Possible Winners in Approval Voting

Nathanaël Barrot[1,2], Laurent Gourvès[2,1], Jérôme Lang[2,1], Jérôme Monnot[2,1], and Bernard Ries[1,2]

[1] PSL, Université Paris-Dauphine, 75775 Paris Cedex 16, France
[2] CNRS, LAMSADE UMR 7243
{nathanael.barrot,laurent.gourves,lang,monnot,ries}@lamsade.dauphine.fr

Abstract. Given the knowledge of the preferences of a set of voters over a set of candidates, and assuming that voters cast sincere approval ballots, what can we say about the possible (co-)winners? The outcome depends on the number of candidates each voter will approve. Whereas it is easy to know who can be a unique winner, we show that deciding whether a set of at least two candidates can be the set of co-winners is computationally hard. If, in addition, we have a probability distribution over the number of candidates approved by each voter, we obtain a probability distribution over winners; we study the shape of this probability distribution empirically, for the impartial culture assumption. We study variants of the problem where the number of candidates approved by each voter is upper and/or lower bounded. We generalize some of our results to multiwinner approval voting.

Keywords: Computational social choice, Approval voting, Voting under incomplete knowledge, Computational complexity.

1 Introduction

While most voting rules take as input a collection of *rankings* over candidates, approval voting stands as an exception and takes as input a collection of *subsets* of candidates [7]. It is well-known that there is no single sincere approval ballot given a voter's preferences over a set of candidates: for any candidate c, approving the set of all candidates that are preferred to c is a sincere ballot [8]. If the voter's preference relation over a set of m candidates is a linear order, this makes m sincere ballots[1].

Assume that we[2] know the preference relation of every voter (each assumed to be a linear order) but that we cannot predict the *threshold* they will fix, that is, the number of candidates they will approve. For each vector of such thresholds (one for each voter), there will be a winner, or, in case of a tie, a set of co-winners, called a co-winning set. We say that a subset of candidates is a

[1] Sometimes, voting for *all* candidates is excluded, which makes only $m-1$ sincere ballots. See for instance [9].

[2] 'We' is generic, and represents anyone who may reason about the outcome of the vote; the chair, for instance.

P. Perny, M. Pirlot, and A. Tsoukiàs (Eds.): ADT 2013, LNAI 8176, pp. 57–70, 2013.

possible co-winning set if it is the set of co-winners for some vector of thresholds, and candidate x is a possible unique winner if $\{x\}$ is a possible co-winning set. The properties of the set of possible approval winners has been addressed first in [9], with the restriction that voters cannot approve all candidates (nor none). They show that the set of possible approval winners contains the Condorcet winner (if any) and also the winner(s) of many voting rules. Another related work is [28], who gives a geometric interpretation of the set of possible approval winners. None of both works characterizes possible winning sets, nor addresses the computational difficulties of identifying them.

We go further in several respects. First, we consider a more general setting where the number of approved candidates can be anything between a fixed lower bound and a fixed upper bound. In the case where voters are totally free of the number of approvals, that is, when these bounds are respectively 1 and m candidates[3], characterizing the set of candidates that can be a unique winner (without a tie) turns out to be straightforward: x is a possible unique winner if it is Pareto-undominated in the original profile. We give a similar characterization when the bounds are different. Then we consider the problem of recognizing co-winning sets, and show that it is NP-complete, even for sets of size two. Next, we consider a probabilistic version of the problem, starting with a probability distribution over approval vectors; we focus on the uniform distribution, and in this case we first observe that the probability that a candidate is in the co-winning set is proportional to its Borda score; then, assuming impartial culture, we study experimentally the shape of the probability distribution over winners.

This work is related to (at least) four research streams. The first of these is a series of works in social choice theory that relate approval voting to the classical Arrovian model, which considers social choice functions mapping a collection of weak orders into a nonempty subset of candidates (whereas approval voting generates the social outcome by aggregating collections of subsets of candidates). For this, the key notion is that of sincere ballot, already evoked above. Most works in this research stream (with the exception of [9] and [28] cited above) study the conditions under which approval voting can, or cannot, be considered strategyproof, and the extent to which strategic behaviour may lead to an undesirable outcome; see [29,30,23,15,26,24,14].

The second related research stream is the characterization and computation of possible and necessary winners given some incomplete information about the votes. The main difference with our setting is that in all these works (up to one exception, discussed below), the voting rule used takes a classical profile, that is, a collection of rankings, as input, and the incomplete information consists of a collection of *partial orders*: a possible (resp. necessary) winner is then a candidate that wins in some completion (respectively, all completions) of this collection of partial orders [21,32,4,3,33,10,5,1,22,18]. An exception is [33], which, in Section 4, states a characterization of possible winners in approval voting, given an

[3] Approving no candidate and approving all of them are equivalent, in the sense that whatever the remaining votes, the outcome will be the same. Therefore, without loss of generality, we exclude the possibility for a voter to approve 0 candidate.

initial approval ballot over an initial set of candidates, and given a number of new candidates to be added; the nature of the incomplete information about approval ballots in their setting and ours (an approval profile over a subset of candidates vs. a ranking profile over all candidates) is totally different, and results cannot easily be compared.

The third related research stream is a series of works that focuses on the computational aspects of strategic behaviour in approval voting; see in particular [16,2]. The reason why it relates to our work is that we also find computationally hard problems in approval voting; but, once again, our problems do not come from any form of strategic behaviour. Lastly, the computational aspect of strategic behaviour in *multiwinner* versions of approval voting was considered in [25].

Lastly, our Section 4, where we study the complexity of identifying possible outcomes in multiwinner approval voting, relates to the computational study of multiwinner election schemes, such as full proportional representation [27,6,31], Condorcet winning sets [13] or other approaches to committee selection [11,12,20] (we discuss [11] in more detail in Section 4).

The paper is organized as follows. In Section 2 we introduce the necessary background. In Section 3 we define possible and necessary (co)winners and give some characterizations as well as some hardness results about the identification of possible winning sets. In Section 4 we consider multiwinner elections, and generalize some of our results of Section 3. In Section 5 we present further research issues.

2 Preliminaries

We are given n *voters* $N = \{1, \ldots, n\}$ and m *candidates* (or *alternatives*) $X = \{x_1, \ldots, x_m\}$. A *ranking profile* $P = (P_i)_{i \in N}$ is a collection of linear orders (also called rankings) over X. P_i is also denoted by $\succ_i$.

An *approval ballot* is a nonempty subset of X. An *approval profile* is a collection $A = \langle A_1, \ldots, A_n \rangle$ where $A_i \subseteq X$ is the set of candidates approved by voter i. Such an approval ballot is called *sincere*, if for every voter i and every candidate x_j approved by i there exists no candiate x not approved by i such that $x \succ_i x_j$. We denote by k_i, for $i = 1, \ldots, n$ and $1 \leq k_i \leq m$, the number of candidates approved by voter i. Hence, in a sincere approval ballot, each voter i approves its k_i best candidates according to the ranking given by P_i.

Given an approval profile A, the *approval score* of candidate x_j, denoted by $app_A(x_j)$, or, when there is no ambiguity, $app(x_j)$, is the number of voters i such that $x_j \in A_i$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The set of *approval co-winners for* A, denoted by $App(A)$, is the set of candidates with maximal approval score. If $App(A)$ is a singleton $\{a\}$ then a is said to be a *single winner* for A[4].

For a ranking profile P, voter $i \in N$ and candidate $x \in X$, $\mathrm{rk}_P(i, x) \in \{1, \ldots, m\}$ denotes the rank of x in the ranking P_i. For $X' \subset X$, let $P_{X'}$ be the

[4] Approval voting is here considered an *irresolute* voting rule; a *resolute* version of approval voting can be defined by applying a tie-breaking priority mechanism.

restriction of P to candidates in X'. We denote by $pl(x, X')$ the plurality score approval score of candidate $x \in X'$ in profile $P_{X'}$, that is, the number of voters in $P_{X'}$ who rank x on top. For $X' \subset X$ and $x \in X \setminus X'$, we write $X' \succ_i x$ if $\forall x' \in X'$, $x' \succ_i x$ and candidates among X' are ranked arbitrarily. Finally, we say that candidate x *dominates* candidate x' according to profile P if $\forall i \in N$, $x \succ_i x'$.

Approval voting can also be used for *multiwinner* elections. Here the goal is to elect a set of alternatives, or a *committee*, of fixed size K. There are several procedures for determining a committee using approval voting, which are reviewed in [19]. The most obvious way consist in choosing the candidates with the K highest approval scores (using some tie-breaking mechanism if necessary).

Sometimes, a further constraint on the number of approvals is added: each voter is only allowed to approve at least d and most k candidates, where $k \geq d \geq 1$; a typical choice, often implemented in real-world elections, consists in fixing d to 1 and k to an arbitrary constant (such as, in multi-winner elections, the number of positions to be filled). The corresponding voting rule, mapping any collection of n subsets of X of cardinality between d and k, is called $[d, k]$-*approval voting.* Notice that approval voting is equivalent to $[1, m]$-approval voting.

3 Single-Winner Approval Voting

3.1 Restriction-Free Approval Voting

We start by defining the set of approval ballots that are compatible with a ranking profile.

Definition 1

- *A* threshold vector *(for N and X) is a vector $\boldsymbol{k} = \langle k_1, \ldots, k_n \rangle \in \{1, \ldots, m\}^n$.*
- *Let $P = \langle P_1, \ldots, P_n \rangle$ be a ranking profile over X, and $\boldsymbol{k}$ a threshold vector. For all $i \leq n$, let $(P_i)^{1 \to k_i}$ be the subset of X defined by*

$$(P_i)^{1 \to k_i} = \{x \in X \mid \mathrm{rk}(x, P_i) \leq k_i\}$$

The approval profile induced by P and $\boldsymbol{k}$*, denoted by $A^{P,\boldsymbol{k}}$, is defined as*

$$A^{P,\boldsymbol{k}} = \langle (P_1)^{1 \to k_1}, \ldots, (P_n)^{1 \to k_n} \rangle$$

- *The set of all approval profiles compatible with P is defined as*

$$CAP(P) = \{A^{P,\boldsymbol{k}} \mid \boldsymbol{k} \in \{1, \ldots, m\}^n\}$$

Example 1. Let $m = n = 3$, $P = \langle x_1 \succ x_2 \succ x_3, x_1 \succ x_2 \succ x_3, x_3 \succ x_1 \succ x_2 \rangle$ and $\boldsymbol{k} = \langle 2, 1, 2 \rangle$; then $A^{P,\boldsymbol{k}} = \langle \{x_1, x_2\}, \{x_1\}, \{x_1, x_3\} \rangle$.

Definition 2. *Let P be a ranking profile P over X. A subset $X' \subseteq X$ is called a* possible co-winner set *for P if there exists a threshold vector $\boldsymbol{k}$ such that $X' = App(A^{P,\boldsymbol{k}})$. The set of all possible co-winner sets for P is denoted by $PCS(P)$. $x \in X$ is a* possible single winner *for P if $\{x\} \in PCS(P)$, a* possible co-winner *if it belongs to some possible co-winner set, a* necessary co-winner *if it belongs to all co-winner sets for P, and a* necessary single winner *if $PCS(P) = \{x\}$.*

Example 1, continued. $\{x_1\}$ is a possible co-winner set (and hence x_1 a possible single winner) for P, obtained for instance for $\boldsymbol{k} = \langle 1, 3, 3 \rangle$, for $\boldsymbol{k} = \langle 1, 2, 2 \rangle$, and for many other threshold vector; and

$$PCS(P) = \{\{x_1, x_2, x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_1\}, \{x_3\}\}$$

whereas the possible single winners for P are x_1 and x_3.

Without any restriction on the allowed thresholds, the notions of possible co-winner, necessary co-winner and necessary single winner turn out to trivialize: all candidates are possible co-winners, no candidate is a necessary single winner, and a x is a necessary co-winner if and only if it is ranked on top of all votes.

We now consider the following question: given a ranking profile P and a subset of X' of candidates, is X' a possible winner set for P? We call this problem the POSSIBLE CO-WINNER SET PROBLEM FOR APPROVAL VOTING[5]. This problem turns out to be easy in the case where X' is a singleton:

Theorem 1. *x is a possible single winner for P if and only if no candidate in $X \setminus \{x\}$ dominates x in P.*

Proof. Assume no y dominates x in P. Define $\boldsymbol{k}$ by $k_i = \mathrm{rk}_P(i, x)$ for any $i \in N$. x is approved n times in $A^{P,\boldsymbol{k}}$; if $y \neq x$ is also approved n times $A^{P,\boldsymbol{k}}$, then for all i, $\mathrm{rk}_P(i, y) \leq k_i$, *i.e.*, y would dominates x in P; therefore, $App(A^{P,\boldsymbol{k}}) = \{x\}$. Conversely, if y dominates x in P, then for all $\boldsymbol{k}$, y will be approved at least as many times as x in $A^{P,\boldsymbol{k}}$, therefore x cannot be a posssible single winner.

As a consequence, the restriction of the possible co-winner set problem to singletons can be solved in polynomial time. This property does not generalize to subsets of arbitrary size. Indeed, the possible co-winner set problem is computationally hard, even under the restriction to sets of candidates of fixed size $\ell \geq 2$. We first prove the following lemma.

Lemma 1. *If $X' \in PCS(P)$, then there exists a solution $(k_i)_{i \in N}$ satisfying the following properties:*

(a) For any $i \in N$, $k_i \in \{\mathrm{rk}_P(i, x) : x \in X'\}$.
(b) The score of any co-winner is at least $\max_{x \in X'} pl(x, X')$.

Proof. Let $X' \in PCS(P)$. *(a)*: Let $(k_i)_{i \in N}$ be any solution such that the candidates in $X' = \{x_1, \ldots, x_\ell\}$ are exactly the co-winners for profile P. Consider a voter i and, without loss of generality, assume that $x_1 \succ_i \cdots \succ_i x_\ell$. Moreover, assume $k_i \notin \{\mathrm{rk}_P(i, x) : x \in X'\}$. If $\mathrm{rk}_P(i, x_j) < k_i < \mathrm{rk}_P(i, x_{j+1})$ with $j \in \{1, \ldots, \ell - 1\}$, then we replace k_i by $\mathrm{rk}_P(i, x_j)$. If $k_i < \mathrm{rk}_P(i, x_1)$ or $k_i > \mathrm{rk}_P(i, x_\ell)$, we replace k_i by $\mathrm{rk}_P(i, x_\ell)$. It is not difficult to see that X' remains exactly the co-winner set. By repeating this procedure for each voter, we obtain the expected result. *(b)*: Using (a), we know that there is a solution $(k_i)_{i \in N}$ such that the global score of candidate $x_j \in X'$ is at least $pl(x_j, X')$. Since the candidates in X' are the co-winners, we must have that each candidate of X' is approved at least $\max_{x \in X'} pl(x, X')$ times.

[5] From now on we will generally omit "for approval voting".

Theorem 2. *Let $\ell \geq 2$. Given a profile P and a subset of candidates X' such that $|X'| = \ell$, determining whether X' is a possible co-winner set for P is **NP**-complete.*

Proof. The problem is clearly in **NP** for all $\ell \geq 2$. Let us give a proof for the case when $\ell = 2$ and explain then how to generalize to all other cases. The proof of the **NP**-completeness is based on a reduction from EXACT 3-SET COVER (X3C in short). In an instance of X3C, we are given a family of m sets $\mathcal{S} = \{S_1, \ldots, S_m\}$ over a ground set $Y = \{y_1, \ldots, y_{3n}\}$ such that $\cup_{i=1}^{m} S_i = Y$ and $|S_i| = 3$, for $i = 1, \ldots, m$. The question is whether there exists a subset $J \subseteq \{1, \ldots, m\}$ of size n such that $\sum_{j \in J} S_j = Y$? This problem is known to be **NP**-complete [17].

Let $I = (\mathcal{S}, Y)$, with $\mathcal{S} = \{S_1, \ldots, S_m\}$ and $Y = \{y_1, \ldots, y_{3n}\}$, be an instance of X3C. We build an instance of POSSIBLE CO-WINNER SET FOR APPROVAL VOTING, with $\ell = 2$, as follows. There are $2m - n$ voters $N = \{1, \ldots, m - n\} \cup \{1', \ldots, m'\}$ and $m + 2n + 2$ candidates $X = E \cup Y \cup \{a, b\}$ where $E = \{e_1, \ldots, e_{m-n}\}$ and we set $X' = \{a, b\}$ as the target candidates. The profile P is given by:

- For $1 \leq i \leq m - n$, $E \setminus \{e_i\} \succ_i Y \succ_i a \succ_i e_i \succ_i b$.
- For $1 \leq j \leq m$, $b \succ_{j'} Y \setminus S_j \succ_{j'} E \succ_{j'} a \succ_{j'} S_j$.

This clearly gives us an instance I' of POSSIBLE CO-WINNER SET. We claim that there exists a subset $J \subseteq \{1, \ldots, m\}$ with$|J| = n$ such that $\sum_{j \in J} S_j = Y$ if and only if $\{a, b\}$ is a possible co-winner set for P.

Suppose that I is a yes-instance of X3C, *i.e.*, there exists $J \subseteq \{1, \ldots, m\}$ with$|J| = n$ such that $\sum_{j \in J} S_j = Y$. We set $k_{j'} = \mathrm{rk}_P(j', a)$ for $j \in J$. For the remaining voters $i \in N \setminus \{j' : j \in J\}$, we set $k_i = \min\{\mathrm{rk}_P(i, a), \mathrm{rk}_P(i, b)\}$. a and b are approved m times while candidates in $E \cup Y$ are approved at most $m - 1$ times. Thus $X' = \{a, b\}$ is a possible co-winner set.

Conversely, assume that I' is a yes-instance of POSSIBLE CO-WINNER SET. Using (a) and (b) of Lemma 1, there exists $\boldsymbol{k}$ with $k_i \in \{\mathrm{rk}_P(i, a); \mathrm{rk}_P(i, b)\}$ for any $i \in N$ and a, b must be approved at least m times. Thus, there exists $J \subseteq \{1, \ldots, m\}$ such that $k_{j'} = \mathrm{rk}_P(j', a)$ for $j \in J$ and $k_{j'} = \mathrm{rk}_P(j', b)$ for $j \notin J$. In particular, we deduce that $\sum_{j \in J} S_j = X$ since otherwise any candidate of $X \setminus (\sum_{j \in J} S_j)$ necessarily dominates a; hence, $|J| \geq n$. Moreover, if $|J| \geq n + 1$, then a gets approved at least $m + 1$ times. Thus, there exists at least one voter $i \in \{1, \ldots, m - n\}$ such that $k_i = \mathrm{rk}_P(i, b)$ (since a and b must get approved the same number of times). But then $app(e_i) \geq app(a)$, a contradiction. So we conclude that $|J| = n$ and $\sum_{j \in J} S_j = Y$: I is a yes-instance of X3C.

This shows the **NP**-completeness of POSSIBLE CO-WINNER SET FOR APPROVAL VOTING, restricted to co-winner sets of size 2. Now it is not difficult to see that, if we proceed exactly the same way and replace everywhere in the previous proof a by $\{a_1, \ldots, a_{\ell-1}\}$ and we set $X' = \{b\} \cup \{a_1, \ldots, a_{\ell-1}\}$, for $\ell \geq 3$, we can show the **NP**-completeness of POSSIBLE CO-WINNER SET FOR APPROVAL VOTING restricted to co-winner sets of size ℓ.

3.2 Approval Voting with Restriction on the Number of Approvals

We now consider, more generally, $[d, k]$-approval voting. The definitions are natural generalizations of those in Section 3.1, with the difference that each k_i should be such that $d \leq k_i \leq k$. The set of all $[d, k]$-approval profiles compatible with P is defined by $CAP_{d,k}(P) = \{A^{P,\boldsymbol{k}} \mid \boldsymbol{k} \in [d, k]^n\}$, and the set of possible $[d, k]$-approval co-winner sets for P is denoted by $PCS_{d,k}(P)$.

Example 1, continued.

- $PCS_{1,2}(P) = \{\{x_1\}, \{x_1, x_2\}\}$;
- $PCS_{2,3}(P) = \{\{x_1\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_2, x_3\}\}$

Again, in order to check whether x is a possible single winner, it is enough to check it for a specific choice of $\boldsymbol{k}$, namely, the best possible choice for x.

Theorem 3. *$\{x\} \in PCS_{d,k}(P)$ if $App(A^{P,\boldsymbol{k}}) = \{x\}$ for $\boldsymbol{k}$ defined by $k_i = \mathrm{rk}_P(i, x)$ when $\mathrm{rk}_P(i, x) \in [d, k]$, and $k_i = d$ otherwise.*

Proof. ($\Leftarrow$) is direct from the definition. For ($\Rightarrow$), suppose $\{x\} \in PCS_{d,k}(P)$. Then there is a vector $(k'_i)_{i \in N}$ for which x is the single winner. If $\boldsymbol{k}' = \boldsymbol{k}$ then we are done. Otherwise, take the voter i with minimum index that satisfies $k'_i \neq k_i$. If $k'_i < k_i$ then doing $k'_i \leftarrow k_i$ increases the score of a subset of candidates by one unit, and this subset includes x. If $k'_i \geq k_i$ then by doing $k'_i \leftarrow k_i$ the score of x remains unchanged, while the score of some other candidates decreases. In all, x remains the single winner by the operation $k'_i \leftarrow k_i$. Repeating the operation until $\boldsymbol{k}' = \boldsymbol{k}$ leads to the result. □

Theorem 3 generalizes Lemma 2 in [9]; for $d = 1$ and $k = m - 1$, we recover their notion of *critical strategy profile for x*: every voter who ranks i as his worst candidate approves only one candidate; the other voters vote for i and all candidates above. Then x is a possible $1, m - 1$- approval winner (called AV outcome in [9]) if x wins at his critical strategy profile.

As a corollary, we get simple characterizations of possible and necessary co-winners and single winners, which we state without proof: let $D^+_P(x, y) = \{i \mid \mathrm{rk}_i(P, x) \leq k, \mathrm{rk}_i(P, y) > d \,\text{and}\, x \succ_i y\}$ and $D^-_P(x, y) = \{i \mid \mathrm{rk}_i(P, x) \leq d \,\text{and}\, \mathrm{rk}_i(P, y) > k\}$. Then x is a possible $[d, k]$-approval co-winner (respectively, possible single winner, necessary co-winner, single winner) for P if and only if for all $y \neq x$, $|D^+_P(x, y)| \geq |D^-_P(y, x)|$ (respectively, $|D^+_P(x, y)| > |D^-_P(y, x)|$, $|D^-_P(x, y)| \geq |D^+_P(y, x)|$, $|D^-_P(x, y)| > |D^+_P(y, x)|$).

Theorem 2 immediately extends to $[1, k]$-approval for $k = m - 2$ because in the proof of Theorem 2 we do not approve more $m - 2$ candidates for each voter.

Theorem 4. *For any integer $\ell \geq 2$, the problem of checking whether X' is a $[1, m - 2]$-approval possible co-winner set is **NP**-complete, even under the restriction $|X'| = \ell$.*

Remark: The proof of Theorem 2 can be adapted in such a way that for any integers $\ell \geq 2$ and $d \geq 2$, checking whether X' is a $[d, k]$- approval possible co-winner set, under the restriction $|X'| = \ell$, is **NP**-complete for some k. Moreover, using Algorithm 1 described in Subsection 3.3 we can prove that checking whether X' is a $[d, k]$- approval possible co-winner set is polynomial whenever $k - d$ is upper bounded by a constant.

3.3 The Probability of Possible Co-winner Sets

Definition 3. *Let p be a probability distribution on all threshold vectors. Given a profile P and a subset of candidates $X' \subseteq X$, the probability that X' is the co-winner set (for approval) is equal to $\sum_{\boldsymbol{k}|App(A^{P,\boldsymbol{k}})=X'} pr(\boldsymbol{k})$.*

A simple assumption consists in assuming that $\pi(i, r)$ approves his r most preferred candidates with a given probability $\pi(i, r)$, and that voters' choices are probabilistically independent. Under this assumption, we show how to compute efficiently the probability of each co-winner subset.

We first show how to enumerate all possible scores and their probabilities. Given a voter i and a threshold $k_i \in [d..k]$, we define $\text{TRACE}(i, k_i)$ as the m-dimensional 0-1 vector whose coordinate j is equal to 1 if candidate x_j belongs to the k_i most preferred candidates of i, and 0 otherwise. For example, there are 4 candidates and voter i's preference profile is $x_2 \succ_i x_3 \succ_i x_1 \succ_i x_4$; we have $\text{TRACE}(i, 1) = (0, 1, 0, 0)$, $\text{TRACE}(i, 2) = (0, 1, 1, 0)$, $\text{TRACE}(i, 3) = (1, 1, 1, 0)$ and $\text{TRACE}(i, 4) = (1, 1, 1, 1)$.

We suppose wlog. that the voters provide their ballots sequentially, by ascending index, and a list L_i contains all possible scores after voter i's turn. Therefore L_i is defined as L_{i-1} to which one adds the possible ballots of voter i.

An element of a list is a couple composed of an m-dimensional vector (a score for each candidate) and a probability. We suppose that a list never contains two elements with the same vector. In addition, a list is sorted by its elements' vectors which are sorted in lexicographic order (e.g. $(1, 3, 4) <_{lex} (1, 4, 0)$). A possible list can be $\langle((1, 3, 4), 0.24), ((1, 4, 0), 0.36), ((2, 0, 1), 0.15)\rangle$.

We use a subroutine $\text{MERGE-LISTS}(L, L')$ that merges the lists L and L'. If several elements have the same vector then they are combined in a unique element whose probability is the sum of all condensed elements' probabilities. For example, $\text{MERGE-LISTS}(\langle((1, 4, 0), 0.36),\ ((2, 0, 1), 0.15)\rangle,\ \langle((1, 2, 6), 0.41), ((1, 4, 0), 0.06)\rangle)$ is equal to $\langle((1, 2, 6), 0.41),\ ((1, 4, 0), 0.42),\ ((2, 0, 1), 0.15)\rangle$. $\text{MERGE-LISTS}(L, L')$ needs $|L| + |L'|$ operations.

Given a list L, a vector vec and its probability π, $L \oplus (vec, \pi)$ means that we add vec to every vector of L (component by component) and we multiply every probability by π. For example, $\langle((1, 4, 0), 0.3), ((2, 0, 1), 0.1)\rangle \oplus ((1, 0, 1), 0.3)$ gives $\langle((2, 4, 1), 0.09), ((3, 0, 2), 0.03)\rangle$. $L \oplus (vec, \pi)$ requires $|L|$ operations.

Algorithm 1 gives the exhaustive list of outcomes with their probabilities, where $\mathbf{0}$ denotes the m-dimensional vector whose coordinates are all equal to 0.

Then we can retrieve from L_n the winner sets and their probabilities. The size of L_i is at most $(k - d + 1)|L_{i-1}|$, and $|L_0| = 1$ so $|L_n| \leq (k - d + 1)^n \leq m^n$.

Algorithm 1. All possible scores with probabilities

$L_0 \leftarrow \langle(\mathbf{0}, 1)\rangle$
for $i = 1$ to n **do**
  $L' \leftarrow \langle\rangle$
  for $r = d$ to k **do**
    $L' \leftarrow$ MERGE-LISTS$(L', L_{i-1} \oplus (\text{TRACE}(i, r), \pi(i, r)))$
  end for
  $L_i \leftarrow L'$
end for
return L_n

Meanwhile the final score of any candidate belongs to $[0..n]$ so there are at most $(n+1)^m$ distinct vectors of scores and $|L_n| \leq (n+1)^m$. Thus Algorithm 1 is exponential in the input size but it is polynomial when n or m is a fixed constant.

As a short example, consider an instance with $m = n = 3$, $d = 1$ and $k = 3$. The profiles are $x_1 \succ_1 x_2 \succ_1 x_3$, $x_3 \succ_2 x_1 \succ_2 x_2$ and $x_2 \succ_3 x_3 \succ_3 x_1$. We suppose that for every voter i, the probabilities that the k_i first candidates are approved are 0.3, 0.5 and 0.2 when k_i is equal to 1, 2 and 3 respectively. These probabilities are independent. Hence voter 1 approves $\{x_1, x_2\}$ with probability 0.5. And voter 2 approves $\{x_1, x_2\}$ with probability 0 since it is not a sincere vote. Running Algorithm 1 yields the values given in Table 1.

Table 1. Output of Algorithm 1 on the example

detailed scores & corresponding prob.	winner(s) & total prob.	detailed scores & corresponding prob.	winner(s) & total prob.
(322) (312) (211) 0.082 0.03 0.045	$\{x_1\}$ 0.157	(212) 0.093	$\{x_1, x_3\}$ 0.093
(121) (231) (232) 0.045 0.03 0.062	$\{x_2\}$ 0.137	(122) (233) 0.093 0.02	$\{x_2, x_3\}$ 0.113
(112) (123) (223) 0.045 0.03 0.012	$\{x_3\}$ 0.087	(111) (222) (333) 0.027 0.235 0.008	$\{x_1, x_2, x_3\}$ 0.27
(221) (332) 0.123 0.02	$\{x_1, x_2\}$ 0.143		

3.4 Experimental Analysis

Finally, we provide an experimental analysis of the sensitivity of the winner to the choice of the thresholds. We generate $5 * 10^4$ ranking profiles with an uniform distribution (*impartial culture assumption*). For each profile we generate $5 * 10^4$ threshold vectors with a uniform distribution, for each of these vectors we compute the winner (ties being broken randomly), and we obtain the winning probability of each candidate. We reorder these winning probabilities decreasingly. Then we compute the average, over all generated profiles, of the largest winning probability. The results of these experiments are summarized in Table 2. We observe that the largest winning probability is above 50% with a low number of candidates and any number of voters. This probability decreases when the number of candidates increases.

Table 2. Largest winning probability with uniformly drawn profiles and thresholds

	n = 5	n = 20	n = 50	n = 100
m = 5	55.9	58.5	55.4	54.7
m = 20	32.8	35.0	34.6	35.1
m = 50	23.9	27.1	27.3	27.3
m = 100	18.3	22.4	22.7	22.8

We also compute the average of the second and third largest winning probabilities. Figure 1 shows us the evolution of the largest, second and third largest winning probabilities as a function of the number of candidates, with $n = 5$. Finally, in Figure 2 we represent the largest winning probability as a function of the number of voters. The largest winning probability appears to be independent from the number of voters.

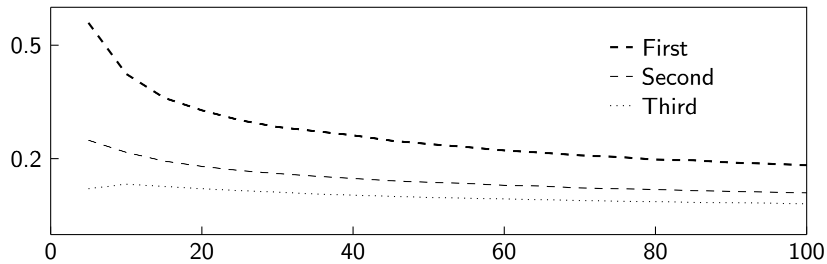

Fig. 1. Largest, second and third largest winning probabilities as a function of the number of candidates, with $n = 5$

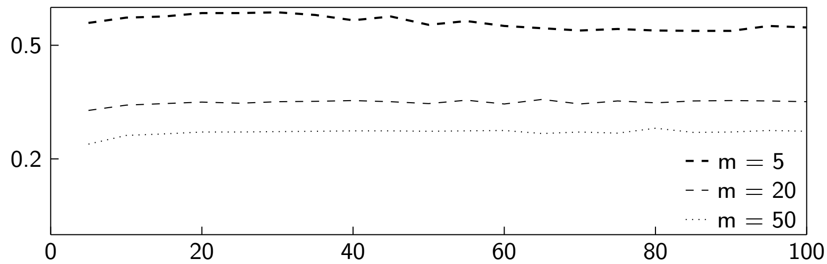

Fig. 2. Largest winning probability as a function of the number of voters

4 Multiwinner Approval Voting

We now briefly reconsider some of the questions addressed in Section 3 in the context of *multiwinner* approval voting. We are now given an integer $K \leq m$, and look for a committee of size K.

Definition 4. *For any approval profile A, $App_K(A)$ is the set of all committees $X' \subseteq X$ of size K such that for any $x \in X$ and $z \notin X'$, $app_A(x) \geq app_A(z)$. Let P a ranking profile over X. $X' \subseteq X$ is a* possible winning K-committee *for P if $X' \in App_K(A^{P,\boldsymbol{k}})$ for some threshold vector $\boldsymbol{k}$. The set of all possible winning K-committees for P is denoted by $PCS_K(P)$. $x \in X$ is* possibly *(resp.* necessarily*) elected w.r.t. K and P if it belongs to some (resp. all) possible winning K-committee(s) for P. Let $Poss_K(P)$ and $Nec_K(P)$ be the set of possibly (resp.* necessarily*) elected candidates w.r.t. K and P. These definitions naturally generalize to $[d,k]$-approval voting.*

The following result generalizes Theorem 1.

Theorem 5. *$x \in Poss_K(P)$ if and only if x is not Pareto-dominated by K candidates or more.*

Proof. Suppose that x is member of a possible winning K-committee, then the candidates that dominate x are also in this winning K-committee, therefore at most $K-1$ candidates dominate x. Conversely, assume that x is dominated by $K-1$ candidates or less. For all $i \in N$, let $k_i = \mathrm{rk}_P(i,x)$. Only x and the candidates that dominates x have an approval score equal to n, therefore x belongs to a winning K-committee.

Theorem 6. *$x \in Nec_K(P)$ if and only if x dominates at least $n-K$ candidates.*

The proof is similar to the proof of Theorem 5.

We now consider the following problem: given a ranking profile P over X and a subset $X' \subset X$ of size K, is K a possible winning K-committee for P? We first establish the following lemma, for $K=2$.

Lemma 2. *Let $X' = \{x'_1, x'_2\}$. If $X' \in PCS_2(P)$, then $X' \in App_2(A^{P,\boldsymbol{k}})$ for some $\boldsymbol{k}$ satisfying $|\{i \in N : \mathrm{rk}_P(i,x'_1) \leq k_i\}| = |\{i \in N : \mathrm{rk}_P(i,x'_2) \leq k_i\}|$.*

Proof. Let $X' \in PCS_2(P)$ and let $\boldsymbol{k}$ such that $X' \in App_2(A^{P,\boldsymbol{k}})$. If $app_{A^{P,\boldsymbol{k}}}(x'_1) = app_{A^{P,\boldsymbol{k}}}(x'_2)$, we are done. Otherwise, assume without loss of generality that $app_{A^{P,\boldsymbol{k}}}(x'_1) > app_{A^{P,\boldsymbol{k}}}(x'_2)$. There exists a subset $N' \subset N$ of size $app_{A^{P,\boldsymbol{k}}}(x'_1) > app_{A^{P,\boldsymbol{k}}}(x'_2)$ such that for voters $i \in N', \mathrm{rk}_P(i,x'_1) \leq k_i < \mathrm{rk}_P(i,x'_2)$. We build a new vector $\boldsymbol{k}$ as follows: (i) for $i \in N'$, $k'_i = 0$; (ii) for $i \in N \backslash N'$, $k'_i = k_i$. We have $app_{A^{P,\boldsymbol{k}'}}(x'_1) = app_{A^{P,\boldsymbol{k}}}(x'_1) - (app_{A^{P,\boldsymbol{k}}}(x'_1) - app_{A^{P,\boldsymbol{k}}}(x'_2)$ and $app_{A^{P,\boldsymbol{k}'}}(x'_2) = app_{A^{P,\boldsymbol{k}}}(x'_2)$, therefore $app_{A^{P,\boldsymbol{k}'}}(x'_2) = app_{A^{P,\boldsymbol{k}'}}(x'_1)$.

Theorem 7. *Determining whether X' is a possible winning 2-committee is **NP**-complete.*

Proof. Hardness is shown by a reduction from the problem of determining whether a set of 2 candidates is a possible winning set in single-winner approval (**NP**-complete, cf. Theorem 2). Let $I = (P, N, X, X')$ be an instance of this problem, with $X' = \{x'_1, x'_2\}$. From I, we build an instance I' of the possible winning 2-committee problem, with the same N, X, P, and X'. We claim that X' is a possible winning set in I if X' is a possible winning 2-committee in I'. Clearly, if $X' = App(A^{P,\boldsymbol{k}})$, then $X' = App_2(A^{P,\boldsymbol{k}})$. Conversely, assume that $X' = App_2(A^{P,\boldsymbol{k}})$. By lemma 2, we know that there exists $\boldsymbol{k'}$ such that $X' = App_2(A^{P,\boldsymbol{k}})$ and $app_{A^{P,\boldsymbol{k'}}}(x'_1) = app_{A^{P,\boldsymbol{k'}}}(x'_2)$, therefore, $X' = App_2(A^{P,\boldsymbol{k'}})$.

Unsurprisingly, this difficulty carries on to committees of larger size (the proof, by reduction from the POSSIBLE WINNING 2-COMMITTEE, is easy and omitted):

Theorem 8. *For any integer $K \geq 2$, determining whether X' is a possible winning K-committee is **NP**-complete.*

This complexity result extends to $[1, k]$-approval:

Theorem 9. *For any integer $K \geq 2$, and $k \geq 3$, determining whether X' is a possible winning K-committee for $[1, k]$-approval is **NP**-complete.*

A related series of results on the complexity of multiwinner elections with approval ballots is in [11] (Theorems 3.4 to Corollary 3.9). There the setting is different from ours: each voter approves *exactly* t candidates; if voter i approves $A_i \subseteq X$ (with $|A_i| = t$), then given two k-committees X and Y, i is assumed to prefer X over Y ($X \gg_i Y$) if $|X \cap A_i| > |Y \cap A_i|$. A k-committee X is a *popular* k-committee if it majority-wise defeats all other k-committees (that is, if it a Condorcet winner in the set of all k-committees for the profile $\langle \gg_1, \ldots, \gg_n \rangle$). Darmann shows that deciding whether a k-committee is a popular committee is NP-hard as soon as $2 \leq t \leq m - 2$ (finding such a committee is probably even harder). Unlike ours, the hardness results in [11] are not due to the uncertainty about the number of approvals and they do not imply, nor are implied by, any of our results.

5 Further Issues

When thresholds vectors are generated with a uniform probability, the winning probability of a candidate for a given profile is proportional to its Borda score; more generally, if the probabilities on the number of approvals for voters are i.i.d., the winning probability of a candidate for a profile is proportional to its score for some positional scoring rule. This connection is worth exploring further.

Another interesting topic that we did not explore is the control of an election by a chair who has the power to fix the lower and upper bounds d and k on the number of approvals. Assume that the chair moreover knows the voters' rankings and has some subjective probability distribution on the number of candidates the voters will approve (to be conditioned by the bounds d and k). Clearly, the choice of d and k has an influence on the winning probability of a candidate; this election control is computationally hard if computing winning probabilities is computationally hard — a question that we have not addressed yet.

References

1. Bachrach, Y., Betzler, N., Faliszewski, P.: Probabilistic possible-winner determination. In: Proc. of AAAI 2010 (2010)
2. Baumeister, D., Erdèlyi, G., Hemaspaandra, E., Hemaspaandra, L., Rothe, J.: Computational aspects of approval voting. In: Laslier, J.-F., Sanver, R. (eds.) Handbook of Approval Voting, pp. 199–251. Springer (2010)
3. Baumeister, D., Rothe, J.: Taking the final step to a full dichotomy of the possible winner problem in pure scoring rules. In: Proceedings of ECAI 2010 (2010)
4. Betzler, N., Dorn, B.: Towards a dichotomy of finding possible winners in elections based on scoring rules. In: Královič, R., Niwiński, D. (eds.) MFCS 2009. LNCS, vol. 5734, pp. 124–136. Springer, Heidelberg (2009)
5. Betzler, N., Hemmann, S., Niedermeier, R.: A multivariate complexity analysis of determining possible winners given incomplete votes. In: Proceedings of IJCAI 2009, pp. 53–58 (2009)
6. Betzler, N., Slinko, A., Uhlmann, J.: On the computation of fully proportional representation. Journal of Artificial Intelligence Research (2013)
7. Brams, S., Fishburn, P.: Approval voting. American Political Review 72(3), 831–847 (1978)
8. Brams, S., Fishburn, P.: Approval Voting, 2nd edn. Birkhäuser (1987)
9. Brams, S., Sanver, R.: Critical strategies under approval voting: Who gets ruled in and ruled out. Electoral Studies 25(2), 287–305 (2006)
10. Chevaleyre, Y., Lang, J., Maudet, N., Monnot, J., Xia, L.: New candidates welcome! possible winners with respect to the addition of new candidates. Mathematical Social Sciences 64(1), 74–88 (2012)
11. Darmann, A.: Popular committees. Mathematical Social Sciences (to appear, 2013)
12. Delort, C., Spanjaard, O., Weng, P.: Committee selection with a weight constraint based on a pairwise dominance relation. In: ADT, pp. 28–41 (2011)
13. Elkind, E., Lang, J., Saffidine, A.: Choosing collectively optimal sets of alternatives based on the condorcet criterion. In: IJCAI 2011, pp. 186–191 (2011)
14. Endriss, U.: Sincerity and manipulation under approval voting. Theory and Decision (2011)
15. Endriss, U., Pini, M.S., Rossi, F., Venable, K.B.: Preference aggregation over restricted ballot languages: Sincerity and strategy-proofness. In: Boutilier, C. (ed.) IJCAI 2009, pp. 122–127 (2009)
16. Erdélyi, G., Nowak, M., Rothe, J.: Sincere-strategy preference-based approval voting broadly resists control. In: Ochmański, E., Tyszkiewicz, J. (eds.) MFCS 2008. LNCS, vol. 5162, pp. 311–322. Springer, Heidelberg (2008)
17. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York (1979)
18. Kalech, M., Kraus, S., Kaminka, G.A., Goldman, C.V.: Practical voting rules with partial information. Autonomous Agents and Multiagent Systems 22(1), 151–182 (2011)
19. Kilgour, M.: Approval balloting for multi-winner elections. In: Laslier, J.-F., Sanver, R. (eds.) Handbook of Approval Voting, pp. 105–124. Springer (2010)

20. Klamler, C., Pferschy, U., Ruzika, S.: Committee selection with a weight constraint based on lexicographic rankings of individuals. In: Rossi, F., Tsoukias, A. (eds.) ADT 2009. LNCS, vol. 5783, pp. 50–61. Springer, Heidelberg (2009)
21. Konczak, K., Lang, J.: Voting procedures with incomplete preferences. In: IJCAI 2005 Multidisciplinary Workshop on Advances in Preference Handling (2005)
22. Lang, J., Pini, M.S., Rossi, F., Salvagnin, D., Venable, K.B., Walsh, T.: Winner determination in voting trees with incomplete preferences and weighted votes. In: Autonomous Agents and Multi-Agent Systems (2011)
23. Laslier, J.-F.: The leader rule – a model of strategic approval voting in a large electorate. Journal of Theoretical Politics 21, 113–136 (2009)
24. Laslier, J.-F., Sanver, R.: The basic approval voting game. In: Laslier, J.-F., Sanver, R. (eds.) Handbook of Approval Voting. Springer (2010)
25. Meir, R., Procaccia, A., Rosenschein, J., Zohar, A.: The complexity of strategic behavior in multi-winner elections. JAIR 33, 149–178 (2008)
26. Nuñez, M.: Condorcet consistency of approval voting: a counter example in large Poisson games. Journal of Theoretical Politics 22, 64–84 (2010)
27. Procaccia, A., Rosenschein, J., Zohar, A.: On the complexity of achieving proportional representation. Social Choice and Welfare 30(3), 353–362 (2008)
28. Saari, D.: Systematic analysis of multiple voting rules. Social Choice and Welfare 34(2), 217–247 (2010)
29. Sertel, M., Yılmaz, B.: The majoritarian compromise is majoritarian-optimal and subgame-perfect implementable. Social Choice and Welfare 16(4), 615–627 (1999)
30. De Sinopoli, F., Dutta, B., Laslier, J.-F.: Approval voting: three examples. International Journal of Game Theory 35(1), 27–38 (2006)
31. Skowron, P., Faliszewski, P., Slinko, A.: Achieving fully proportional representation is easy in practice. In: AAMAS 2013, pp. 399–406 (2013)
32. Xia, L., Conitzer, V.: Determining possible and necessary winners under common voting rules given partial orders. In: Proceedings of AAAI 2008, pp. 196–201 (2008)
33. Xia, L., Lang, J., Monnot, J.: Possible winners when new alternatives join: new results coming up! In: Sonenberg, L., Stone, P., Tumer, K., Yolum, P. (eds.) AAMAS. IFAAMAS, pp. 829–836 (2011)