CENTRE DE RECHERCHE SUR LES STRATÉGIES ÉCONOMIQUES

On the price of diversity for multiwinner elections under (weakly) separable scoring rules

Mostapha Diss, Clinton Gubong Gassi, Eric Kamwa

January 2024

Working paper No. 2024 – 02

CRESE
30, avenue de l'Observatoire
25009 Besançon
France
http://crese.univ-fcomte.fr/

UFR SJEPG Sciences juridiques économiques politiques et de gestion

Université de Franche-Comté

On the price of diversity for multiwinner elections under (weakly) separable scoring rules

Mostapha Diss[a,b], Clinton Gubong Gassi[a,c], and Eric Kamwa[d,†]

[a]Université de Franche-Comté, CRESE EA3190, F-25000 Besançon, France. Email: mostapha.diss@univ-fcomte.fr

[b]Africa Institute for Research in Economics and Social Sciences (AIRESS), University Mohamed VI Polytechnic, Rabat, Morocco.

[c]Department of Mathematics - University of Yaounde I. BP 47 Yaounde, Cameroon. Email: clinton.gassi@univ-fcomte.fr

[d]LC2S UMR CNRS 8053, Université des Antilles, Schoelcher Cedex, Martinique, France. Email: eric.kamwa@univ-antilles.fr

[†]Corresponding author

Abstract

We consider a model of multi-winner elections, where each voter expresses a linear preference over a finite set of alternatives. Based on voters' preferences, the primary goal is to select a subset of admissible alternatives, forming what is referred to as a committee. We explore (weakly) separable committee scoring rules, the voting mechanisms that assess each alternative individually using a scoring vector and select the top k alternatives, where k represents the committee's size. Furthermore, we operate under the assumption that alternatives are categorized based on specific attributes. Within each attribute category, there exists a targeted minimum number of alternatives that the selected committee should encompass, emphasizing the necessity for diversity. In this context, we assess the cost associated with imposing such a diversity constraint on the voting process. This assessment is conducted through two methodologies, referred to as the "price of diversity" and the "individual price of diversity". We set the upper bounds for both prices across all (weakly) separable committee scoring rules. Additionally, we show how the maximum price of diversity can be used to discriminate between different voting rules in this context. Ultimately, we illustrate that concentrating on the candidates' performance yields a more accurate estimation of the price of diversity compared to a focus on the enforced diversity constraint.

Keywords: Group decisions and negotiations; Voting; multiwinner elections; scoring rules; price of diversity.

JEL classification: D71, D72.

1 Introduction

Multi-winner elections pose common challenges in social choice theory, where a set of individuals have to aggregate their preferences to select a predefined-sized subset of alternatives from a larger set. This voting scenario is prevalent in real-life situations such as parliamentary elections, candidate shortlisting for competitions, or curating a set of movies for in-flight entertainment. This process is commonly referred to as "committee selection." Formally, a finite set of voters express their preferences over a finite set of alternatives (or candidates) in order to choose a fixed-size subset, known as a "committee." Numerous research endeavors in this field focus on the ordinal setting, where voters establish linear orders over the set of alternatives, ranking them from most to least preferred without ties. The majority of these studies aim to extend single-winner voting rules to the multi-winner framework. Without been exhaustive, the reader can refer to the work of Bock et al. (2019),Brams et al. (2019), Diss and Doghmi (2016), Diss et al. (2020), Elkind et al. (2017), Faliszewski et al. (2018, 2019), Kilgour (2010), Kilgour and Marshall (2012), and Skowron et al. (2019). Remarkably, Elkind et al. (2017) and Skowron et al. (2019) have introduced and characterized the family of committee scoring rules as an extension of the well-established family of positional scoring rules for single-winner elections, initially characterized by Young (1974, 1975). The extensive family of committee scoring rules encompasses the distinctive class of "weakly separable committee scoring rules." In this class, candidates can be individually rated according to a single-winner scoring vector, and the winning committee of size k comprises the k candidates with the highest scores. Notably, this class of committee scoring rules is known by its comprehensibility and ease of implementation. For a more in-depth exploration of this specific category of rules, we refer the reader to Faliszewski et al. (2019, 2018). Note that the subclass of weakly separable committee scoring rules can be viewed as the intersection between committee scoring rules and the "candidates-wise" procedures defined by Kilgour and Marshall (2012). A voting procedure is considered candidate-wise if the score of a given committee is the sum of the scores of each candidate within the committee, with each candidate being treated as a 1-size committee.

The committee selection framework often introduces constraints driven by various concepts and objectives, as discussed by prominent scholars such as Brams (1990), Kamada and Kojima (2015), and Lu and Boutilier (2011), among others. The paper at hand specifically delves into committee selection under diversity constraints, focusing on weakly separable committee scoring rules. The foundational premise here is the existence of a distinguishing attribute –be it gender, age, ethnicity, etc.–that allows the set of candidates to be partitioned into several disjoint classes or types. Each candidate is assigned a label based on this attribute, and the objective is to ensure a specified level of diversity within the selected committee, termed as "diverse committee selection".[1] Various contributions have tackled the challenge of selecting a diverse committee, employing diverse definitions of "diversity constraint". For example, Ianovski (2022) introduced "interval constraints" and "dominance constraints", where the

[1]It is important to note that, for certain attributes like spoken languages, classes may overlap as candidates could belong to multiple classes. However, our focus is on attributes where each candidate falls into a single class (e.g., gender, age, or religion).

former specifies the range within which the number of candidates to be selected from each class must lie, and the latter determines, for any pair of classes, which one merits a higher number of candidates to be selected. Ianovski (2022) explored the computational complexity of selecting a committee that upholds candidate excellence under a defined objective function. Similarly, Aziz (2019) considered that the set of candidates is structured into non-disjoint classes based on a specific attribute, defining a diversity constraint as a vector of integers indicating the minimum number of candidates to be chosen from each class. The author assumed that the preferences of the voters have already been aggregated into a single weak order on the set of candidates and provided an algorithm (a voting procedure) that combines both excellence, expressed by the positions of the candidates in the social weak order, and diversity. Numerous other studies contribute to this framework, including works by Thejaswi et al. (2021) and Kagita et al. (2021), among others. Notably, all the cited works focus on the single-attribute setting. In contrast, some studies delve into the multi-attribute setting, where candidates are labeled according to more than one attribute. While beyond the scope of this paper, interested readers can explore works by Bei et al. (2022), Bredereck et al. (2017), Celis et al. (2017), and Do et al. (2021), among others.

The task of selecting a diverse committee naturally prompts the question: what is the cost associated with enforcing a diversity constraint in the committee selection process? Clearly, the imposition of diversity has repercussions on the "quality" of the selected committee, as it narrows down the set of admissible committees. Consequently, the committee selected with a diversity constraint might have a lower score compared to the committee selected without any diversity constraint. Exploring the impact of the diversity constraint on the quality of the selected committee, this paper delves into the concept referred to as the "price of diversity" in committee selection. To the best of our knowledge, the sole exploration of the cost of the diversity constraint in committee selection comes from Bredereck et al. (2017). They defined the price of diversity as the ratio between the score of the selected committee without any diversity constraint and the score of the selected committee considering the diversity constraint. The authors concentrated on a specific scenario where the set of candidates is partitioned into only two classes, and the diversity constraint necessitates the selected committee to be balanced, i.e., containing an equal number of candidates from each of the two classes. In this context, they demonstrated that, for any scoring rule based on any submodular and monotonic function,[2] the price of diversity cannot exceed 2. A related study by Benabbou et al. (2020) tackled the price of diversity in assignment problems, which is distinctly different from the committee selection task. In the assignment problem, a group of agents is divided into multiple groups, and a set of items is divided into various blocks. Assigning an item to an agent involves a utility for the agent, and the total utility of the assignment is the sum of the utilities of all agents. A quota-based diversity constraint specifies the maximum number of agents from each group allowed in each block of items. The price of diversity is then the ratio between the total utility of the assignment without any diversity constraint and the total utility of the assignment when such a diversity constraint is imposed.

[2]The formal definition is provided in Footnote 5.

In this paper, we explore a framework where the set of candidates is divided into at least two (disjoint) groups based on a specific attribute. The diversity constraint is expressed as a vector of integers specifying the minimum number of candidates required from each group. To maintain consistency in the diversity concept, we assume that the quota bound for each class is at least one, ensuring that at least one candidate is needed from each class. This condition aligns with scenarios such as the apportionment problem in party-list elections in certain countries (refer to the examples provided by Balinski and Young, 1994 for more details). Our focus is on studying the cost of the diversity constraint when selecting a diverse committee under (weakly) separable committee scoring rules. To measure this cost, we propose two approaches. The first is the well-known "price of diversity" as defined by Bredereck et al. (2017). Introducing a second approach called the "individual price of diversity", we base it on the "harms" suffered by the candidates due to the diversity constraint. Initially, we provide a bound on the price of diversity for the entire class of (weakly) separable committee scoring rules. Additionally, we present another bound on the price of diversity that depends on the specific rule under consideration. This allows for discriminating between voting rules by assessing their sensitivity to diversity enforcement. Moreover, we demonstrate that, instead of solely focusing on the applied diversity constraint, evaluating the performance of each type can offer a more nuanced understanding of the price of diversity. Finally, we define the individual price of diversity and show that, while it consistently exceeds the price of diversity, it cannot surpass the tight bound provided for the price of diversity across the entire class of (weakly) separable committee scoring rules.

The structure of the paper is as follows: In Section 2, we establish the formal setting and introduce preliminary definitions. Section 3 delves into the definitions of the two approaches employed to analyze the cost of the diversity constraint in our model, presenting all the key results of the paper. Finally, Section 4 offers conclusions and outlines potential avenues for future research.

2 Formal Setting

In this paper, we adopt the following conventions: for any integer $l \in \mathbb{N}^*$, we denote the set $\{1, \dots, l\}$ simply as $[l]$. For a given set Z, we write 2^Z to denote the family of all its subsets, and 2^Z_k to denote the set of all its subsets of size k. Finally, the notation $\lvert Z \rvert$ is used to express the cardinality of Z.

We examine the following scenario, involving a non-empty and finite set $A = \{a_1, a_2, \dots, a_m\}$ of alternatives/candidates,[3] and a non-empty and finite set $N = \{1, 2, \dots, n\}$ of voters/individuals, where $m \geq 3$ and $n \geq 2$. We make the assumption that each voter ranks, without ties, all the candidates from the most preferred to the least preferred. Thus, the preference of each individual is a linear order on A, signifying a complete, anti-symmetric, and transitive binary relation on the set of candidates A. For a voter $i \in N$, the strict order of i is denoted by $\succ_i$. The n-tuple $\succ = (\succ_1, \succ_2, \dots, \succ_n)$ encompassing all voters' preference

[3] Alternatives may sometimes be represented by lowercase letters a, b, c, etc.

orders is referred to as a preference profile or simply a profile. The set of all linear preferences over A is denoted by $\mathcal{P}$, and the set of all possible profiles with n voters is denoted by $\mathcal{P}^n$. For any pair of candidates a and b, we use the notation $a \succ_i b$ to indicate that voter i strictly prefers candidate a to candidate b. The rank of any alternative $a \in A$ in the preference relation $\succ_i$ of voter i is denoted by $r(\succ_i, a)$ and is defined as:

$$r(\succ_i, a) = \left|\left\{b \in A : b \succ_i a\right\}\right| + 1 = m - \left|\left\{b \in A : a \succ_i b\right\}\right|. \tag{1}$$

Let us reiterate that within our framework, the objective is to identify a fixed-size subset of candidates that most accurately captures the preferences of the voters. For any integer $k \in [m-1]$, we define a "committee" of size k as any k-element subset of A, denoted by an element from the set 2^A_k.[4] As per Elkind et al. (2017), the rank of a committee $W \in 2^A_k$ in voter i's preference relation $\succ_i$, denoted by $r(\succ_i, W)$, is represented by the increasing sequence obtained through the sorting of the set $\left\{r(\succ_i, a) : a \in W\right\}$. The rank of committee W in the preference relation of voter i is then a k-tuple $I = (i_1, \ldots, i_k)$ with $i_1 < \cdots < i_k$. We denote the set of all possible committee ranks in a voter preference relation by $[m]_k$. To illustrate, let us consider an example.

Example 1 *Let us consider the set of voters $N = \{1, \ldots, 7\}$, the set of candidates $A = \{a_1, a_2, a_3, a_4, a_5\}$, and the preference profile*

$$\succ = \begin{bmatrix} a_1 & a_3 & a_5 & a_2 & a_2 & a_4 & a_3 \\ a_2 & a_2 & a_4 & a_1 & a_5 & a_3 & a_1 \\ a_3 & a_4 & a_3 & a_4 & a_4 & a_1 & a_4 \\ a_4 & a_5 & a_1 & a_3 & a_1 & a_2 & a_5 \\ a_5 & a_1 & a_2 & a_5 & a_3 & a_5 & a_2 \end{bmatrix},$$

where for each $i \in \{1, \ldots, 7\}$, the i-th column of the matrix represents the preference relation $\succ_i$ of individual i. Let us consider the second voter and the committee $\{a_1, a_2, a_3\}$). Upon a simple check, we find that $r(\succ_2, a_1) = 5$, $r(\succ_2, a_2) = 2$, and $r(\succ_2, a_3) = 1$. Consequently, we can deduce that $r(\succ_2, \{a_1, a_2, a_3\}) = (1, 2, 5)$.

We posit that the set of candidates is partitioned into $l \geq 2$ classes (or types) denoted as $A_1, \ldots, A_l$ based on a specific attribute such that $A = \bigcup_{j=1}^{l} A_j$. In the course of this paper, we employ the term "class" to denote the subset A_j and the term "type" to refer to the label $j \in [l]$. As stated in the introductory section, while the classes can be non-disjoint, our focus in this paper is on attributes for which each alternative belongs to only one class (e.g., gender, religion, age, etc.). Formally, this implies that $A_j \cap A_{j'} = \emptyset$ for all $j, j' \in [l]$ with $j \neq j'$. For any candidate $a \in A$, we use $j(a)$ to represent the type of a; in other words, $j(a)$ is the integer from the set $[l]$ such that $a \in A_{j(a)}$. It is important to note that the partitioning of the set of candidates does not influence the ranking of any voter. Each voter

[4] It is worth noting that we assume the committee size satisfies $k \in [m-1]$, as the case $k = m$ is straightforward.

provides his/her (hereafter her) sincere ranking, irrespective of the types of candidates. Indeed, acknowledging the potential for voters to strategically vote based on different candidate types would undermine the genuine impact of the diversity constraint. Now, let us introduce the class of rules that we focus on in this paper.

Definition 1 *A multi-winner voting rule (or committee selection rule) is a mapping F that takes any profile $\succ = (\succ_1, \ldots, \succ_n) \in \mathcal{P}^n$ and any positive integer $k \in [m-1]$ as input, and outputs the set $F(\succ, k)$ of one or more committees of size k.*

Note that in the case of ties between two or more committees, there is usually a tie-breaking mechanism that will output a single committee. Now, the committee scoring rules are formally defined as follows:

Definition 2 *A committee scoring function is a function f defined from the set $[m]_k$ to $\mathbb{R}_+$ and that satisfies the following condition: for all $I = (i_1, \cdots, i_k), J = (j_1, \cdots, j_k) \in [m]_k$, $I \succeq J \Rightarrow f(I) \geq f(J)$, where $\succeq$ is the dominance relationship on $[m]_k$ defined by $I \succeq J \Leftrightarrow i_t \leq j_t$ for all $t \in [k]$. The committee scoring rule F associated with the committee scoring function f is the multi-winner voting rule that selects the committee(s) with the highest score under the committee scoring function f.*

In this paper, we are interested in a particular class of committee scoring rules, namely the "(weakly) separable committee scoring rules," which we define below.

Definition 3 *A scoring vector of length m is a vector $w = (w_1, \ldots, w_m)$ such that $w_1 \geq w_2 \geq \cdots \geq w_m$ and $w_1 > w_m$. The point received by each candidate $a \in A$ from each individual preference $\succ_i$ is $w_{r(\succ_i, a)}$. The score $s_w(\succ, a)$ of each alternative $a \in A$, with respect to w, across the profile $\succ = (\succ_1, \ldots, \succ_n)$ is the sum of the points it receives from all the voters. That is,*

$$s_w(\succ, a) = \sum_{i=1}^{n} w_{r(\succ_i, a)}. \tag{2}$$

Given a committee W of size k, the score obtained by W over the profile $\succ$ with respect to the scoring vector w is given by

$$s_w(\succ, W) = \sum_{a \in W} s_w(\succ, a). \tag{3}$$

Definition 4 *A committee scoring rule F is said to be "(weakly) separable" if there exists a scoring vector w such that, for any profile $\succ$ and any committee size k, $F(\succ, k)$ consists of the committee(s) with the highest score $s_w(\succ, W)$.*

To put it differently, a committee scoring rule F is considered (weakly) separable if we can calculate the score of each candidate independently, utilizing a single-winner scoring vector, and subsequently select the k candidates with the highest scores. When the scoring vector w does not depend on the committee size k, we refer to F as "separable" without the "weakly" qualification. It is important to note that any separable committee scoring rule is inherently weakly separable, but the converse is not necessarily true. A (weakly) separable committee scoring rule, associated with a scoring vector w, is then a committee scoring rule with the underlying scoring function defined as $f(i_1, \cdots, i_k) = \sum_{t=1}^{k} w_{i_t}$. Let us give some examples of well-known (weakly) separable committee scoring rules that we focus on in this paper.

k-Plurality: The k-Plurality rule is defined by the scoring vector $w = (1, 0, \ldots, 0)$. Each voter gives one point to her top candidate, and 0 to all the others, and the k candidates with the highest aggregated scores are selected;

k-Borda rule: The k-Borda rule is defined by the scoring vector $w = (m-1, m-2, \ldots, 0)$. In this rule, each voter assigns $m - t$ points to the candidate ranked at the t position in her ranking. Subsequently, the k candidates with the highest aggregated scores are selected.

k-Antiplurality rule: The k-Antiplurality rule is defined by the scoring vector $w = (1, \ldots, 1, 0)$. Each voter gives zero point to her worst candidate and one point to all the others. Subsequently, the k candidates with the highest aggregated scores are selected.

Bloc rule: The Bloc rule uses the scoring vector $w = (\underbrace{1, \ldots, 1}_{k-times}, 0 \ldots, 0)$. Each voter gives one point to each of her top k candidates and zero point to all the others. The k candidates with the highest aggregated scores are selected.

Let us quickly illustrate how the four well-known (weakly) separable committee scoring rules we have just listed operate.

Example 2 *Let us examine the same profile as outlined in Example 1. Suppose the committee size is $k = 2$. The scores of the candidates based on the 2-Plurality rule, the 2-Anti-plurality rule, the 2-Borda rule, and the Bloc rule are computed as follows:*

	$s(\succ, a_1)$	$s(\succ, a_2)$	$s(\succ, a_3)$	$s(\succ, a_4)$	$s(\succ, a_5)$
The 2-Plurality rule	1	2	2	1	1
The 2-Antiplurality rule	6	5	6	7	4
The 2-Borda rule	14	15	16	16	9
The Bloc rule ($k = 2$)	3	4	3	2	2

The winning committee for the 2-Plurality rule is $\{a_2, a_3\}$ while the committees $\{a_1, a_4\}$ and $\{a_3, a_4\}$ tie for the 2-Antiplurality rule; the committee $\{a_3, a_4\}$ is the winning committee for the 2-Borda rule; finally, the committees $\{a_1, a_2\}$ and $\{a_2, a_3\}$ tie for the Bloc rule.

It is worth noting that the k-Plurality rule, the k-Antiplurality rule, and the k-Borda rule are all separable, and consequently, they are weakly separable. However, the Bloc rule is weakly separable but not separable, as its scoring vector depends on the committee size k. In the sequel, when the scoring vector of a (weakly) separable committee scoring rule is known in advance, we can express the score of a candidate (and a committee) as $s(\succ, a)$ (and $s(\succ, W)$), omitting the vector w for simplicity in notation.

In the context of a given preference profile, the objective is to choose a fixed-size committee, considering both the individual scores that reflect the excellence of the candidates and a diversity constraint based on various types. A "diversity constraint" is defined by a quota vector $q = (q_1, \ldots, q_l) \in \mathbb{N}^l$, specifying the minimum number of candidates from each class that the selected committee must include. For each type $j \in [l]$, we denote $\alpha_j = \frac{q_j}{k}$ as the minimum proportion of candidates of type j required in the selected committee. In our setting, the set of candidates is assumed to be known in advance. Therefore, the required quota q_j for a class A_j cannot exceed the total number of candidates from that particular class, i.e., $q_j \leq |A_j|$ for all $j \in [l]$. Notably, given a committee of size k, we have $\sum_{j=1}^{l} q_j \leq k$, indicating that the sum of the minimum quotas cannot surpass the fixed committee size. Furthermore, the concept of "diversity is meaningful if each class is represented by at least one member in the selected committee, i.e., $q_j \neq 0$ for all $j \in [l]$. For diversity consistency, we focus on diversity constraints where $q_j \geq 1$ for all $j \in [l]$. This implies that the number of types should be less than or equal to the committee size k, i.e., $l \leq k$. For any committee W and any type j, we use W^j to represent the set of alternatives of type j within W, denoted as $W^j = W \cap A_j$.

Definition 5 *Let $W \in 2^A_k$ be a committee and q be a diversity constraint. We say that W is a q-diverse committee if $|W^j| \geq q_j$ for all $j \in [l]$. We denote by $2^A_{k,q}$ the set of all q-diverse committees of size k.*

To integrate both diversity and excellence, a (weakly) separable committee scoring rule will contemplate the set of q-diverse committees, denoted as $2^A_{k,q}$, as the set of admissible committees. The rule then selects the committee(s) with the highest score from this set.

Definition 6 *The constrained (weakly) separable committee scoring rule F associated with a scoring vector w assigns to each profile $\succ$, each committee size k, and each diversity constraint q, the set $F(\succ, k, q)$ of committee(s) from $2^A_{k,q}$ with the highest score with respect to w.*

It is evident that the diversity constraint, necessitating a reduction in the domain of acceptable committees, may result in the selection of a committee that is "less optimal" compared to the committee chosen without any diversity constraint. Our interest lies in assessing the cost, in terms of excellence, imposed by the diversity constraint when a constrained (weakly) separable committee scoring rule is employed to select a diverse committee.

3 Price of diversity for (weakly) separable scoring rules

Recall that the idea of capturing and analyzing the cost induced by the diversity constraint in the selection of diverse committees was introduced by Bredereck et al. (2017). They defined the "price of diversity" as the ratio between the score of the selected committee without any diversity constraint and the score of the selected committee when the diversity constraint is enforced. The authors specifically focused on rules within the class of submodular and monotonic objective functions,[5] which includes the class of (weakly) separable committee scoring functions. They have demonstrated that in the case of a binary attribute (i.e., $l = 2$), if the diversity constraint necessitates the selected committee to be balanced (i.e., $q_1 = q_2 = k/2$), the price of diversity cannot surpass 2. This specific scenario obviously corresponds to a particular case in our model, prompting our initial task to extend this result to our broader setting. Additionally, beyond the price of diversity defined by Bredereck et al. (2017), we believe it would be insightful to delve into how each candidate is affected by diversity enforcement. This approach provides an alternative method for gauging the cost induced by the diversity constraint.

3.1 Price of diversity

In this section, we introduce the initial scheme for quantifying the cost of the diversity constraint in our context, aligning with the definition provided by Bredereck et al. (2017).

Definition 7 *Let F be a (weakly) separable committee scoring rule associated with a scoring vector w. Given the committee size k and the diversity constraint q, the price of diversity induced by q on F with respect to the preference profile $\succ$ is given by*

$$POD(F, \succ, q) = \frac{s(\succ, T)}{s(\succ, R)}, \tag{4}$$

where $T \in F(\succ, k)$ and $R \in F(\succ, k, q)$.

Before proceeding to the first result of this section, let us state the following useful claim.

Claim 1 *Every winning committee $T \in F(\succ, k)$ allows to build a winning q-diverse committee $R \in F(\succ, k, q)$.*

Proof. Let $T \in F(\succ, k)$ be a winning committee. If T is a q-diverse committee, then $R = T$. Otherwise, we can construct from T a q-diverse committee R with the highest score (among the q-diverse committees) such that:

- for every $j \in [l]$, $|T^j| < q_j \Rightarrow |R^j| = q_j$ $(T^j \subset R^j)$;
- for every $j \in [l]$, $|T^j| = q_j \Rightarrow R^j = T^j$;
- for every $j \in [l]$, $|T^j| > q_j \Rightarrow |T^j| \geq |R^j| \geq q_j$ $(R^j \subseteq T^j)$.

This construction is carried out iteratively by replacing a candidate with the lowest individual score in a subset T^j, where $|T^j| > q_j$, with a candidate holding the highest individual score from a class $A^{j'} \setminus T$ for a type j' with $|T^{j'}| < q_{j'}$. ■

[5] An objective function f is said to be submodular if, for any two subsets B and C from 2^A, we have $f(B \cup C) - f(B \cap C) \leq f(B) + f(C)$; it is said to be monotonic if, for any subsets B and C from 2^A, we have $B \subseteq C \Rightarrow f(B) \leq f(C)$. Obviously, every (weakly) separable committee scoring function is submodular and monotonic.

Literally, the committee R is constructed from T by systematically substituting candidates from the overrepresented classes with the lowest individual scores with candidates from the underrepresented classes having the highest individual scores, until the diversity constraint is fulfilled. Let us give a simple example.

Example 3 *Consider the set of six candidates* $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ *partitioned as* $A = A_1 \cup A_2 \cup A_3$ *with* $A_1 = \{a_1, a_2, a_3\}, A_2 = \{a_4, a_5\}$*, and* $A_3 = \{a_6\}$*. Assume that the committee size is* $k = 3$ *and that the diversity constraint is* $q = (1, 1, 1)$*, requiring exactly one candidate from each class. Consider the following preference profile with four voters*

$$\succ = \begin{bmatrix} a_1 & a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 & a_2 \\ a_3 & a_3 & a_3 & a_3 \\ a_4 & a_4 & a_4 & a_4 \\ a_5 & a_5 & a_5 & a_5 \\ a_6 & a_6 & a_6 & a_6 \end{bmatrix}$$

Under the 3*-Borda rule, we have* $s(\succ, a_1) = 20$*,* $s(\succ, a_2) = 16$*,* $s(\succ, a_3) = 12$*,* $s(\succ, a_4) = 8$*,* $s(\succ, a_5) = 4$*, and* $s(\succ, a_6) = 0$*. The winning committee without the diversity constraint is then* $T = \{a_1, a_2, a_3\}$*, which is not a* q*-diverse committee. However, from* T *we can construct the* q*-diverse committee* R *with the highest score by replacing* a_3 *and* a_2 *by* a_4 *and* a_6*, and we get* $R = \{a_1, a_4, a_6\}$*.*

As ties may occur among committees with the highest score, it is preferable to select the one that most closely aligns with the diversity constraint. The next definition is therefore useful.

Definition 8 *Let* k *be the committee size,* q *be the diversity constraint,* $\succ$ *be a preference profile, and* F *be a weakly separable committee scoring rule. We will call optimal committee under* F *with respect to* $\succ$*, the committee* $T \in F(\succ, k)$ *that minimizes the value* $\sum_{j \in [l]} \big| |W^j| - q_j \big|$*, and we will call diverse optimal committee any* q*-diverse committee* $R \in F(\succ, k, q)$*.*

In other words, the optimal committee is the committee that belongs to $F(\succ, k)$ and minimizes the number of substitutions needed to reach the diversity constraint.[6]

Example 4 *Consider the same set of candidates, the same committee size, and the same*

[6] The tie-breaking rule is only used when there is more than one optimal committee.

diversity constraint as in Example 3. Consider the following preference profile with four voters

$$\succ = \begin{bmatrix} a_1 & a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 & a_2 \\ a_3 & a_3 & a_4 & a_4 \\ a_4 & a_4 & a_3 & a_3 \\ a_5 & a_5 & a_5 & a_5 \\ a_6 & a_6 & a_6 & a_6 \end{bmatrix}$$

Under the 3-Borda rule, the committees $\{a_1, a_2, a_3\}$ and $\{a_1, a_2, a_4\}$ have the maximum score of 46, and the optimal committee is $T = \{a_1, a_2, a_4\}$, since it needs one substitution to satisfy the diversity constraint, whereas the committee $\{a_1, a_2, a_3\}$ needs two substitutions.

Let us now state the first result of this section.

Proposition 1 *Let F be a (weakly) separable committee scoring rule, k be the committee size, and $q = (q_1, \cdots, q_l)$ be the diversity constraint. For any preference profile $\succ$, we have*

$$POD(F, \succ, q) \leq \frac{1}{\min\limits_{j \in [l]} \alpha_j},$$

where $\alpha_j = q_j/k$ for all $j \in [l]$.

Proof. Let F be the considered (weakly) separable committee scoring rule and $\succ$ a preference profile.

If there is a q-diverse committee T in $F(\succ, k)$, then the price of diversity is equal to 1, since T also belongs to $F(\succ, k, q)$ and the result holds.

Now, assume that there is no q-diverse committee in $F(\succ, k)$, and let $T \in F(\succ, k)$ be the optimal committee with respect to $\succ$, and R be the diverse optimal committee obtained from T. It follows that for each $j \in [l]$ we have:

$$\frac{1}{q_j} s(\succ, R^j) \geq \frac{1}{k} s(\succ, T^j), \tag{5}$$

where for all $W \in 2^A_k$, $s(\succ, W^j) = \sum\limits_{a \in W^j} s(\succ, a)$ is the total score collected by all the members of W^j belonging to W.

Indeed, let $j \in [l]$.

- If $|T^j| < q_j$, then $s(\succ, R^j) \geq s(\succ, T^j)$ since the transformation in Claim 1 increases the number of alternatives picked from A_j ($T^j \subset R^j$). Therefore, it holds that

  $$\frac{1}{q_j} s(\succ, R^j) \geq \frac{1}{q_j} s(\succ, T^j) \geq \frac{1}{k} s(\succ, T^j),$$

  since

  $$q_j \leq k.$$

- If $|T^j| > q_j$, let us set $t_j = |T^j|$ and $r_j = |R^j|$. Without loss of generality, we can write $T^j = \{a_1, \ldots, a_{t_j}\}$ with $s(\succ, a_1) \geq \cdots \geq s(\succ, a_{t_j})$, and then $R^j = \{a_1, \ldots, a_{r_j}\}$ with $r_j \leq t_j$ since the transformation in Claim 1 does not increase (and probably decreases) the total number of alternatives selected from A_j ($R^j \subset T^j$). The difference between the average marginal contributions of candidates with type j, in R and T respectively, is

$$\begin{aligned}
\frac{s(\succ, R^j)}{r_j} - \frac{s(\succ, T^j)}{t_j} &= \frac{s(\succ, a_1) + \cdots + s(\succ, a_{r_j})}{r_j} - \frac{s(\succ, a_1) + \cdots + s(\succ, a_{t_j})}{t_j} \\
&= \frac{(t_j - r_j)[s(\succ, a_1) + \cdots + s(\succ, a_{r_j})] - r_j[s(\succ, a_{r_j+1}) + \cdots + s(\succ, a_{t_j})]}{r_j t_j} \\
&\geq \frac{(t_j - r_j)[s(\succ, a_1) + \cdots + s(\succ, a_{r_j})] - r_j(t_j - r_j)s(\succ, a_{r_j})}{r_j t_j}, \text{ since}
\end{aligned}$$

$s(\succ, a_t) \leq s(\succ, a_{r_j}),\ \forall t \geq r_j$.

$$\frac{s(\succ, R^j)}{r_j} - \frac{s(\succ, T^j)}{t_j} \geq \frac{r_j(t_j - r_j)s(\succ, a_{r_j}) - r_j(t_j - r_j)s(\succ, a_{r_j})}{r_j t_j} = 0, \text{ since}$$

$s(\succ, a_t) \geq s(\succ, a_{r_j}),\ \forall t \leq r_j$.

Hence, it holds that

$$\frac{1}{q_j} s(\succ, R^j) \geq \frac{1}{r_j} s(\succ, R^j) \geq \frac{1}{t_j} s(\succ, T^j) \geq \frac{1}{k} s(\succ, T^j),$$

since

$$r_j \geq q_j \text{ and } t_j \leq k.$$

- The equality holds for any type $j \in [l]$ such that $|T^j| = q_j$, since the transformation in Claim 1 does not affect the subset T^j. Then, for such a type, we have

$$\frac{1}{q_j} s(\succ, R^j) = \frac{1}{q_j} s(\succ, T^j) \geq \frac{1}{k} s(\succ, T^j)$$

As a result, we have $s(\succ, R^j) \geq \frac{q_j}{k} s(\succ, T^j) = \alpha_j s(\succ, T^j)$, for all $j \in [l]$, which implies that $s(\succ, R^j) \geq \min\limits_{j \in [l]} \alpha_j s(\succ, T^j)$ for all $j \in [l]$. Summing over all the types, we obtain $s(\succ, R) \geq \min\limits_{j \in [l]} \alpha_j \times s(\succ, T)$. Thus,

$$POD(F, \succ, q) = \frac{s(\succ, T)}{s(\succ, R)} \leq \frac{1}{\min\limits_{j \in [l]} \alpha_j}.$$

■

It is noteworthy that the result provided in Proposition 1 is inspired from Benabbou et al. (2020), where a bound on the price of diversity in assignment problems is established. In our setup related to committee selection, we offer a similar result, even though the models and tasks are entirely different.[7] This result also generalizes the finding of Bredereck et al. (2017)

[7]In Benabbou et al. (2020), each proportion α_{pq} is the maximal proportion of agents of type p to match with items of group q, as required by the diversity constraint.

for (weakly) separable committee scoring rules. In the case of a binary attribute, if the diversity constraint demands a balanced committee (i.e., $q_1 = q_2 = k/2$), the required proportion for each class is $1/2$, and thus $\frac{1}{\min\limits_{j\in[l]}\alpha_j} = 2$. However, Proposition 1 provides the maximum price of diversity caused by a given diversity constraint q on any (weakly) separable committee scoring rule F, regardless of the preference profile; that is, for any diversity constraint q and any (weakly) separable committee scoring rule F, it holds that

$$\max_{\succ\in\mathcal{P}^n} POD(F, \succ, q) = \frac{1}{\min\limits_{j\in[l]}\alpha_j}. \tag{6}$$

Therefore, we can derive the maximum price of diversity for every (weakly) separable committee scoring rule, irrespective of the preference profile and the diversity constraint. This is stated in the following corollary.

Corollary 1 *For any (weakly) separable committee scoring rule F and any committee size k, the maximum price of diversity when selecting a diverse committee of size k is k. That is, for any (weakly) separable committee scoring rule F, we have*

$$\max_{q\in\mathbb{N}^l}\Big(\max_{\succ\in\mathcal{P}^n} POD(F, \succ, q)\Big) = k.$$

Proof. From Proposition 1 and Equation (6), we have

$$\max_{\succ\in\mathcal{P}^n} POD(F, \succ, q) = \frac{1}{\min\limits_{j\in[l]}\alpha_j} = \frac{1}{\frac{1}{k}\times\min\limits_{j\in[l]}q_j} \leq k,$$

since $q_j \geq 1$, for all $j \in [l]$. ■

The result presented in Corollary 1 is applicable to the entire class of (weakly) separable committee scoring rules, without considering the specific impact of the rule being used. Nevertheless, it is essential to understand the potential worst-case loss of excellence when selecting a scoring rule. Hence, it would be valuable to examine the maximum price to be paid based on the voting procedure. This analysis could help differentiate between various rules and identify which rule is more susceptible to the diversity constraint.

Let F be a (weakly) separable committee scoring rule defined through the scoring vector $w = (w_1, \ldots, w_m)$. Recall that for any preference profile $\succ$ and any committee $W \in 2^A_k$, the top scored candidate from W according to the profile $\succ$ has an individual score of at least $s(\succ, W)/k$, which is the average of the scores collected by all of the candidates in W. Let $T \in F(\succ, k)$ be the optimal committee with respect to the profile $\succ$ and R be the diverse optimal committee obtained from T when a diversity constraint is imposed. Recall that the set $T \cap R$ is the set of candidates selected in T that still belong to R. Since the diversity constraint is assumed to require at least one candidate from each class, then it holds that $|T \cap R| \geq 1$, where the instance $|T \cap R| = 1$ gives the maximum price induced by the diversity constraint, and in this case the single candidate belonging to $T \cap R$ is necessarily the top scored candidate across the preference profile. This scenario occurs when all the members of T are of the same type, otherwise $T \cap R$ would contain at least two candidates. Indeed, since $q_j \geq 1$ for all $j \in [l]$, then if the optimal committee T contains candidates from two different types, then the best candidate from each of the two types selected in T should still belong to R, and $T \cap R$ would contain at least two candidates.

Proposition 2 below establishes the maximum value of the price of diversity, contingent on the scoring vector that defines the rule.

Proposition 2 *Let F be a (weakly) separable scoring rule defined by a scoring vector $w = (w_1, \ldots, w_m)$. Then for any preference profile $\succ$, and any diversity constraint q, we have*

$$POD(F, \succ, q) \leq \frac{\overline{w}_k}{\frac{\overline{w}_k}{k} + \displaystyle\sum_{t=m-k+2}^{m} w_t},$$

where $\overline{w}_k = \sum_{t=1}^{k} w_t$.

Proof. Let F be a weakly separable scoring rule defined by a scoring vector $w = (w_1, \ldots, w_m)$, $\succ$ be a preference profile, and q be a diversity constraint. Let $T \in F(\succ, k)$ be the optimal committee under F and $R \in F(\succ, k, q)$ be the diverse optimal committee obtained from T. Obviously, we have

$$s(\succ, T) \leq n \sum_{t=1}^{k} w_t = n\overline{w}_k, \tag{7}$$

since the right-hand part of Equation (7) is the maximum score of any committee with respect to the scoring vector w.

Furthermore, we have $R = (T \cap R) \cup (R \setminus T)$, and since F is (weakly) separable, we have $s(\succ, R) = s(\succ, R \cap T) + s(\succ, R \setminus T)$. Recall that the maximum value of the price of diversity is reached when $|R \cap T| = 1$, which means that only one candidate from T still belongs to R. Moreover, as mentioned previously, this unique candidate is necessarily the top scored candidate across the preference profile $\succ$, then it holds that $s(\succ, R \cap T) \geq s(\succ, T)/k$.

On the other hand, the remaining $k - 1$ alternatives from $R \setminus T$ receive a total score of at least $n \displaystyle\sum_{t=m-k+2}^{m} w_t$, which is the lowest score of any committee of size $k - 1$. We therefore deduce that

$$s(\succ, R) \geq \frac{s(\succ, T)}{k} + n \sum_{t=m-k+2}^{m} w_t. \tag{8}$$

From Equation (8), it follows that

$$POD(F, \succ, q) \leq \frac{s(\succ, T)}{\frac{s(\succ, T)}{k} + n \displaystyle\sum_{t=m-k+2}^{m} w_t}. \tag{9}$$

The right-hand part of Equation (9) is an increasing function of $s(\succ, T)$; thus, by merging

Equations (7) and (9), we have

$$POD(F, \succ, q) \leq \frac{\overline{w}_k}{\frac{\overline{w}_k}{k} + \sum_{t=m-k+2}^{m} w_t}.$$

■

From Proposition 2, we can easily derive the following corollary.

Corollary 2 *For a given (weakly) separable committee scoring rule F defined with a scoring vector $w = (w_1, \cdots, w_m)$, choosing a size-k diverse committee using F involves a maximum price of diversity given by*

$$POD_{max}(F) = \max_{q \in \mathbb{N}^l} \Big(\max_{\succ \in \mathcal{P}^n} POD(F, \succ, q) \Big) = \frac{\overline{w}_k}{\frac{\overline{w}_k}{k} + \sum_{t=m-k+2}^{m} w_t}.$$

Obviously, it can be checked that the maximum price of diversity associated to a given rule and provided in Corollary 2 cannot exceed the previous one given in Corollary 1 for the entire class of (weakly) separable scoring rules, since $POD_{max}(F) \leq k$ for each rule F. We are now able to calculate the maximum price of diversity for several (weakly) separable committee scoring rules, representing the worst loss of excellence according to each rule when a diversity constraint is incorporated into the committee selection process. By some straightforward calculations, we can obtain the maximum price of diversity for each of the four well-known rules presented in Section 2. This is stated in the following corollary.

Corollary 3 *The maximum prices of diversity for the four well-known (weakly) separable scoring rules are given by:*

$$POD_{max}(k - Plurality) = k;$$

$$POD_{max}(Bloc) = k;$$

$$POD_{max}(k - Borda) = \frac{k\ (2m - k - 1)}{(2m - k - 1) + (k - 1)(k - 2)};$$

$$POD_{max}(k - AP) = \frac{k}{k - 1}.$$

We can observe that the upper bound given by Corollary 1 for the whole class of (weakly) separable scoring rules is not always reached by all the rules. The k-Borda and k-Antiplurality rules illustrate this point. We can then compare these four rules based on the maximum price of diversity induced by the diversity constraint.

Remark 1 *The comparison of the four rules studied above according to the maximum price of diversity is given by*

$$POD_{max}(k - Plurality) = POD_{max}(Bloc) \geq POD_{max}(k - Borda) \geq POD_{max}(k - AP).$$

Based on the above comparison, it can be deduced that the k-Plurality and Bloc rules are the most sensitive to the diversity constraint, while the k-Antiplurality rule is the least sensitive. This is primarily due to the scoring vector $w^{AP} = (1, \cdots, 1, 0)$ associated with the k-Antiplurality rule, where, for example, replacing a candidate ranked first by all voters with a candidate ranked at position $(m-1)$ by all voters does not change the price of diversity for this rule. Moreover, it can be observed that the maximum price of diversity for the k-Plurality and Bloc rules increases with the committee size, while for the k-Antiplurality rule, it decreases with the committee size. On the other hand, for the k-Borda rule, the maximum price of diversity increases and reaches a peak for a certain value of k (depending on the number of candidates). Thus, $POD_{max}(k - Borda)$ increases with the committee size until a certain point, after which it decreases.[8] This dynamic is explained by the fact that the Borda rule assigns points based on the ranking of candidates, and as the committee becomes larger, the total committee score increases. However, at a certain point, the individual scores of candidates start to offset each other, leading to a decrease in $POD_{max}(k - Borda)$. It is noteworthy that the maximum price of diversity is the same for all the aforementioned rules when considering the smallest committee size, $k = 2$. The aforementioned observed behaviors highlight the complexity of the impact of diversity on the committee's quality, especially for more sophisticated voting rules like the Borda rule.

3.2 The impact of types performances on the price of diversity

Recall that Proposition 1 provides the maximum price of diversity caused by a specific diversity constraint on any (weakly) separable committee scoring rule. In this section, we assert that a more accurate estimation of the price of diversity can be achieved by solely considering the performances of candidates of each type. In other words, for a given preference profile $\succ$ and any rule F, we can determine the maximum price of diversity induced by any diversity constraint q, potentially offering a more refined estimate than the one provided in Proposition 1.

Formally, consider F as a (weakly) separable committee scoring rule, $\succ$ as a preference profile, k as the committee size, and $T \in F(\succ, k)$ as the selected committee without any diversity constraint. Define $[l]_T = j \in [l] : T \cap A_j \neq \emptyset$ as the set of types represented by at least one candidate in T. For each type $j \in [l]_T$, let $\lambda_j(F, \succ)$ denote the proportion of scores of candidates with type j in T. This is expressed as:

$$\lambda_j(F, \succ) = \frac{s(\succ, T^j)}{s(\succ, T)}. \tag{10}$$

In this manner, the proportion λ_j associated with each type j aims to capture the strength of candidates of type j in T. We denote by $\lambda(F, \succ)$ the proportion of scores of the weakest

[8] The maximum price of diversity is reached at $k* = \frac{-2m-1+\sqrt{8m^3-16m^2+2m+6}}{2m-5}$ where it is equal to $\frac{\left(-4m^2+10m-6+\sqrt{8m^3-16m^2+2m+6}\right)\left(-2m-1+\sqrt{8m^3-16m^2+2m+6}\right)}{-16m^3+32m^2-4m-12+12m\sqrt{8m^3-16m^2+2m+6}-18\sqrt{8m^3-16m^2+2m+6}}$. In fact, $k*$ and the maximum price are increasing functions of m.

type in T, which is expressed as:

$$\lambda(F, \succ) = \min_{j \in [l]_T} \lambda_j(F, \succ). \tag{11}$$

The following proposition provides an alternative bound on the price of diversity based on the performances of multiple types, potentially offering a better estimate than the one given by Proposition 1.

Proposition 3 *Let F be a (weakly) separable committee scoring rule and $\succ$ a preference profile. For any diversity constraint q, we have*

$$POD(F, \succ, q) \le \frac{k\,(k-l+1)}{l\,\lambda(F, \succ)}.$$

Proof. Let $T \in F(\succ, k)$ be the selected committee without any diversity constraint. Let q be an arbitrary diversity constraint and R be the diverse optimal committee obtained from T. Recall that if T is a committee that already satisfies the diversity constraint q, then $R = T$, $POD(F, \succ, q) = 1$, and the result holds. We are then interested in the case where T is not a q diverse committee. For any $j \in [l]$, the maximum number of candidates that can be required from class A_j is $(k - l + 1)$, since at least one candidate is required from each of the other $(l-1)$ classes. Then, for all $j \in [l]$, it holds that

$$\alpha_j \le \frac{k-l+1}{k}.$$

Therefore, it holds that

$$\frac{\lambda_j(F, \succ)}{\alpha_j} \ge \frac{k-l+1}{k}\lambda_j(F, \succ) \ge \frac{k}{k-l+1}\lambda(F, \succ),$$

which implies that

$$s(\succ, T^j) \ge \alpha_j s(\succ, T)\frac{k}{k-l+1}\lambda(F, \succ). \tag{12}$$

On the other hand, it follows from Equation (5) that

$$s(\succ, R^j) \ge \frac{q_j}{k} s(\succ, T^j) \ge \frac{1}{k} s(\succ, T^j), \tag{13}$$

since $q_j \ge 1$ whatever the imposed diversity constraint q. It follows from Equations (12) and (13) that

$$s(\succ, R^j) \ge \alpha_j \frac{s(\succ, T)}{k-l+1}\lambda(F, \succ).$$

Summing over all types, we obtain

$$s(\succ, R) \ge \frac{s(\succ, T)}{k-l+1}\lambda(F, \succ)\sum_{j=1}^{l}\alpha_j \ge \frac{s(\succ, T)}{k-l+1}\lambda(F, \succ) \times \frac{l}{k}.$$

As a result, we get,

$$POD(F, \succ) = \frac{s(\succ, T)}{s(\succ, R)} \leq \frac{k(k-l+1)}{l\lambda(F, \succ)}.$$

■

The following example shows that the bound on the price of diversity provided by Proposition 3 can be significantly lower than that provided by Proposition 1.

Example 5 *Consider the set of six candidates $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ partitioned into four classes as $A = \{a_1, a_2\} \cup \{a_3, a_4\} \cup \{a_5\} \cup \{a_6\}$, and the following preference profile*

$$\succ = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_1 & a_4 & a_3 \\ a_3 & a_4 & a_2 & a_1 \\ a_4 & a_3 & a_1 & a_2 \\ a_5 & a_5 & a_5 & a_5 \\ a_6 & a_6 & a_6 & a_6 \end{bmatrix}$$

Let $k = 4$ be the committee size and assume that the rule under consideration is the 4-Plurality rule. If we have the information that diversity constraint to be applied is $q = (1, 1, 1, 1)$, then Proposition 1 says that the maximum price of diversity is 4. However, if we focus only on the preferences of voters giving the scores of candidates, without knowing the diversity constraint, the optimal committee is $T = \{a_1, a_2, a_3, a_4\}$ and $[j]_T = \{1, 2\}$ since the only types represented in T are the types 1 and 2. In this case, it can be checked that $\lambda_1(F, \succ) = \lambda_2(F, \succ) = 1/2$ and we can deduce from Proposition 5 that the maximum price of diversity would be 2.

Roughly speaking, knowing the performances of the candidates (and then of different types) given by the profile can sometimes give a better estimation of the cost of diversity than knowing the diversity constraint to be applied.

3.3 Individual price of diversity

We commence this section by reiterating that the price of diversity aims to gauge the influence of the diversity constraint on the committee's performance. However, the diversity constraint has the consequence of favoring certain candidates while penalizing others. This occurs as some candidates yield their positions to weaker candidates due to the diversity constraint. Consequently, we posit that an alternative approach to assessing the cost of the diversity constraint should involve measuring the "harm" suffered by penalized candidates. To address this, we introduce the "individual price of diversity," which is contingent on how candidates experience the impact of the diversity constraint."

If the optimal committee T fails to meet the diversity constraint q, then we can quantify the harm suffered by each candidate from T, specifically, each candidate excluded due to the diversity constraint. To achieve this, for each $a \in T$, we set

$$H_a = \{x \in T : s(\succ, x) \geq s(\succ, a)\}, \tag{14}$$

as the set of candidates in T that are at least as good as a. Let R be the diverse optimal committee obtained from T and $R_a = H_a \cap R$. Intuitively, R_a is the set of candidates that are at least as good as a and that are actually selected in the diverse optimal committee R. In other words,

$$R_a = \{x \in R : s(\succ, x) \geq s(\succ, a)\}. \tag{15}$$

The harm suffered by candidate a with respect to the diversity constraint q, denoted by $\chi_a(F, \succ, q)$, can be evaluated as the ratio of the aggregated score of all candidates that are at least as good as a to the aggregated score of those candidates that are actually selected in the diverse optimal committee R, i.e.,

$$\chi_a(F, \succ, q) = \frac{\sum_{x \in H_a} s(\succ, x)}{\sum_{x \in R_a} s(\succ, x)}. \tag{16}$$

In essence, $\chi_a(F, \succ, q)$ aims to capture how the total performance of candidates higher scored than a has been weakened by the diversity constraint q.

Definition 9 *Let F be a (weakly) separable committee scoring rule, $\succ$ be a preference profile, and q be a diversity constraint. The individual price of diversity induced by q on F with respect to the profile $\succ$ is defined by*

$$IPOD(F, \succ, q) = \max_{a \in T} \chi_a(F, \succ, q). \tag{17}$$

A comparable methodology has been developed by Yang et al. (2019) to investigate "in-group fairness" in committee selection with diversity constraints. Indeed, the candidates chosen in the diverse optimal committee may not necessarily be the top candidates overall across the profile, and this unfairness is not uniformly experienced by all types. The authors introduced a similar parameter and explored this phenomenon through experimental analysis. Note that for any candidate $a \in T$, the lowest value of $\chi_a(F, \succ, q)$ is 1, corresponding to the case where a, as well as all candidates that are at least as good as a, still belong to R; that is, when $H_a = R_a$. Moreover, the harm $\chi_a(F, \succ, q)$ suffered by candidate a increases as the number of candidates at least as good as a and selected in R decreases. Note also that we can only focus on the harm suffered by the candidates belonging to $T \setminus R$, i.e. the candidates belonging to the optimal committee but not to the selected diverse optimal committee with the diversity constraint, since any candidate belonging to both committees is not affected by the diversity constraint at all. Incidentally, since the individual price of diversity is the maximum harm suffered by the candidates from T, it can be checked that this maximum harm cannot be that of a candidate that belongs to both T and R.

Example 6 *Examine the identical profile, committee size, and diversity constraint as presented in Example 3. Under the 3-Borda rule, we have $s(\succ, a_1) = 20, s(\succ, a_2) = 16, s(\succ, a_3) = 12, s(\succ, a_4) = 8, s(\succ, a_5) = 4, s(\succ, a_6) = 0$, and the optimal committee is $T = \{a_1, a_2, a_3\}$.*

However, the diverse optimal committee is $R = \{a_1, a_4, a_6\}$ and the set of candidates offended by the diversity constraint is $T \setminus R = \{a_2, a_3\}$. Furthermore, the harms suffered by a_2 and a_3 respectively are $\chi_{a_2}(3-Borda, \succ, q) = \frac{s(\succ,\{a_1,a_2\})}{s(\succ,\{a_1\})} = \frac{36}{20}$ and $\chi_{a_3}(3-Borda, \succ, q) = \frac{s(\succ,\{a_1,a_2,a_3\})}{s(\succ,\{a_1\})} = \frac{48}{20} = \frac{12}{5}$. Thus the individual price of diversity induced by the diversity constraint q on the 3-Borda rule according to the profile $\succ$ is $IPOD(3-Borda, \succ, q) = \chi_{a_3}(3-Borda, \succ, q) = \frac{12}{5}$.

The following proposition asserts that the individual price of diversity is consistently higher than the price of diversity for any (weakly) separable committee scoring rule, regardless of the preference profile.

Proposition 4 *For any (weakly) separable committee scoring rule F, any preference profile $\succ$, and any diversity constraint q, we have*

$$POD(F, \succ, q) \leq IPOD(F, \succ, q).$$

Proof. Let $\succ$ be a preference profile and $T \in F(\succ, k)$ be the optimal committee with respect to $\succ$. Without loss of generality, we can write $T = \{a_1, \cdots, a_k\}$ with $s(\succ, a_t) \geq s(\succ, a_{t+1})$, for all $t \in [k-1]$. It follows that

$$POD(F, \succ, q) = \frac{s(\succ, T)}{s(\succ, R)} \leq \frac{s(\succ, H_{a_k})}{s(\succ, R_{a_k})} = \chi_{a_k}(F, \succ, q) \leq \max_{a \in T} \chi_a(F, \succ, q) = IPOD(F, \succ, q),$$

since $H_{a_k} = T$ and $R_{a_k} \subseteq R$. ■

Proposition 3 clearly shows that the price of diversity is more meaningful when it is measured individually across candidates, and the following example further illustrates this point.

Example 7 *Consider the set of candidates $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ partitioned as $A = A_1 \cup A_2 \cup A_3$ with $A_1 = \{a_1, a_2, a_3\}, A_2 = \{a_4, a_5\}$, and $A_3 = \{a_6\}$. Assume that the committee size is $k = 3$, the diversity constraint is $q = (1, 1, 1)$, and consider the following preference profile with three voters*

$$\succ = \begin{bmatrix} a_1 & a_2 & a_2 \\ a_3 & a_1 & a_3 \\ a_2 & a_3 & a_1 \\ a_4 & a_4 & a_4 \\ a_5 & a_5 & a_5 \\ a_6 & a_6 & a_6 \end{bmatrix}$$

Under the 3-Borda rule, the optimal committee is $T = \{a_1, a_2, a_3\}$ with the maximum score of 36. Moreover, candidates a_2 and a_3 will be replaced by a_4 and a_5 due to the diversity constraint q, and the individual price of diversity is $IPOD(F, \succ, q) = \chi_{a_3} = \frac{s(\succ,T)}{s(\succ,a_2)} = \frac{36}{13}$. Thus, we have $POD(F, \succ, q) = \frac{s(\succ,T)}{s(\succ,R)} = \frac{36}{19} \leq IPOD(F, \succ, q)$.

However, we it can be checked that even if the individual price of diversity in Example 7 above is greater than the price of diversity defined in Section 3.1, it remains lower than the bound provided by Proposition 1, since $IPOD(F, \succ, q) = \frac{36}{13} < 3 = \frac{1}{\underset{j \in [l]}{\min} \alpha_j}$. The next proposition actually shows that the bound on the price of diversity given by Proposition 1 is very tight, as even the individual price of diversity cannot exceed this bound.

Proposition 5 *For any (weakly) separable committee scoring rule F, any preference profile $\succ$, and any diversity constraint q, we have*

$$IPOD(F, \succ, q) \leq \frac{1}{\underset{j \in [l]}{\min} \alpha_j}.$$

Proof. Let k and $q = (q_1, \cdots, q_l)$ be the given committee size and diversity constraint, respectively. Let F be a (weakly) separable committee scoring rule, $\succ$ be a preference profile, and $T \in F(\succ, k)$ be the optimal committee under F with respect to $\succ$.

- If T is q-diverse, then $T = R$ and $H_a = R_a$ for all $a \in T$, which means that $\chi_a(F, \succ, q) = 1$ for all $a \in T$ and the result holds.

- If T is not q-diverse, consider the candidate $a \in T$ such that $\chi_a(F, \succ, q) = \underset{x \in T}{\max} \chi_x(F, \succ, q)$. It cannot be that $H_a = R_a$, as in this case, we would have $\chi_a(F, \succ, q) = 1$, implying that $\chi_x(F, \succ, q) = 1$ for all $x \in T$, which means that $T = R$, an impossible scenario. Therefore, it follows that $H_a \neq R_a$, and, consequently, there is a candidate $a_0 \in H_a$ such that $a_0 \notin R_a$. Recall that R_a is the set of candidates from H_a that still belong to R (where R is the diverse optimal committee obtained from T). The fact that $a_0 \notin R_a$ means that the quota $q_{j(a_0)}$ has already been reached in R_a, and then $|R_a| \geq q_{j(a_0)} \geq \underset{j \in [l]}{\min} q_j$. Consequently, it follows that

  $$IPOD(F, \succ, q) = \chi_a(F, \succ, q) = \frac{s(\succ, H_a)}{s(\succ, R_a)} = \frac{\sum\limits_{x \in H_a} s(\succ, x)}{\sum\limits_{x \in R_a} s(\succ, x)} \leq \frac{k s(\succ, a_1)}{\underset{j \in [l]}{\min} q_j s(\succ, a_k)} \leq \frac{k}{\underset{j \in [l]}{\min} q_j},$$

  since

  $$s(\succ, a_k) \leq s(\succ, x) \leq s(\succ, a_1),$$

  for all $x \in T$. As a result, we obtain

  $$IPOD(F, \succ, q) \leq \frac{k}{\underset{j \in [l]}{\min} q_j} = \frac{1}{\underset{j \in [l]}{\min} \alpha_j}.$$

■

It is evident that, in contrast to the price of diversity, the maximum value of the individual price of diversity does not provide a basis for distinguishing between various rules. This is due to a reasoning analogous to that presented in Proposition 2, which results in $IPOD_{max}(F) = k$ for any rule F.

4 Concluding comments

In summary, our focus in this study was on assessing the cost of diversity in committee selection under (weakly) separable committee scoring rules. This scenario is common in various real-world problems, and the introduced metrics, namely the price of diversity and individual price of diversity, aim to quantify the loss of excellence when a diverse committee is required. We presented an upper bound for both metrics, highlighting their worst-case scenarios, and demonstrated the utility of the maximum price of diversity in distinguishing between different rules. Additionally, we illustrated that an evaluation based solely on the candidates' performances derived from the preference profile can provide a more accurate estimation of the maximum price of diversity.

Our study opens the door to several important avenues for future research. One such direction is to explore a "qualitative" approach for measuring the price of diversity. Utilizing the score function alone may not provide a comprehensive understanding of the loss of "quality" among candidates. For instance, Corollary 3 reveals that the k-Antiplurality rule, among the four (weakly) separable committee scoring rules in the literature, yields the lowest maximum price of diversity. This outcome is largely influenced by its associated scoring vector $w^{AP} = (1, \cdots, 1, 0)$. In cases where replacing a candidate ranked first by all voters with one ranked at position $(m-1)$ by all voters does not affect the price of diversity, it raises concerns about fairness. Considering alternative definitions of the price of diversity based on candidates' positions in voters' preferences, such as using metric rationalization, could offer insights into the closeness of the selected committee to voters' preferences. We refer the reader to, for instance, Andjiga et al. (2014) and Elkind et al. (2015), among many others. Additionally, employing a probabilistic approach may help assess the frequency with which various rules approach the bounds of the price of diversity. Indeed, if a voting rule results in reaching the maximum bound of a given price of diversity but with a low probability, we may question the significance of such determination. While obtaining bounds is commendable, gaining insight into the probability of its achievement is even more valuable. We refer the reader to, for instance, Diss and Merlin (2021) and Gehrlein and Lepelley (2011, 2017), among others. Another challenging task is extending this analysis to other categories of committee selection rules, particularly those handling approval preferences as in Brams (1990), Kagita et al. (2021), Kilgour (2010), Kilgour and Marshall (2012), among others.

References

Andjiga, N.G., Mekuko, A. Y., and Moyouwou, I. (2014). Metric rationalization of social welfare functions. *Mathematical Social Sciences*, 72:14–23.

Aziz, H. (2019). A rule for committee selection with soft diversity constraints. *Group Decision and Negotiation*, 28(6):1193–1200.

Balinski, M.L. and Young, H.P (1994) Apportionment. *Handbooks in operations research and management science*, 6:529–560.

Bei, X., Liu, S., Poon, C. K., and Wang, H. (2022). Candidate selections with proportional fairness constraints. *Autonomous Agents and Multi-Agent Systems*, 36(1):1–32.

Benabbou, N. and Chakraborty, M. and Ho, X.V. and Sliwinski, J. and Zick, Y. (2020). The price of quota-based diversity in assignment problems. *ACM Transactions on Economics and Computation (TEAC)*, 8(3):1–32.

Bock H., Dayb W., and Mcmorris F.(1998). Consensus rules for committee elections. *Mathematical Social Sciences*, 35(3):219–232.

Brams, S. J. (1990). Constrained approval voting: A voting system to elect a governing board [with comment]. *Interfaces*, 20(5):67–80.

Brams S. J., Kilgour D. M., and Potthoff R.F.(2019). Multiwinner approval voting,an apportionment approach. *Public Choice*, 178(1-2):67–93.

Bredereck, R. and Faliszewski, P. and Igarashi, A. and Lackner, M. and Skowron, P. (2017). Multiwinner elections with diversity constraints. *CoRR abs/1712.02712*, https://arxiv.org/pdf/1711.06527.pdf.

Celis, L. E., Huang, L., and Vishnoi, N. K. (2017). Multiwinner voting with fairness constraints. *arXiv preprint*, https://arxiv.org/pdf/1710.10057.pdf.

Chamberlin, J. R. and Courant, P. N. (1983). Representative deliberations and representative decisions: Proportional representation and the Borda rule. *The American Political Science Review*, 77(3):718–733.

Diss, M., and Doghmi, A. (2016). Multi-winner scoring election methods, Condorcet consistency and paradoxes. *Public Choice*, 169:97–116.

Diss, M., Kamwa, E., and Tlidi, A. (2020). On some k-scoring rules for committee elections: agreement and Condorcet Principle. *Revue d'Economie Politique*, 130(5):699–725.

Diss, M. and Merlin, V. (2021). *Evaluating Voting Systems with Probability Models, Essays by and in honor of William V. Gehrlein and Dominique Lepelley.* Studies in Choice and Welfare, Springer.

Do, V., Atif, J., Lang, J., and Usunier, N. (2021). Online selection of diverse committees. *arXiv preprint*, https://arxiv.org/pdf/2105.09295.pdf.

Elkind, E. and Faliszewski, P. and Skowron, P. and Slinko, A. (2017). Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632.

Elkind, E., Faliszewski, P., and Slinko, A. (2015). Distance rationalization of voting rules. *Social Choice and Welfare*, 45(2):345–377.

Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. (2018). Multiwinner analogues of the Plurality rule: Axiomatic and algorithmic perspectives. *Social Choice and Welfare*, 51(3):513–550.

Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. (2019). Committee scoring rules: Axiomatic characterization and hierarchy. *ACM Transactions on Economics and Computation*, 7(1):1–39.

Gehrlein, W.V. and Lepelley, D. (2011). *Voting paradoxes and group coherence.* Studies in Choice and Welfare, Springer.

Gehrlein, W.V. and Lepelley, D. (2017). *Elections, Voting Rules and Paradoxical Outcomes.* Studies in Choice and Welfare, Springer.

Goto, M., Kojima, F., Kurata, R., Tamura, A., and Yokoo, M. (2017). Designing matching mechanisms under general distributional constraints. *American Economic Journal: Microeconomics*, 9(2):226–62.

Ianovski, E. (2022). Electing a committee with dominance constraints. *Annals of Operations Research*, 318:985–1000.

Izsak, R., Talmon, N., and Woeginger, G. (2018). Committee selection with intraclass and interclass synergies. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32.

Kagita, V. R., Pujari, A. K., Padmanabhan, V., Aziz, H., and Kumar, V. (2021). Committee selection using attribute approvals. In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 683–691.

Kamada, Y. and Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *The American Economic Review*, 105(1):67–99.

Kilgour, D. M. (2010). Approval balloting for multi-winner elections. In: *J. Laslier and M. R. Sanver (Eds.), Handbook on approval voting* (pp. 105-124), Heidelberg: Springer

Kilgour, D. M. and Marshall, E. (2012). Approval balloting for fixed-size committees. In: *Electoral Systems*, pp. 305–326. Springer.

Lang, J. and Skowron, P. (2018). Multi-attribute proportional representation. *Artificial Intelligence*, 263:74–106.

Lu, T. and Boutilier, C. (2011). Budgeted social choice: From consensus to personalized decision making. In: *Twenty-Second International Joint Conference on Artificial Intelligence.*

Relia, K. (2021). Dire committee: Diversity and representation constraints in multiwinner elections. *arXiv preprint*, https://arxiv.org/pdf/2107.07356.pdf.

Skowron, P., Faliszewski, P., and Slinko, A. (2019). Axiomatic characterization of committee scoring rules. *Journal of Economic Theory*, 180:244–273.

Thejaswi, S., Ordozgoiti, B., and Gionis, A. (2021). Diversity-aware k-median: Clustering with fair center representation. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 765–780. Springer.

Yang, K., Gkatzelis, V., and Stoyanovich, J. (2019). Balanced Ranking with Diversity Constraint. *arXiv preprint arXiv:1906.01747.*

Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory*, 9(1):43–52.

Young, H. P. (1975). Social choice scoring functions. *SIAM Journal on Applied Mathematics*, 28(4):824–838.

Funding

Mostapha Diss would like to acknowledge financial support from Région Bourgogne Franche-Comté within the program ANER 2021-2024 (project DSG).

Compliance with Ethical Standards

The authors declare that they have no conflict of interest.