

# Multigoal Committee Selection

Maciej Kocot<sup>1</sup>, Anna Kolonko<sup>1</sup>, Edith Elkind<sup>2</sup>, Piotr Faliszewski<sup>1\*</sup> and Nimrod Talmon<sup>3</sup>

<sup>1</sup>AGH University, Krakow, Poland

<sup>2</sup>University of Oxford, Oxford, UK

<sup>3</sup>Ben-Gurion University, Be'er Sheva, Israel

mkocot@op.pl, aniakol18@gmail.com, elkind@cs.ox.ac.uk, faliszew@agh.edu.pl, talmonn@bgu.ac.il

## Abstract

We study the problem of computing committees that perform well according to several different criteria, which are expressed as committee scoring rules. We analyze the computational complexity of computing such committees and provide an experimental evaluation of the compromise levels that can be achieved between several well-known rules, including  $k$ -Borda, SNTV, Bloc, and the Chamberlin–Courant rule.

## 1 Introduction

We study the problem of computing committees that perform well according to several different criteria, expressed as committee scoring rules [Elkind *et al.*, 2017b]. The following example illustrates our problem.

Let us consider a setting where a conference organizer plans a dinner for the participants and asks them to rank dishes from the catering menu. The organizer intends to use a multiwinner voting rule to aggregate these preferences and to form the catering order; at the dinner, the participants would be able to choose freely among the catered items. If the organizer used a rule focused on *individual excellence* of the candidates (such as  $k$ -Borda [Debord, 1992]), then many people (let us call them “the majority group”) would have lots of options to choose from, but some others might not be able to find anything acceptable. On the other hand, if the organizer used a rule focused on the *diversity* of the committee (such as the Chamberlin–Courant rule,  $\beta$ -CC [Chamberlin and Courant, 1983]), then everyone would be able to eat something, but perhaps many participants, possibly including those in the majority group, would only have one appealing option available to them. Naturally, the organizer would prefer some sort of compromise between these two extremes. In this paper we propose a principled way of finding such compromises.

Our main idea is that given a set of committee scoring rules—such as  $k$ -Borda and  $\beta$ -CC—we seek committees that achieve high scores (i.e., are judged highly) according to each of them. More formally, let  $\mathcal{R}$  be a collection of committee scoring rules (our approach is not limited to committee

scoring rules, but they suffice for our purposes and we restrict to them for the sake of focus).<sup>1</sup> In the  $\mathcal{R}$ -MULTIGOAL COMMITTEE problem we are given an election (i.e., a set of candidates and a collection of voters with their preferences), a target committee size, and lower bounds on the scores according to the rules from  $\mathcal{R}$ ; we ask if there is a committee that meets or exceeds the respective lower bound with respect to each of the rules. In particular, by solving a sequence of such problems, we can find a committee for which it is impossible to improve the score under any of the rules without impairing the scores under the other rules (i.e., we can find a committee on the Pareto frontier defined by our rules).

Using our framework, the conference organizer would be able to analyze the full spectrum of possible menus, ranging from the one focused on the majority group (chosen using  $k$ -Borda), through various levels of compromise where the members of the majority group would have fewer and fewer options to choose from, but smaller and smaller minorities would be catered for, down to the fully diverse solutions (computed according to  $\beta$ -CC). Multigoal committee selection framework would not relieve the organizer from the burden of choosing the final menu—which would require knowledge of the event—but it would allow him or her to make an informed decision.

### 1.1 Our Contribution

While we pay special attention to the  $\{k$ -Borda,  $\beta$ -CC}-MULTIGOAL COMMITTEE problem, we are also interested in combinations of other voting rules. Below we summarize our main contributions:

1. We show that if  $\mathcal{R}$  is a fixed-sized collection of *weakly separable* committee scoring rules (such as SNTV, Bloc, or  $k$ -Borda) and the scores assigned by the rules are polynomially bounded, then  $\mathcal{R}$ -MULTIGOAL COMMITTEE is in P. The two conditions—fixed cardinality of  $\mathcal{R}$  and polynomially-bounded scores—are necessary, as violating them may lead to NP-hardness.
2. As the  $\beta$ -CC rule is NP-hard to compute [Procaccia *et al.*, 2008; Lu and Boutilier, 2011], so is the  $\{k$ -Borda,  $\beta$ -CC}-MULTIGOAL COMMITTEE problem. We show

\*Contact Author

<sup>1</sup>Readers not familiar with committee scoring rules may think of  $\mathcal{R}$  as a collection of functions that evaluate committees.

that many of the known means of dealing with the computational hardness of  $\beta$ -CC, such as using approximation algorithms, using FPT algorithms, or considering restricted domains, are still applicable to the  $\{k$ -Borda,  $\beta$ -CC $\}$ -MULTIGOAL COMMITTEE problem, but sometimes require new insights.

3. We experimentally evaluate what sorts of committees appear on the Pareto curves corresponding to pairs of several well-known rules. E.g., we find that there often exist committees that achieve  $k$ -Borda and  $\beta$ -CC scores exceeding 90% of the optimal ones. This shows that at least in some settings the tension between individual excellence and diversity is lower than one might expect.

We omit some proofs, due to limited space.

## 1.2 Related Work

The problem of computing compromise committees between  $k$ -Borda and  $\beta$ -CC was recently introduced by Faliszewski et al. [2017], who defined a parametrized family of rules, with extreme values of the parameter corresponding, respectively, to  $k$ -Borda and  $\beta$ -CC; thus, their definition was syntactic in nature. In contrast, we focus on the semantics of the computed committees and guarantee requested score values. A similar syntactic approach was recently used [Faliszewski and Talmon, 2018] to define a compromise between  $\beta$ -CC and the Monroe rule.

Our work is also inspired by that of Lackner and Skowron [2019], who considered two important approval-based multiwinner rules (approval voting, which models individual excellence, and the approval-based Chamberlin–Courant rule, which models diversity), and investigated whether other approval-based voting rules offer good performance when viewed as approximation algorithms for these two rules. The difference between our work and theirs is that they focused on performance of specific voting rules with respect to their criteria, whereas we seek committees that perform optimally. Lackner and Skowron considered the approval setting because they were interested in proportional representation of the voters and the approval setting is more convenient in this context (in particular, due to the work on justified representation [Aziz et al., 2017; Sánchez-Fernández et al., 2017]). We use the ordinal setting because it offers a better model of individual excellence (diversity seems to be modeled equally well in both settings).

Finally, we mention that the idea of studying Pareto optimality for multiwinner elections was recently put forward by Aziz et al. [2016]. The difference between their work and ours is that they consider committees that are Pareto optimal with respect to the preferences of individual voters, whereas we consider committees that are Pareto optimal with respect to a set of given voting rules. Our paper is also similar in spirit to the work of Elkind and Erdélyi [2012] and Erdélyi, Hemaspaandra and Hemaspaandra [2014], who study the complexity of finding election attacks (such as manipulation, control, or bribery) that perform well against any (single-winner) voting rule from a given family of rules. We discuss further related work throughout the paper, in the relevant contexts.

## 2 Preliminaries

Let  $\mathbb{R}_+$  denote the set of non-negative real numbers. For an integer  $t$ , we write  $[t]$  to mean  $\{1, \dots, t\}$ . For a logical expression  $F$ , by  $[F]$  we mean 1 if  $F$  is true and 0 otherwise.

**Elections.** An election  $E = (C, V)$  consists of a set of candidates  $C = \{c_1, \dots, c_m\}$  and a collection of voters  $V = (v_1, \dots, v_n)$ , where each voter ranks all the candidates from the most to the least appealing one in his or her preference order. A multiwinner voting rule  $R$  is a function that takes an election  $E$  and an integer  $k$  and returns a non-empty family  $R(E, k)$  of size- $k$  sets of candidates, i.e., of size- $k$  committees that win the election. (We disregard tie-breaking issues, but point to the works of Obraztsova et al. [2011] and Obraztsova and Elkind [2011] for some related discussion.) We focus on committee scoring rules, formally introduced by Elkind et al. [2017b] and studied axiomatically by Skowron et al. [2019] and Faliszewski et al. [2019] (see also the works of Aziz et al. [2017] and Lackner and Skowron [2018] for their analogues in the approval setting). Among committee scoring rules, we restrict our attention to (weakly) separable and representation-focused rules (defined below).

**Single-winner scoring functions.** Consider an election  $E = (C, V)$  with  $m$  candidates. We write  $\text{pos}_v(c)$  to denote the position of candidate  $c$  in the preference order of voter  $v$  (the top-ranked candidate has position 1 and the bottom-ranked one has position  $m$ ). A single-winner scoring function  $\gamma_m: [m] \rightarrow \mathbb{R}_+$  is a function that maps each possible position to a score value. We define the  $\gamma_m$ -score of a candidate  $c \in C$  to be  $\gamma_m\text{-score}_E(c) = \sum_{v \in V} \gamma_m(\text{pos}_v(c))$ , i.e., to be the sum of the scores associated with the positions in which  $c$  is ranked. From our point of view, the most important single-winner scoring functions are the Borda functions, defined as  $\beta_m(i) = m - i$ , and the  $t$ -Approval functions, defined for each positive integer  $t$  as  $\alpha_t(i) = [i \leq t]$  ( $\alpha_1$  is the Plurality scoring function).

**(Weakly) separable rules.** Let  $\gamma = (\gamma_{m,k})_{k \leq m}$  be a family of single-winner scoring functions, with one function for each possible number of candidates  $m$  and each possible committee size  $k$ . We define the weakly separable rule  $R_\gamma$  as follows. Let  $E = (C, V)$  be an election with  $m$  candidates and let  $k$  be the given committee size. For each size- $k$  committee  $S$  we define the  $R_\gamma$ -score of  $S$  to be the sum of the  $\gamma_{m,k}$ -scores of its members:  $R_\gamma\text{-score}_E(S) = \sum_{c \in S} \gamma_{m,k}\text{-score}_E(c)$ .  $R_\gamma(E, k)$  is the family of committees with the highest  $R_\gamma$ -score. If the functions  $\gamma_{m,k}$  do not depend on  $k$  (i.e., for each  $m$  we have  $\gamma_{m,1} = \dots = \gamma_{m,m}$ ), then we refer to  $R_\gamma$  as *separable*. We are particularly interested in the following (weakly) separable rules:

1. the  $k$ -Borda rule, based on the Borda functions,
2. the  $t$ -Bloc rules, where for each positive integer  $t$ , the  $t$ -Bloc rule uses  $t$ -Approval functions (1-Bloc is known as the single non-transferable vote rule, SNTV), and
3. the Bloc rule, which uses  $k$ -Approval functions.

Intuitively, the difference between Bloc and the  $t$ -Bloc rules is that under the former each voter names members of his or her most preferred committee (his or her top- $k$  candidates) and

each of these candidates gets a single point from the voter; under the latter, each voter gives points only to his or her  $t$  top-ranked candidates, irrespective of the committee size. As a consequence,  $t$ -Bloc rules—as well as the  $k$ -Borda rule—are separable, and Bloc is only weakly separable.

**Example 1.** Consider an election  $E = (C, V)$  with  $C = \{a, b, c, d, e, f\}$  and the following four voters:

$$\begin{aligned} v_1: a \succ b \succ c \succ d \succ e \succ f & \quad v_3: e \succ a \succ f \succ d \succ b \succ c \\ v_2: a \succ b \succ c \succ d \succ e \succ f & \quad v_4: f \succ d \succ a \succ e \succ b \succ c \end{aligned}$$

We consider committees of size 3. According to  $k$ -Borda, committee  $\{a, b, d\}$  wins with score 37; according to SNTV, committee  $\{a, e, f\}$  wins with score 4; and according to Bloc there is a 3-way tie between committees  $\{a, b, c\}$ ,  $\{a, b, f\}$ , and  $\{a, c, f\}$ , all getting a score of 8.

**Representation-focused rules.** Representation-focused rules are variants of  $\beta$ -CC [Chamberlin and Courant, 1983]. Consider some election  $E = (C, V)$  and committee  $S$ . For each voter  $v \in V$ , we refer to the member of  $S$  that  $v$  ranks highest as  $v$ 's *representative* (in  $S$ ), and we denote this candidate as  $\text{rep}_v(S)$ . For a family  $\gamma = (\gamma_{m,k})_{k \leq m}$  of single-winner scoring functions, an election  $E = (C, V)$ , and committee  $S$ , we define the  $\gamma$ -CC score of  $S$  to be  $\sum_{v \in V} \gamma_{|C|,|S|}(\text{rep}_v(S))$  (i.e., each voter contributes exactly the score of his or her representative); the  $\gamma$ -CC rule outputs the committees with the highest  $\gamma$ -CC-score. We are mostly interested in the classic Chamberlin–Courant rule, based on the Borda scoring function and denoted  $\beta$ -CC, but we mention that SNTV is representation-focused as well.

**Example 2.** Consider again the election from Example 1. The  $\beta$ -CC winning committee is  $\{a, e, f\}$  with score 20 (indeed, this is also an SNTV winning committee; this is not surprising as for this committee all voters rank their representatives in the top position, and hence this committee wins under every representation-focused rule).

**Approximate committees.** Let  $R$  be a multiwinner voting rule that chooses committees with the highest scores (computed in some way) and let  $E$  be an election. For a number  $p \in [0, 1]$ , we say that a committee  $S$  is a  $p$ -approximate committee with respect to  $R$  if its  $R$ -score is at least a  $p$ -fraction of the score of an  $R$ -winning committee.

**The complexity of winner determination.** SNTV,  $k$ -Borda, Bloc, and all the  $t$ -Bloc rules are polynomial-time computable; i.e., for each of them there is a polynomial-time algorithm that given an election, a committee size, and a subset of candidates decides if this subset can be extended to a committee with at least a given score (indeed, this is the case for all weakly separable rules, whose scoring functions are polynomial-time computable). On the other hand, the problem of deciding if there is a  $\beta$ -CC committee with at least a given score is NP-hard [Procaccia *et al.*, 2008; Lu and Boutilier, 2011]. Fortunately, there are numerous ways of circumventing this result, including approximation algorithms [Lu and Boutilier, 2011; Skowron *et al.*, 2015a], FPT algorithms [Betzler *et al.*, 2013], polynomial-time algorithms for various restricted domains [Betzler *et al.*, 2013; Skowron *et al.*, 2015b; Peters, 2018], and heuristics [Faliszewski *et al.*, 2018].

### 3 The $\mathcal{R}$ -MULTIGOAL COMMITTEE Problem

We first define our problem formally, then consider its computational complexity, and finally show experimental results.

**Definition 1.** Let  $\mathcal{R} = \{R_1, \dots, R_d\}$  be a set of  $d$  committee scoring rules. In the  $\mathcal{R}$ -MULTIGOAL COMMITTEE problem we are given an election  $E = (C, V)$ , a committee size  $k$ , and a vector  $(t_1, \dots, t_d)$  of non-negative numbers; we ask if there is a committee  $S$  of size  $k$  such that for each  $i \in [d]$  the score of  $S$  with respect to  $R_i$  is at least  $t_i$ .

This definition has a few subtle points and requires some discussion. First, while the definition is phrased for general committee scoring rules, we restrict our attention to (weakly) separable and representation-focused rules. Second, we note that it is necessary to specify how the scores are computed for each of the rules from  $\mathcal{R}$ , because each rule can be defined in many different ways. Whenever we speak of the rules from Section 2, we use the scores from their definitions.

Third, we view the  $\mathcal{R}$ -MULTIGOAL COMMITTEE problem as a computational tool that allows us to analyze the (scores of the) committees that lie on the Pareto frontier stretched between the winning committees of the rules from  $\mathcal{R}$ . E.g., let  $\mathcal{R}$  contain the rules  $R_1, R_2$  and  $R_3$ . Using  $\mathcal{R}$ -MULTIGOAL COMMITTEE, we may look for a committee that achieves some scores  $t_1, t_2$ , and  $t_3$  according to these rules. If such a committee exists, then—e.g., using binary search—we can find the largest value  $t_1^*$  such that a committee with scores at least  $t_1^*, t_2$ , and  $t_3$  exists. We can do the same for the other rules to eventually obtain scores  $t_1^*, t_2^*$ , and  $t_3^*$  such that for each  $i$  we have  $t_i^* \geq t_i$  and there is no committee with at least as high scores for all the rules and a strictly higher score for at least one of them. Depending on the order in which we consider the rules and on the strategy for increasing the scores, we may reach different points of the Pareto frontier. In our experiments, we focus on pairs of rules (e.g.,  $k$ -Borda and  $\beta$ -CC or SNTV and Bloc) and explore the entire Pareto curve between their winning committees.

Fourth, it may not be obvious what values  $t_1, \dots, t_d$  to use at the beginning of the procedure from the previous paragraph. One possibility is to find the largest  $p \in [0, 1]$  for which there is a committee which is  $p$ -approximate for each of the rules in  $\mathcal{R}$ .

**Example 3.** Consider again the election from Example 1. Committee  $\{a, d, e\}$  has  $k$ -Borda score 36 and  $\beta$ -CC score 19. As the highest possible  $k$ -Borda score is 37 and the highest possible  $\beta$ -CC score is 20 (recall Example 1 and 2), our committee is 97%-approximate for  $k$ -Borda and 95%-approximate for  $\beta$ -CC.

### 4 The Case of (Weakly) Separable Rules

We say that a weakly separable rule, defined through a family  $\gamma = (\gamma_{m,k})_{k \leq m}$  of single-winner scoring functions, is *polynomially restricted* if the functions in  $\gamma$  are both polynomial-time computable and polynomially bounded (i.e., there is a polynomial  $p$  such that for each  $m, k$ , and  $i \in [m]$  we have  $\gamma_{m,k}(i) \leq p(m)$ ). It turns out that if  $\mathcal{R}$  is a fixed set of polynomially-restricted weakly separable rules, then  $\mathcal{R}$ -MULTIGOAL COMMITTEE can be solved in polynomial time.

**Theorem 1.** Let  $\mathcal{R} = \{R_1, \dots, R_d\}$  be a family of polynomially restricted, weakly separable rules. Then,  $\mathcal{R}$ -MULTIGOAL COMMITTEE is in P.

*Proof sketch.* Let  $\gamma^{(1)}, \dots, \gamma^{(d)}$  be the families of scoring functions used by the rules in  $\mathcal{R}$ . Consider an instance of  $\mathcal{R}$ -MULTIGOAL COMMITTEE with election  $E = (C, V)$ ,  $C = \{c_1, \dots, c_m\}$ , committee size  $k$ , and vector  $(t_1, \dots, t_d)$  of the lower bounds for scores under the respective rules. Now, for each rule  $R_i$  and each candidate  $c_j$ , let  $s_{i,j}$  be the  $\gamma^{(i)}$ -score of  $c_j$ . Our problem then boils down to deciding if there is a committee  $S$  of size  $k$  such that for each  $i \in [d]$  we have  $\sum_{c_j \in S} s_{i,j} \geq t_i$ . This problem can be solved in time  $O(m \cdot k \cdot \prod_{\ell=1}^d t_\ell)$  using dynamic-programming (the algorithm is analogous to the pseudo-polynomial algorithm for KNAPSACK, but extended to  $d$  dimensions; see, e.g., the overview of Kellerer et al. [2004]). Since each  $t_i$  is polynomially bounded in  $m$  (as our rules are polynomially restricted), this suffices to establish polynomial running time.  $\square$

Theorem 1 is very useful, as it allows us to solve  $\mathcal{R}$ -MULTIGOAL COMMITTEE for fixed combinations of rules such as  $k$ -Borda, SNTV, and Bloc. Yet, if the rules in  $\mathcal{R}$  are not polynomially restricted or the size of  $\mathcal{R}$  may grow with the instance size, then Theorem 1 does not apply and  $\mathcal{R}$ -MULTIGOAL COMMITTEE may become NP-hard. Below we show two such examples.

Let  $\delta = (\delta_{m,k})_{k \leq m}$  and  $\rho = (\rho_{m,k})_{k \leq m}$  be two families of single-winner scoring rules, defined for each even number  $2m$  of candidates and each committee size  $k$  as follows (the case of odd number of candidates is irrelevant for our proof):

$$\begin{aligned} \delta_{2m,k}(i) &= 2^{m-i-1}[i < m], \\ \rho_{2m,k}(i) &= 2^{m-2}[i \leq m] + 2^{2m-i-1}[m < i < 2m]. \end{aligned}$$

Intuitively, for  $i$  between 1 and  $m$  the values of  $\delta_{2m,k}$  change exponentially and the values of  $\delta_{2m,k}$  stay constant, and for  $i$  between  $m+1$  and  $2m$  this behavior is reversed. Both functions are polynomial-time computable but neither is polynomially bounded, and so the rules  $R_\delta$  and  $R_\rho$  are not polynomially restricted (but both are separable).

**Theorem 2.**  $\{R_\delta, R_\rho\}$ -MULTIGOAL COMMITTEE is NP-complete.

The proof of Theorem 2 (omitted) follows by a reduction from KNAPSACK (see, e.g., [Kellerer et al., 2004]). The main idea is to use the  $\delta_{2m,k}$ - and  $\rho_{2m,k}$ -scores to implement the (binary-encoded) weights and profits of the items from the KNAPSACK instance in a bit-by-bit manner.

Theorem 2 shows that indeed we need the rules in  $\mathcal{R}$  to be polynomially restricted for Theorem 1 to hold. Now we show that we also need  $\mathcal{R}$  to be fixed. For a positive integer  $\ell$ , let  $\ell$ -MULTIBLOC be the  $\{1\text{-Bloc}, \dots, \ell\text{-Bloc}\}$ -MULTIGOAL COMMITTEE problem. Intuitively, under  $\ell$ -MULTIBLOC we ask if there is a committee with a given number of first positions, a given number of top-two positions, and so on, up to the  $\ell$ -th position. Let MULTIBLOC be the  $\ell$ -MULTIBLOC problem where  $\ell$  is part of the input.

**Theorem 3.** MULTIBLOC is NP-complete.

## 5 $k$ -Borda and Chamberlin–Courant

We now focus on the  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE problem. The idea is to compute committees of individually very good candidates (as measured by their Borda scores) that cover the preferences of all the voters as well as possible (as measured by the Chamberlin–Courant scores). Since deciding if there is a committee with at least a given  $\beta$ -CC score is NP-hard, so is the  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE problem; thus, we seek algorithmic workarounds to bypass this issue.

### 5.1 Approximation Algorithms

Perhaps the simplest idea for an approximation algorithm is to combine the polynomial-time winner-determination algorithm of  $k$ -Borda with an approximation algorithm for  $\beta$ -CC, choosing some candidates according to the former and the remaining ones according to the latter. Formally, our algorithm proceeds as follows. Let  $E = (C, V)$  be an election with  $m$  candidates and  $n$  voters, let  $k$  be the desired committee size, and let  $k'$  and  $k''$  be two non-negative integers such that  $k = k' + k''$  (we may try all possible combinations of  $k'$  and  $k''$  and choose one that gives the best result). Let  $S'$  be a size- $k'$  committee of candidates with the highest Borda scores (i.e.,  $S' \in k\text{-Borda}(E, k')$ ), and let  $S''$  be a size- $k''$  committee computed using Algorithm P of Skowron et al. [2015a] (this is the best known polynomial-time approximation algorithm for  $\beta$ -CC). Our algorithm returns the committee  $S = S' \cup S''$  (if  $S'$  and  $S''$  are not disjoint then we complement the committee with sufficiently many arbitrary candidates). Carefully choosing the values of  $k'$  and  $k''$ , based on Skowron et al.'s analysis of Algorithm P, we obtain the following result.

**Proposition 1.** For each positive value  $\varepsilon$ , there is a polynomial-time algorithm that, given an election  $E = (C, V)$  and a positive integer  $k$ ,  $k \leq |C|$ , computes a size- $k$  committee  $S$  that is  $(1 - \varepsilon)$ -approximate with respect to both  $k$ -Borda and  $\beta$ -CC.

### 5.2 Fixed-Parameter Tractability

If approximate solutions are not acceptable in a given setting, then we may resort to fixed-parameter tractable (FPT) algorithms. For the parametrization by the committee size already the  $\beta$ -CC rule is W[2]-hard [Betzler et al., 2013], and so the same holds for  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE. Yet, for the parametrizations by the number of candidates and by the number of voters we do find FPT algorithms. The former one tries all possible committees; for the latter we need new ideas.

**Proposition 2.**  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE is in FPT for the parametrization by the number of candidates, but it is W[2]-hard for the parametrization by the committee size.

We now focus on the parametrization by the number of voters. The algorithm for  $\beta$ -CC, due to Betzler et al. [2013], proceeds by guessing a partition of the voters into groups, based on their representatives, and choosing the Borda winner of each group as its representative. We use a similar approach,

but in our case the representatives do not need to be the Borda winners, and we need to explicitly ensure that each group has a different representative.

**Theorem 4.** *There is an FPT algorithm for  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE parametrized by the number of voters.*

*Proof.* Consider an instance of  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE with election  $E = (C, V)$ , where  $C = \{c_1, \dots, c_m\}$  and  $V = (v_1, \dots, v_n)$ , committee size  $k$ , and lower bounds  $t_{kB}$  and  $t_{CC}$  on the  $k$ -Borda and  $\beta$ -CC scores, respectively. Before describing the algorithm, we introduce some notation: Let  $\prec$  be an order such that  $c_1 \prec c_2 \prec \dots \prec c_m$ . For a size- $r$  committee  $S = \{d_1, \dots, d_r\}$ , where  $d_1 \prec \dots \prec d_r$ , and a partition  $\mathcal{V}' = (V'_1, \dots, V'_{k'})$  of  $V$  into  $k'$  groups, we define  $\text{score}_{\mathcal{V}'}(S)$  to be  $\sum_{i=1}^r \beta\text{-score}_{(C, V_i)}(d_i)$ . I.e.,  $\text{score}_{\mathcal{V}'}(S)$  is the  $\beta$ -CC score of committee  $S$  in the election  $(C, V_1 \cup \dots \cup V_r)$ , provided that  $d_1$  is the representative of the voters in  $V_1$ ,  $d_2$  is the representative of the voters in  $V_2$ , and so on (note that this is a lower bound on the actual  $\beta$ -CC score of  $S$  in this election).

Our algorithm works as follows (we explain Step 2 below):

1. We guess a number  $k' \leq \min\{k, n\}$ , a partition  $\mathcal{V} = (V_1, \dots, V_{k'})$  of  $V$  into  $k'$  voter groups, and a permutation  $\mathcal{V}' = (V'_1, \dots, V'_{k'})$  of  $\mathcal{V}$ .
2. Among all size- $k'$  committees  $S'$  with  $\text{score}_{\mathcal{V}'}(S') \geq t_{CC}$  we pick one with the highest  $k$ -Borda score; we try another guess if no such committee exists.
3. We extend committee  $S'$  to committee  $S$  by adding  $k - k'$  not-yet-included candidates with the highest  $k$ -Borda scores. If the  $k$ -Borda score of  $S$  is at least  $t_{kB}$  then we accept (its  $\beta$ -CC score must be at least  $t_{CC}$  by definition). Otherwise, we try another guess.

We reject if we ran out of guesses without accepting. To compute committee  $S'$  from Step 2, we use dynamic programming: For each  $r \in [k']$ ,  $\ell \in [m]$ , and  $x \in [n(m-1)]$  we define  $f(r, \ell, x)$  to be the highest  $k$ -Borda score that can be achieved by a committee of the form  $\{d_1, \dots, d_r\}$  such that (1)  $d_1 \prec \dots \prec d_r = c_\ell$  and (2)  $\text{score}_{\mathcal{V}'}(\{d_1, \dots, d_r\}) = x$ ; we define  $f(r, \ell, x)$  to be  $-\infty$  if a required committee does not exist, and we define it to be 0 if  $r = \ell = x = 0$ . Intuitively,  $r$  is the size of the subcommittee under consideration,  $\ell$  is the index of its last member with respect to the order  $\prec$  ( $c_\ell$  is intended as a representative of the voters in group  $V'_r$ ), and  $x$  is the expected  $\beta$ -CC score of the committee (computed using the  $\text{score}_{\mathcal{V}'}(\cdot)$  function). One can verify that for given values of  $r, \ell$ , and  $x$ , we can express  $f(r, \ell, x)$  recursively as

$$f(r, \ell, x) = \max_{0 \leq \ell' < \ell} f(r-1, \ell', x - \beta\text{-score}_{(C, V'_r)}(c_\ell)).$$

Based on this formula, and using standard dynamic programming techniques, we can compute committee  $S'$  in Step 2 in polynomial time. Thus, the algorithm runs in FPT-time with respect to the number of voters (in particular, we have  $O(k \cdot n! \cdot n^n)$  guesses; all other steps are polynomial-time computable). For correctness, note that whenever the algorithm accepts, it means that it has found a committee satisfying the given constraints. For the other direction, let us

assume that  $W = \{d_1, \dots, d_k\}$  is a committee witnessing that we have a *yes*-instance; for the sake of brevity, let us assume that  $k \leq n$ , that every voter is represented by some committee member from  $W$ , and, w.l.o.g., that  $d_1 \prec \dots \prec d_k$ . For each  $i \in [k]$ , we define  $U'_i$  to be a set of voters whose representative in  $W$  is  $d_i$ . We see that our algorithm considers  $\mathcal{V}' = (U'_1, \dots, U'_k)$  in Step 2 and then accepts (or accepts even before considering this  $\mathcal{V}'$ ). This completes the proof.  $\square$

In fact, Theorem 4 holds for every pair of rules where the former is representation-focused and the latter is (weakly) separable, provided that both rules use polynomial-time-computable scoring functions and that the functions used for the former are polynomially bounded. Further, it also holds for weighted elections, provided that the weights are represented in unary. Yet, if the scoring rule used for the representation-focused rule is not polynomially bounded or the weights are not represented in unary, then our proof does not go through: In such cases our dynamic program requires exponential time. Formally, we have the following theorem (which can be extended to weighted elections):

**Theorem 5.** *There is a representation-focused rule  $R$  (defined using polynomial-time computable scoring functions) such that  $\{R, k\text{-Borda}\}$ -MULTIGOAL COMMITTEE is W[1]-hard when parametrized by the number of voters.*

It is interesting to contrast Theorem 5 with the fact that the FPT algorithm of Betzler et al. [2013] works for every representation-focused rule defined by polynomial-time computable scoring functions, even if they are not polynomially bounded. This shows that our  $\mathcal{R}$ -MULTIGOAL COMMITTEE problem can be strictly more difficult than each of the winner determination problems for the rules in  $\mathcal{R}$ .

### 5.3 Restricted Domains

If our FPT algorithms are too slow, then we may hope that our elections at least have some convenient structure. It is known that  $\beta$ -CC is tractable when voters' preferences are single-peaked [Betzler et al., 2013] or single-crossing [Skowron et al., 2015b]. We show that these results extend to  $\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE. We start by defining the respective restricted domains—see also the survey by Elkind et al. [2017c]—and then we show tractability for both domains.

**Definition 2.** Let  $\triangleleft$  be a total order on a set of candidates  $C$ . An election  $E = (C, V)$  is *single-peaked with respect to axis  $\triangleleft$*  if for every voter  $v_i$  and every triple of candidates  $a, b, c$  such that  $a \triangleleft b \triangleleft c$  it holds that  $b$  is not the least preferred candidate of voter  $v_i$  in  $\{a, b, c\}$ . An election is *single-peaked* if it is single-peaked with respect to some axis.

**Definition 3.** An election  $E = (C, V)$  with  $V = (v_1, \dots, v_n)$  is *single-crossing* if for every pair of candidates  $a, b$  such that  $v_1$  prefers  $a$  to  $b$  there exists a value  $j \in [n]$  such that voters  $v_1, \dots, v_j$  prefer  $a$  to  $b$  and voters  $v_{j+1}, \dots, v_n$  prefer  $b$  to  $a$ .

**Theorem 6.**  *$\{k\text{-Borda}, \beta\text{-CC}\}$ -MULTIGOAL COMMITTEE is in P when the input election is either single-peaked or single-crossing.*

## 6 Experimental Results

We experimentally test what levels of compromise exist for pairs of rules from the set  $\{\text{SNTV}, \text{Bloc}, k\text{-Borda}, \beta\text{-CC}\}$ .

We consider elections generated according to the 2D Euclidean model (see the work of Enelow and Hinch [1984]): Each candidate and each voter is represented by a point in a 2D Euclidean space, and each voter forms his or her preference order by sorting the candidates' points with respect to their distance from his or her point (from the closest to the farthest one). We generate the candidate points and the voter points by drawing them uniformly at random from a  $[-3, 3] \times [-3, 3]$  square. We consider elections with 100 candidates and 100 voters, and we consider committees of size 10. We chose these parameters as either they or similar ones are often used in the literature [Elkind *et al.*, 2017a; Faliszewski *et al.*, 2017; Celis *et al.*, 2018].

For each ordered pair of rules  $(R^{(1)}, R^{(2)})$  from the set  $\{\text{SNTV}, \text{Bloc}, k\text{-Borda}, \beta\text{-CC}\}$ , we proceed as follows. First, we generate 100 elections as described above. Then, for each value  $p$  between 0 and 1 (with step 0.01) and for each election  $E$  that we generated, we compute the highest value  $q$  such that there is a committee  $S$  that is at least  $p$ -approximate with respect to  $R^{(1)}$  and  $q$ -approximate with respect to  $R^{(2)}$  (to this end, we use simple ILP formulations of our problems; this is necessary when  $\beta\text{-CC}$  is involved, for other cases Theorem 1 would suffice). Finally, for each point  $p$  we report the average of the computed values of  $q$ . We note that the results for pairs  $(R^{(1)}, R^{(2)})$  and  $(R^{(2)}, R^{(1)})$  are not necessarily symmetric. For example, consider committees that achieve at least 80% of the  $k$ -Borda score and maximize the Bloc score; on average, such committees are about 95%-approximate for Bloc. However, not all of these committees are 95%-approximate for Bloc, and so we would not use them all if we reversed the roles of  $k$ -Borda and Bloc in this example.

We show our results in Figure 1. The results are surprisingly positive. For example, for  $k$ -Borda and  $\beta\text{-CC}$  (irrespective of the order in which they are considered) we can typically find committees that are 95%-approximate with respect to both rules. For the other pairs of rules the results are slightly worse, but it is always possible to find a committee that achieves a fairly high score according to both of the rules. SNTV is least compatible with Bloc and  $k$ -Borda, but this is hardly surprising as SNTV assigns points only to those committee members that are ranked in top positions.

To verify the robustness of our results, we considered a few other variants of the Euclidean model, the impartial culture model, and Mallow's model [Mallows, 1957]. In each case, the results were either similar or even better.

Following Elkind *et al.* [2017a], we also show 2D histograms for the committees that we compute (see their work for details of the methodology). Briefly put, for each pair of rules, we have computed 5000 elections and for each election we have computed a committee  $S$  such that  $S$  is  $p$ -approximate for both rules, for  $p$  as large as possible. We partition our  $[-3, 3] \times [-3, 3]$  square into  $120 \times 120$  cells, and count how many members of our committees fall in each cell. We show the results in Figure 2. The darker a given point is, the more members of our committees fell into it.

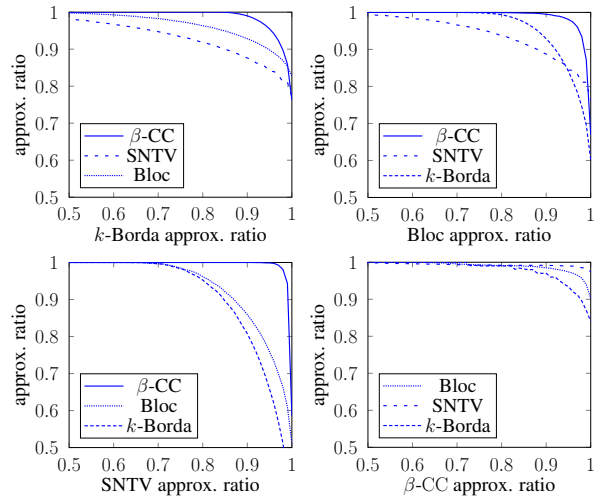


Figure 1: Pareto curves between pairs of rules from the set  $\{\text{SNTV}, \text{Bloc}, k\text{-Borda}, \beta\text{-CC}\}$ . A  $(p, q)$  point in a plot in the top-left figure means that committees that are at least  $p$ -approximate for  $k$ -Borda and maximize the score under a given rule, are, on average,  $q$ -approximate for this rule. The next figures are analogous, but for Bloc (top-right), SNTV (bottom-left), and  $\beta\text{-CC}$  (bottom-right).

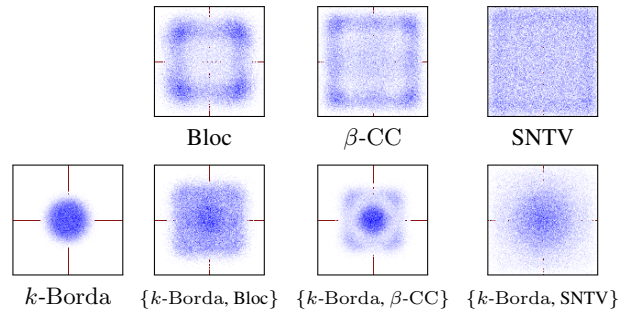


Figure 2: Histograms for MULTIGOAL COMMITTEE variants involving  $k$ -Borda.

## 7 Conclusions

We have studied the problem of computing committees that perform well with respect to given sets of committee scoring rules. Our problem is polynomial-time computable for fixed-size combinations of weakly separable rules (including rules such as SNTV, Bloc, and  $k$ -Borda) and we have found ways of computing committees that perform well according to both  $k$ -Borda and  $\beta\text{-CC}$  (even though the problem is NP-hard). We have also shown experimentally that achieving good compromises between our rules is often possible. Natural future work is to consider other rule sets, and to analyze axiomatic properties of committees computed using our framework.

## Acknowledgments

M. Kocot and A. Kolonko were supported by the AGH University statutory research grant 11.11.230.337. E. Elkind was supported by the ERC grant number 639945 (ACCORD). P. Faliszewski was supported by the National Science Centre, Poland, under project 2016/21/B/ST6/01509.

## References

- [Aziz *et al.*, 2016] Haris Aziz, Jérôme Lang, and Jérôme Monnot. Computing Pareto optimal committees. In *Proceedings of IJCAI-16*, pages 60–66, 2016.
- [Aziz *et al.*, 2017] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017.
- [Betzler *et al.*, 2013] Nadja Betzler, Arkadii Slinko, and Johannes Uhlmann. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research*, 47:475–519, 2013.
- [Celis *et al.*, 2018] L. Elisa Celis, Lingxiao Huang, and Nisheeth K. Vishnoi. Multiwinner voting with fairness constraints. In *Proc. of IJCAI-18*, pages 144–151, 2018.
- [Chamberlin and Courant, 1983] John R. Chamberlin and Paul N. Courant. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *American Political Science Review*, 77(3):718–733, 1983.
- [Debord, 1992] Bernard Debord. An axiomatic characterization of Borda’s  $k$ -choice function. *Social Choice and Welfare*, 9(4):337–343, 1992.
- [Elkind and Erdélyi, 2012] Edith Elkind and Gábor Erdélyi. Manipulation under voting rule uncertainty. In *Proceedings of AAMAS-12*, pages 627–634, 2012.
- [Elkind *et al.*, 2017a] Edith Elkind, Piotr Faliszewski, Jean-François Laslier, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. What do multiwinner voting rules do? An experiment over the two-dimensional Euclidean domain. In *Proceedings of AAAI-17*, pages 494–501, 2017.
- [Elkind *et al.*, 2017b] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632, 2017.
- [Elkind *et al.*, 2017c] Edith Elkind, Martin Lackner, and Dominik Peters. Structured preferences. In U. Endriss, editor, *Trends in Computational Social Choice*. AI Access Foundation, 2017.
- [Enelow and Hinich, 1984] James M. Enelow and Melvin J. Hinich. *The spatial theory of voting: An introduction*. Cambridge University Press, 1984.
- [Erdélyi *et al.*, 2014] Gábor Erdélyi, Edith Hemaspaandra, and Lane A. Hemaspaandra. Bribery and voter control under voting-rule uncertainty. In *Proceedings of AAMAS-14*, pages 61–68, 2014.
- [Faliszewski and Talmon, 2018] Piotr Faliszewski and Nimrod Talmon. Between proportionality and diversity: Balancing district sizes under the Chamberlin–Courant rule. In *Proceedings of AAMAS-18*, pages 14–22, 2018.
- [Faliszewski *et al.*, 2017] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner rules on paths from  $k$ -Borda to Chamberlin–Courant. In *Proceedings of IJCAI-17*, pages 192–198, 2017.
- [Faliszewski *et al.*, 2018] Piotr Faliszewski, Arkadii Slinko, Kolja Stahl, and Nimrod Talmon. Achieving fully proportional representation by clustering voters. *Journal of Heuristics*, 24(5):725–756, 2018.
- [Faliszewski *et al.*, 2019] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Committee scoring rules: Axiomatic characterization and hierarchy. *ACM Transactions on Economics and Computation*, 6(1):Article 3, 2019.
- [Kellerer *et al.*, 2004] Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, 2004.
- [Lackner and Skowron, 2018] Martin Lackner and Piotr Skowron. Consistent approval-based multi-winner rules. In *Proceedings of EC-18*, pages 47–48, 2018.
- [Lackner and Skowron, 2019] Martin Lackner and Piotr Skowron. A quantitative analysis of multi-winner rules. In *Proceedings of IJCAI-19*, 2019.
- [Lu and Boutilier, 2011] Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In *Proc. of IJCAI-11*, pages 280–286, 2011.
- [Mallows, 1957] Colin L. Mallows. Non-null ranking models. *Biometrika*, 44(1–2):114–130, 1957.
- [Obraztsova and Elkind, 2011] Svetlana Obraztsova and Edith Elkind. On the complexity of voting manipulation under randomized tie-breaking. In *Proceedings of IJCAI-11*, pages 319–324, 2011.
- [Obraztsova *et al.*, 2011] Svetlana Obraztsova, Edith Elkind, and Noam Hazon. Ties matter: Complexity of voting manipulation revisited. In *Proceedings of AAMAS-12*, pages 71–78, 2011.
- [Peters, 2018] Dominik Peters. Single-peakedness and total unimodularity: New polynomial-time algorithms for multi-winner elections. In *Proceedings of AAAI-18*, pages 1169–1176, 2018.
- [Procaccia *et al.*, 2008] Ariel D. Procaccia, Jeffrey S. Rosenschein, and Aviv Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353–362, 2008.
- [Sánchez-Fernández *et al.*, 2017] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús A. Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of AAAI-17*, pages 670–676, 2017.
- [Skowron *et al.*, 2015a] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. Achieving fully proportional representation: Approximability result. *Artificial Intelligence*, 222:67–103, 2015.
- [Skowron *et al.*, 2015b] Piotr Skowron, Lan Yu, Piotr Faliszewski, and Edith Elkind. The complexity of fully proportional representation for single-crossing electorates. *Theoretical Computer Science*, 569:43–57, 2015.
- [Skowron *et al.*, 2019] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. Axiomatic characterization of committee scoring rules. *Journal of Economic Theory*, 180:244–273, 2019.