \in \mathcal{A}^*$. Let us now show that such a deterministic optimal policy π that selects

$\mathbf{a}^* \in \mathcal{A}^*$ with probability one is swap stable. For all $i, j \in \mathcal{N}$ with $a_i^* = a_j^*$ we have nothing to check. Hence, consider $i, j \in \mathcal{N}$ such that $\mathbf{a}_i^* = 1$ and $\mathbf{a}_j^* = 0$. We then find that

$$U(\pi + i - j) - U(\pi - i + j) = U(\mathbf{a}^* + i - j)$$
$$- U(\mathbf{a}^* - i + j) = U(\mathbf{a}^*) - U(\mathbf{a}^* - i + j) \geq 0,$$

by optimality of set $\mathbf{a}^*$. We have thus shown that π is swap stable. To show that π is locally stable, note that if π is any (possibly stochastic) optimal policy, we have that $\text{EMC}_i(U, \pi) \leq 0$ for all $i \in \mathcal{N}$. This follows directly from the fact that the support of π must be a subset of $\mathcal{A}^*$ and $U(\mathbf{a}^* + i) - U(\mathbf{a}^*) \leq 0$ for all $i \in \mathcal{N}$. By merit of Lemma 3, we have therefore shown that any optimal policy π is locally stable, including all stochastic optimal policies. □

PROOF OF LEMMA 5. By counterexample. Let $\mathcal{A}^* \triangleq \{\mathbf{a} \in \mathcal{A} : U(\mathbf{a}) = \max_{\mathbf{a}' \in \mathcal{A}} U(\mathbf{a}')\}$ denote the set of utility maximising selections. For $i, j \in \mathcal{N}$, let there be $\mathbf{a} \in \mathcal{A}^*$ with $\mathbf{a}_i = 1$, $\mathbf{a}_j = 0$, and $\mathbf{b} \in \mathcal{A}^*$ with $\mathbf{b}_i = 0$, $\mathbf{b}_j = 1$. Since $\mathbf{a}, \mathbf{b} \in \mathcal{A}^*$, the policy

$$\pi(\mathbf{a}) = {}^2/_3, \quad \pi(\mathbf{b}) = {}^1/_3 = 1 - \pi(\mathbf{a})$$

is optimal. In particular, note that the probability of selecting i under π is strictly greater than the probability of selecting j. Thus, for swap stability to hold, π must satisfy $U(\pi + i - j) \geq U(\pi - i + j)$. Now, let us assume that $U(\mathbf{a} - i + j) = U(\mathbf{a})$ and $U(\mathbf{b} + i - j) < U(\mathbf{b})$, which implies

$$2(U(\mathbf{a}) - U(\mathbf{a} - i + j)) < U(\mathbf{b}) - U(\mathbf{b} + i - j). \tag{9}$$

This can be understood as saying that j is more "compatible" than i, since j achieves maximal utility when added to $\mathbf{a}$, but adding i to $\mathbf{b}$ yields sub-optimal utility. Now, by equation (9), we have that

$$U(\pi + i - j) - U(\pi - i + j) = \frac{2}{3}(U(\mathbf{a}) - U(\mathbf{a} - i + j))$$
$$+ \frac{1}{3}(U(\mathbf{b} + i - j) - U(\mathbf{b})) < 0.$$

Thus, π is not swap stable. □

A.2 Proofs from Section 5

PROOF OF LEMMA 6. Define $Z(\theta) = \sum_{\mathbf{a} \in \mathcal{A}} e^{\beta \theta^\top \mathbf{a}}$ and note that

$$\frac{\partial}{\partial \theta_i} \log Z(\theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{a} \in \mathcal{A} : a_i = 1} e^{\beta \theta^\top \mathbf{a}} = \beta \sum_{\mathbf{a} \in \mathcal{A} : a_i = 1} \pi_\theta(\mathbf{a}),$$

Recall our convention that $\pi_\theta(a_i = x) \triangleq \sum_{\mathbf{a} \in \mathcal{A} : a_i = x} \pi_\theta(\mathbf{a})$ for $x \in \{0, 1\}$. In the following, we omit the factor β as it will be nothing but a constant factor.

$$\frac{\partial}{\partial \theta_i} U(\pi_\theta)$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \frac{\partial}{\partial \theta_i} \pi_\theta(\mathbf{a}) U(\mathbf{a})$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \pi_\theta(\mathbf{a}) \frac{\partial}{\partial \theta_i} \log(\pi_\theta(\mathbf{a})) U(\mathbf{a})$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \pi_\theta(\mathbf{a}) \frac{\partial}{\partial \theta_i} (\theta^\top \mathbf{a} - \log Z(\theta)) U(\mathbf{a})$$
$$= \sum_{\mathbf{a} \in \mathcal{A}} \pi_\theta(\mathbf{a}) \left(a_i - \sum_{\mathbf{a}' \in \mathcal{A} : a_i' = 1} \pi(\mathbf{a}')\right) U(\mathbf{a})$$

$$= \sum_{\mathbf{a} \in \mathcal{A} : a_i = 1} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a}) - \pi_{\boldsymbol{\theta}}(a_i = 1) U(\pi_{\boldsymbol{\theta}})$$

$$= (1 - \pi_{\boldsymbol{\theta}}(a_i = 1))$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 1} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a}) - \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a}) = \pi_{\boldsymbol{\theta}}(a_i = 0) \, e^{\beta \theta_i}$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a} + i) - \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a}) = \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a} + i) - \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a}) = \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi_{\boldsymbol{\theta}}(\mathbf{a}) [U(\mathbf{a} + i) - U(\mathbf{a})] = \pi_{\boldsymbol{\theta}}(a_i = 1)$$

$$\sum_{\mathbf{a} \in \mathcal{A}} \pi_{\boldsymbol{\theta}}(\mathbf{a}) [U(\mathbf{a} + i) - U(\mathbf{a})]$$

(for $\mathbf{a}$ with $a_i = 1$ the summand is zero)

$$= \pi_{\boldsymbol{\theta}}(a_i = 1) \, \text{EMC}_i(U, \pi_{\boldsymbol{\theta}}).$$

□

PROOF OF LEMMA 7. If $\pi_{\boldsymbol{\theta}}(a_i = 0) > 0$, we obtain

$$\frac{\partial}{\partial \theta_i} U(\pi_{\boldsymbol{\theta}}) = \frac{\partial}{\partial \theta_i} \sum_{\mathbf{a} \in \mathcal{A}} \pi_{\boldsymbol{\theta}}(\mathbf{a}) U(\mathbf{a})$$

$$= \sum_{\mathbf{a} \in \mathcal{A}} \frac{\partial}{\partial \theta_i} \pi_{\theta_i}(a_i) \prod_{j \neq i} \pi_{\theta_j}(a_j) U(\mathbf{a})$$

$$= \sum_{\mathbf{a} \in \mathcal{A}} (\mathbb{I}\{a_i = 1\} - \mathbb{I}\{a_i = 0\}) \prod_{j \neq i} \pi_{\theta_j}(a_j) U(\mathbf{a})$$

$$= \sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \prod_{j \neq i} \pi_{\theta_i}(a_i) U(\mathbf{a} + i)$$

$$- \sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \prod_{j \neq i} \pi_{\theta_i}(a_i) U(\mathbf{a})$$

$$= \sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \prod_{j \neq i} \pi_{\theta_j}(a_j) [U(\mathbf{a} + i) - U(\mathbf{a})]$$

$$= \frac{1}{\pi_{\theta_i}(a_i = 0)} \sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \pi(\mathbf{a}) [U(\mathbf{a} + i) - U(\mathbf{a})]$$

$$= \frac{1}{\pi_{\theta_i}(a_i = 0)} \sum_{\mathbf{a} \in \mathcal{A}} \pi(\mathbf{a}) [U(\mathbf{a} + i) - U(\mathbf{a})]$$

(for $\mathbf{a}$ with $a_i = 1$ the marginal contribution is zero)

$$= \frac{\text{EMC}_i(U, \pi_{\boldsymbol{\theta}})}{\pi_{\theta_i}(a_i = 0)}.$$

In line four of our calculations, we needed the following technical identity

$$\sum_{\mathbf{a} \in \mathcal{A}} \mathbb{I}\{a_i = 1\} \prod_{j \neq i} \pi_{\theta_j}(\mathbf{a}_j) U(\mathbf{a}) = \sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 0} \prod_{j \neq i} \pi_{\theta_j}(\mathbf{a}_j) U(\mathbf{a} + i).$$

For completeness, we rigorously prove it here. The left hand side is equal to

$$\sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 1} \prod_{j \neq i} \pi_{\theta_j}(\mathbf{a}_j) U(\mathbf{a}) = \sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 1} \prod_{j \neq i} \pi_{\theta_j}(\mathbf{a}_j) U(\mathbf{a} + i).$$

Now, let $\mathbf{a}$ and $\mathbf{a}'$ only differ in the i-th element, namely, $\mathbf{a}_i = 1$ and $\mathbf{a}'_i = 0$. Clearly,

$$\prod_{j \neq i} \pi_{\theta_j}(a_j) = \prod_{j \neq i} \pi_{\theta_j}(a'_j) \quad \text{and} \quad U(\mathbf{a} + i) = U(\mathbf{a}' + i),$$

which yields

$$\sum_{\mathbf{a} \in \mathcal{A} : a_i = 1} \prod_{j \neq i} \pi_{\theta_j}(a_j) U(\mathbf{a} + i) = \sum_{\mathbf{a} \in \mathcal{A} : a_i = 0} \prod_{j \neq i} \pi_{\theta_j}(a_j) U(\mathbf{a} + i).$$

□

B DETAILS OF THE EXPERIMENTAL SETUP

B.1 Details of the Policy Gradient Algorithm

For the constrained optimisation problem (under statistical parity), we maximise the penalised utility:

$$U'(\pi) = U(\pi) - \lambda f(\pi) \text{ with } f(\pi) = (\pi(\mathbf{a}_i = 1 \mid i \text{ is male})$$

$$- \pi(\mathbf{a}_j = 1 \mid j \text{ is female}))^2.$$

This leads to the constraint policy gradient algorithm.

Algorithm 2 Constrained policy gradient algorithm

1: Input: A population $\mathcal{N}$ with features $\mathbf{x}$, a utility function u, a constraint function f and $\varepsilon > 0$.
2: Initialise θ^0, threshold $\delta > 0$, learning rate $\eta^0 > 0$ and $\lambda^0 \leq 0$
3: **while** $\| \theta^{i+1} - \theta^i \| > \delta$ **do**
4:     Evaluate $\nabla_{\boldsymbol{\theta}}(U(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \lambda^i (f(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \varepsilon))$ and $\nabla_{\lambda}(U(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \lambda(f(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \varepsilon))$ using $\mathbf{x}$
5:     $\theta^{i+1} \leftarrow \theta^i + \eta^i \left\{ \nabla_{\boldsymbol{\theta}} [U(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \lambda(f(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \varepsilon)]_{\boldsymbol{\theta} = \theta^i} \right\}$
6:     $\lambda^{i+1} \leftarrow \lambda^i + \eta^i \left\{ \nabla_{\lambda} [U(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \lambda(f(\pi_{\boldsymbol{\theta}}, \mathbf{x}) - \varepsilon)]_{\lambda = \lambda^i} \right\}$
7:     $i ++$
8: **end while**
9: **return** $\pi_{\theta^{i+1}}$

In the algorithm above, η^i is updated according to a schedule so that $\eta^{i+1} = 0.9 \eta^i$ whenever convergence stalls.

Utility gradient. Note that for all policies we consider $\pi(\mathbf{a} \mid \mathbf{x}) = \prod_{i \in \mathcal{N}} \pi(\mathbf{a}_i \mid \mathbf{x})$. For the separable linear policies, the gradient $\nabla_{\boldsymbol{\theta}} U(\pi_{\boldsymbol{\theta}}, \mathbf{x})$ is given by Lemma 7. Then, to estimate $\text{EMC}_i(U, \pi_{\boldsymbol{\theta}}, \mathbf{x})$, we sample $n = 40$ sets $a_1, \ldots, a_n \in \mathcal{A}$ from $\pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x})$ and approximate the EMC of individual i given $\pi_{\boldsymbol{\theta}}$ and $\mathbf{x}$ by

$$\text{EMC}_i(U, \pi_{\boldsymbol{\theta}}, \mathbf{x}) \approx n^{-1} \sum_{k=1}^{n} [U(\mathbf{a}_k + i, \mathbf{x}) - U(\mathbf{a}_k, \mathbf{x})].$$

Note that due to the separable structure of $\pi_{\boldsymbol{\theta}}$, we have $\pi_{\boldsymbol{\theta}}(\mathbf{a}_i = 1 \mid \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A} : \mathbf{a}_i = 1} \pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x}) = \theta_i$ and are thus not required to approximate this value. Moreover, recall that, using our predictive model, we obtain the expected utility of selecting a set $\mathbf{a}$ given $\mathbf{x}$ by marginalising over the outcomes: $U(\mathbf{a}, \mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{Y}} \mathbb{P}(\mathbf{y} \mid \mathbf{a}, \mathbf{x}) u(\mathbf{a}, \mathbf{y})$. For the logit threshold polices with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{|\mathcal{X}|})$, the gradient can be computed as

$$\frac{\partial}{\partial \theta_i} U(\pi_{\boldsymbol{\theta}}, \mathbf{x})$$

$$\begin{aligned}
&= \sum_{\mathbf{a} \in \mathcal{A}} \frac{\partial}{\partial \theta_i} \pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x}) U(\mathbf{a}, \mathbf{x}) \\
&= \sum_{\mathbf{a} \in \mathcal{A}} \pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x}) U(\mathbf{a}, \mathbf{x}) \frac{\partial}{\partial \theta_i} \log(\pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x})) \\
&= \sum_{\mathbf{a} \in \mathcal{A}} \pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x}) U(\mathbf{a}, \mathbf{x}) \sum_{j=1}^{N} \frac{\partial}{\partial \theta_i} \log(\pi_{\boldsymbol{\theta}}(\mathbf{a}_j \mid \mathbf{x}_j)) \\
&= \sum_{\mathbf{a} \in \mathcal{A}} \pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x}) U(\mathbf{a}, \mathbf{x}) \sum_{j=1}^{N} \mathbf{x}_{j,i} \Big(\mathbb{I}\left\{\mathbf{a}_j = 1\right\} \\
&\qquad \pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 0 \mid \mathbf{x}_j) - \mathbb{I}\left\{\mathbf{a}_j = 0\right\} \pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 1 \mid \mathbf{x}_j) \Big).
\end{aligned}$$

In our experiments, we again sample $n = 40$ sets $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{A}$ from $\pi_{\boldsymbol{\theta}}(\mathbf{a} \mid \mathbf{x})$, and then approximate the gradient by

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} U(\pi_{\boldsymbol{\theta}}, \mathbf{x}) &\approx n^{-1} \sum_{k=1}^{n} U(\mathbf{a}_k, \mathbf{x}) \sum_{j=1}^{N} \mathbf{x}_{j,i} \\
&\left(\mathbb{I}\left\{\mathbf{a}_{k,j} = 1\right\} \pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 0 \mid \mathbf{x}_j) - \mathbb{I}\left\{\mathbf{a}_{k,j} = 0\right\} \pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 1 \mid \mathbf{x}_j) \right).
\end{aligned}$$

Note that we can compute $\pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 0 \mid \mathbf{x}_j)$ directly as $\pi_{\boldsymbol{\theta}}(\mathbf{a}_j = 0 \mid \mathbf{x}_j) = (1 + e^{\boldsymbol{\theta}^\top \mathbf{x}_j})^{-1}$.

Constraint gradient. We can rewrite the penalty term f using the affiliation vectors $M, F \in \{0, 1\}^N$ with $M_i = 1$ if individual i is male and $F_i = 1$ if individual i is female. Then, $\|M\|_1$ denotes the number of males and $\|F\|_1$ denotes the number of females ($\|M\|_1 + \|F\|_1 = \mathcal{N}$). We get the following expression for the penalty term $f(\pi)$:

$$f(\pi) = \left(\frac{\sum_{i : M_i = 1} \pi(a_i = 1 \mid \mathbf{x})}{\|M\|_1} - \frac{\sum_{i : F_i = 1} \pi(a_i = 1 \mid \mathbf{x})}{\|F\|_1} \right)^2 - \varepsilon$$

Its gradient w.r.t. π is then given by

$$\begin{aligned}
\nabla_\pi f(\pi) = 2 &\left(\frac{\sum_{i : M_i = 1} \pi(a_i = 1 \mid \mathbf{x})}{\|M\|_1} - \frac{\sum_{i : F_i = 1} \pi(a_i = 1 \mid \mathbf{x})}{\|F\|_1} \right) \\
&\left(\frac{M}{\|M\|_1} - \frac{F}{\|F\|_1} \right).
\end{aligned}$$

Policy gradient. For linear separable policies, the gradient of π w.r.t. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$ is given by a diagonal $N \times N$-matrix with entries $(\mathbb{I}\{a_i = 1\} - \mathbb{I}\{a_i = 0\})_{i \in [N]}$. For the logit threshold policy with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ $(m = |\mathcal{X}|)$, we have

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}} = \Big[\mathbf{x}_{i,j} \big(&\mathbb{I}\{a_i = 1\} e^{-\boldsymbol{\theta}^\top \mathbf{x}_i} (1 + e^{-\boldsymbol{\theta}^\top \mathbf{x}_i})^{-2} \\
&- \mathbb{I}\{a_i = 0\} e^{\boldsymbol{\theta}^\top \mathbf{x}_i} (1 + e^{\boldsymbol{\theta}^\top \mathbf{x}_i})^{-2} \big) \Big]_{i \in \mathcal{N}}^{j \in [m]}.
\end{aligned}$$

As before, we then sample from $\pi(\mathbf{a} \mid \mathbf{x})$ to approximate the gradient $\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}}$ and thereby obtain $\nabla_{\boldsymbol{\theta}} f(\pi) = \nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}} \nabla_\pi f(\pi)$.

B.2 Simulator Description

In the following, we describe the simulator that was used to generate the data for our experiments. The simulator generates a population of applicants in two steps: 1) generating applicants' features and 2) generating outcomes - grades in three courses given the applicant features (assuming admission). The dataset itself will not be published as it is not publicly open.

Generating students. We wish to generate a population of applicants for a specific study program. To this end, we use ctgan[9], a deep learning based synthetic data generator for tabular data, that can learn from real data and generate synthetic clones with high fidelity. We provide this generator with a clean version of the application data table (described in 6). The data is cleaned by the following steps: First, we keep applications for the selected study program. Then, we filter out all applications that are not through normal admission or do not have a valid application, resulting in roughly 40,000 applications. In some cases, attributes of an applicant might be missing in the application for one program but are present in an application for a different program. We fill missing values of GPA, science points and language points for applicants using the matching values in different study programs in the same year (if exist). After this step, we select the following features: application year, gender, country code for citizenship, country code for educational background, priority, GPA, science points, language points and other points. The age at time of application is also added (calculated as the difference between application year and year of birth). Finally, all applications with remaining missing values are removed, resulting in approximately 30,000 applications. The ctgan generator is trained using this data over 10 epochs with default parameters (see Figure 3 for evaluation).

Generating outcomes. In addition to applicant features, we also generate outcomes for each candidate - grades in three mandatory courses in the study program (one first year course and two second year courses). Naturally, the data only provides grades for admitted students. Thus, we use existing data to train a linear regression model and use it to produce outcomes for generated applicants.

We construct the training data from the two data tables described in 6. The clean version of the application data is used as training features, with the additional following steps: For each applicant, only the last application to this program is kept, the rest are dropped from the table. This results in 15, 000 applicants. We wish to consider the set of accepted students who took the first year course, had data of at least two years in the program, had the right to study and the right to take the exam. This filter results in 900 students. Then, we also filter out students who did not have a grade in the A-F scale in the first year course. We do include students with the mark 'X' stating they did not show up for the exam, and consider it as 'F'. The final set includes 557 students. The training targets are the last grades of each student in the final set in the three selected courses. Before training, all numeric features are scaled between 0 and 1, categorical features are one-hot encoded, and the targets are translated to numeric values in $[0, 1]$ as follows: 'A': 1, 'B':0.8 , 'C':0.6 , 'D':0.4 , 'E':0.2 , 'F':0 . Missing values (which may exist for the second year courses, for example in cases of dropouts) are considered as 0. A linear regression is fitted to the data using default parameters of scikit-learn. Predictions made by the model are clipped to $[0, 1]$, resulting in r2 score of 0.05.

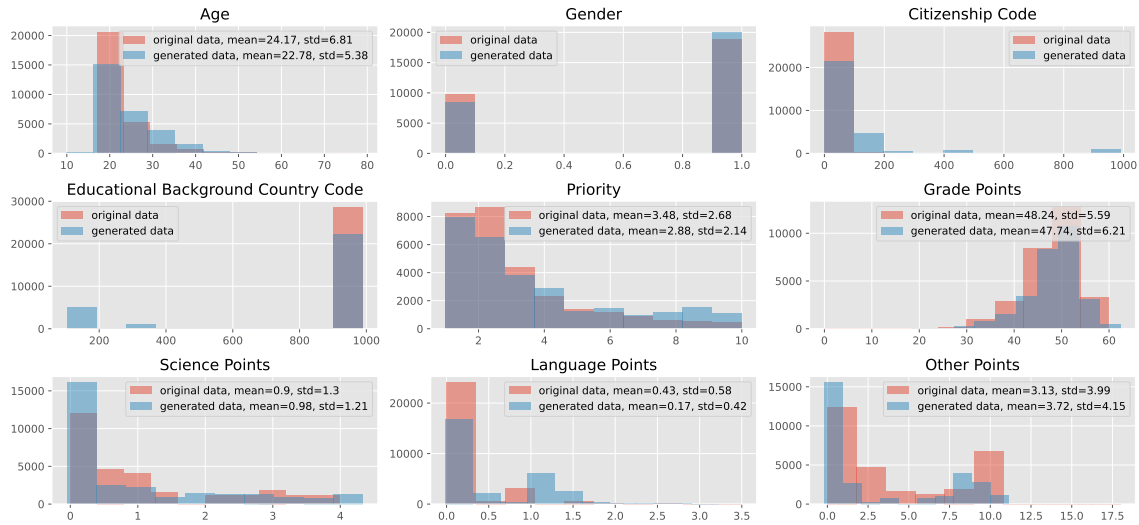---

[9]https://github.com/sdv-dev/CTGAN

Figure 3: Histograms of the different features for the original data (training set) and for the samples generated by CTGAN, using sample size equal to the size of the training set.

Modeling past admissions. To generate the data of students from prior years, we use a model representing the admission policy of the study program that was used in earlier years. This admission policy is modeled by a logistic regression with default parameters trained on approximately 30, 000 samples with 7% positive samples. The training data is identical to the the training data used for generating students, except for the labels which now indicate whether the candidate was offered and accepted the offer to study in the program. Before training, all numeric features are scaled between 0 and 1, categorical features are one-hot encoded. In Norway, a stable marriage algorithm (from the student's perspective) is used to match students to study programs. Thus, a model that can only observe data for one program cannot be completely accurate. Yet, since the features include priority, we can still provide a reasonable prediction. The accuracy achieved by the model given a train-test split of 75 : 25 is 93%. In the admission process we simulate, the k best scores (probability of being admitted) are being selected.

B.3 Additional Results

In addition to Figure 2, we include results comprising the performance of the algorithms with respect to the actual outcomes as well as the predicted outcomes in Figure 4. As shown, the results are almost identical due to the high accuracy (97%) of the regression model. These results also include the performance of the historical policy (described in B.2 under "Modeling past admissions". For these experiments, the policy has selected the top k candidates, where k is set to be identical to the set size selected by the greedy policy. The historical policy does not maximize the utility described in our experiments, thus it seems that it manages to achieve a reasonable utility due to the predetermined set size k. This policy has large values of deviation from meritocracy, which is expected because it does not select necessarily the best candidates, but also takes into account their priorities. In the constrained case, we see a significant increase of $\mathrm{Dev}_{\mathrm{local}}$, along with an increase of $\mathrm{Dev}_{\mathrm{swap}}$ and decreased utility.

(a) Unconstrained Optimisation (True Outcomes)

(b) Constrained Optimisation (True Outcomes)

(c) Unconstrained Optimisation (Predicted Outcomes)
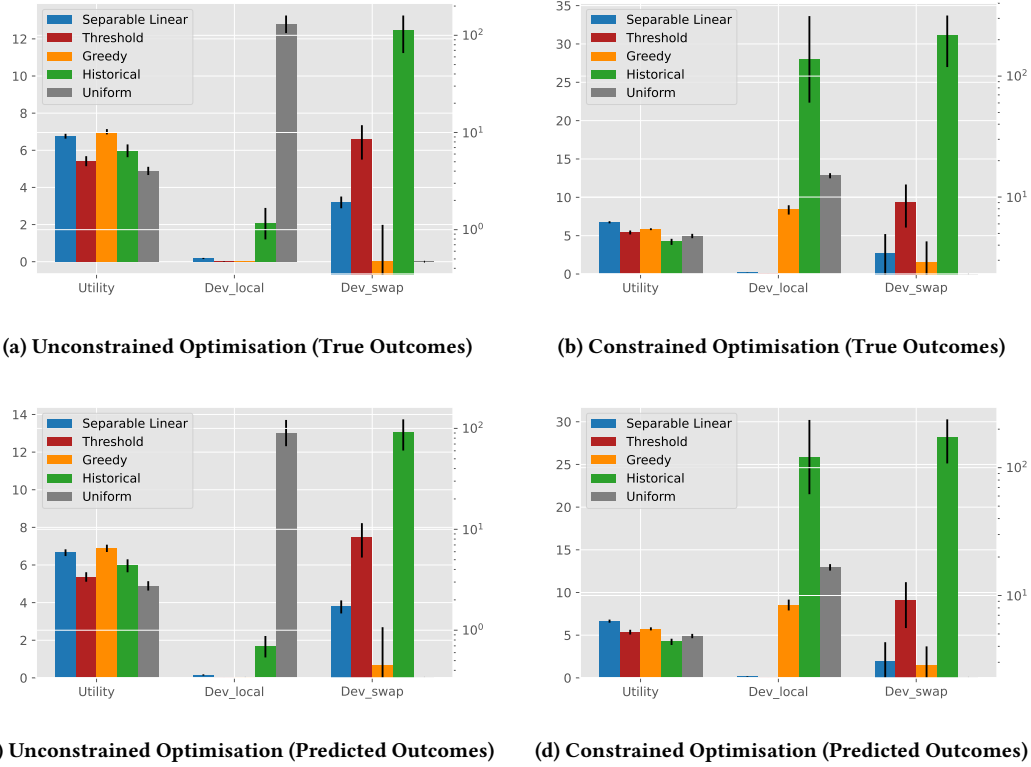
(d) Constrained Optimisation (Predicted Outcomes)

Figure 4: Expected utility, Dev$_{swap}$, and Dev$_{local}$ w.r.t. the true outcomes and the predicted outcomes of individuals for log-linear utility with 200 applicants, selection cost $c = 0.05$, and bias $\varepsilon = 0.1$. The results are averaged over 5 repeats (each with different simulated data). The black lines represent the standard deviation. Dev$_{swap}$ is presented in log scale (appears on the right)