Approval-Based Committee Voting under Incomplete Information

Aviram Imber,[1] **Jonas Israel,**[2] **Markus Brill,**[2] **Benny Kimelfeld**[1]

[1] Technion – Israel Institute of Technology, Haifa, Israel
[2] Research Group Efficient Algorithms, TU Berlin, Germany
aviram.imber@campus.technion.ac.il, j.israel@tu-berlin.de, brill@tu-berlin.de, bennyk@cs.technion.ac.il

Abstract

We investigate approval-based committee voting with incomplete information about the approval preferences of voters. We consider several models of incompleteness where each voter partitions the set of candidates into *approved*, *disapproved*, and *unknown* candidates, possibly with ordinal preference constraints among candidates in the latter category. This captures scenarios where voters have not evaluated all candidates and/or it is unknown where voters draw the threshold between approved and disapproved candidates. We study the complexity of some fundamental computational problems for a number of classic approval-based committee voting rules including Proportional Approval Voting and Chamberlin–Courant. These problems include that of determining whether a given set of candidates is a possible or necessary winning committee and whether it forms a committee that possibly or necessarily satisfies representation axioms. We also consider the problem whether a given candidate is possibly or necessarily a member of the winning committee.

1 Introduction

Approval-based committee (ABC) voting represents a well-studied multiwinner election setting, where a subset of candidates of a predetermined size, a so-called *committee*, needs to be chosen based on the approval preferences of a set of voters (Lackner and Skowron 2021). Traditionally, ABC voting is studied in the context where we know, for each voter and each candidate, whether the voter approves the candidate or not. In this paper, we investigate the situation where the approval information is incomplete. Specifically, we assume that each voter is associated with a set of *approved* candidates, a set of *disapproved* candidates, and a set of candidates where the voter's stand is unknown, hereafter referred to as the *unknown* candidates. Moreover, we may have (partial) ordinal information on voters' preferences among the unknown candidates, restricting the "valid" completions of voters' approval sets.

When the number of candidates is large, unknown candidates are likely to exist because voters are not aware of or not familiar with, and therefore cannot evaluate, all candidates. In particular, this holds in scenarios where candidates join

the election over time, and voter preferences over new candidates have not been elicited (Chevaleyre et al. 2012). Distinguishing between disapproved and unknown candidates accounts for two fundamentally different reasons for non-approvals: either the voter has evaluated the candidate and judged him or her not worth approving (in which case the candidate counts as disapproved), or the voter has not evaluated the candidate (in which case the candidate counts as unknown). Furthermore, incorporating (partial) ordinal preferences among unknown candidates allows us to model situations in which we (partially) know how a voter rank-orders the candidates, but we do not know where the voter draws his or her "approval threshold."

Scenarios involving incomplete knowledge about approval preferences arise naturally in a variety of practical settings. For example, in a scenario where we retrieve information from indirect sources such as social media (say, for the sake of prediction), we may get only sparse information about approval ("Vote for X") and disapproval ("Definitely not Y"), and possibly pairwise preferences due to explicit statements (e.g., "X is at least better than Y"). As another example, in shortlisting scenarios (such as faculty hiring) some voters may know with sufficient confidence that they support or oppose some candidates (e.g., the ones from their own field of expertise) but have no clear opinion on others. This situation also naturally arises when labeling documents for information retrieval, where the committee corresponds to the page of search results: some documents are clearly relevant, some clearly irrelevant, and some are unclear. For the unclear ones, it may be way easier to rank the documents (totally or partially) by relevance rather than to insist on a complete classification into relevant and irrelevant; in fact, this is the motivation behind some successful methodologies for learning ranking functions (*learning to rank*) such as the pairwise and listwise approaches (Cao et al. 2007).

Our basic model of incompleteness is the *poset approval* model that is illustrated in Figure 1(a). This model is a rather direct generalization of the voter model in the seminal work of Konczak and Lang (2005), who study problems of winner determination with incomplete preferences under ranking-based single-winner rules (such as plurality and Borda). For each voter, we are given a set of *approved* candidates and a set of *disapproved* candidates, together with a partial order over the remaining ("unknown") candidates that con-

strains the possible approval ballots of the voter: if an unknown candidate is preferred to another unknown candidate, then the former needs to be approved by the voter whenever the latter is approved. In other words, each possible world is obtained by selecting a linear extension of the partial order and determining a cutoff point—every candidate before the cutoff is approved, and every candidate after the cutoff is disapproved (in addition to the known approvals and disapprovals, respectively). We study in depth two special cases of this model that correspond to the two extremes of posets: zero information (3VA) and full information (linear).

- In the *three-valued approval* (3VA) model,[1] illustrated in Figure 1(b), the partial order of the poset is empty. Hence, every subset of unknown candidates determines a valid possible world where this set is approved and its complement is disapproved.

- In the *linear* model, illustrated in Figure 1(c), the unknown candidates are ordered linearly. Hence, every prefix of this order determines a possible world where the candidates in this prefix are approved and the ones in the remaining suffix are disapproved.

From the computational perspective, the models are fundamentally different. For instance, a voter can have an exponential number of possible worlds (or *completions*) in one case, but only a linear number in the other. Of course, when a problem is tractable in the poset approval model, then it is also tractable in the 3VA and linear models. On the contrapositive, whenever a problem is intractable (e.g., NP-hard) in one of these two models, it is also intractable for the general poset approval model. However, as we illustrate later, a problem may be tractable in 3VA and intractable in the linear model, and vice versa.

We investigate the computational complexity of fundamental problems that arise in ABC voting with incomplete approval profiles. In the first problem, the goal is to determine whether a given set W of candidates is a *possible* winning committee. In the second problem, the goal is to determine whether such given W is a *necessary* winning committee. More formally, a *possible committee* (respectively, *necessary committee*) is a set of candidates that is a winning committee in *some* completion (respectively, *all* completions) of the incomplete approval profile. We also achieve some results on the problem where we are given a candidate and the goal is to determine whether the candidate is possibly or necessarily a member of a winning committee.

In the applications described earlier it is often enough to find only one winning committee, e.g., because we only want to have one set of top search results and do not care about there being other, equally good such sets. Thus finding a necessary winning committee is enough and we do not need to elicit any more information. Conversely, if for example in a hiring process we know that some candidates are not in any possible committee any more, we can already inform them that they will definitely not get the position and we can stop eliciting preferences over those candidates.

We consider a class of voting rules that was introduced by Thiele (1895). This broad class of approval-based committee voting rules contains several classic rules such as Approval Voting (AV), Chamberlin–Courant (CC), and Proportional Approval Voting (PAV). For all Thiele rules except AV, the problem of determining winning committees is intractable even before we allow for incompleteness (Lackner and Skowron 2021). Consequently, computing either possible or necessary committees is intractable as well, under each of our models of incompleteness. Therefore, we focus on the case where the committee size is small (i.e., constant). In this case, winning committees can be computed in polynomial time for all considered rules.

In addition, we study a problem of fair representation of voters given incomplete approvals. In particular, we focus on justified representation (JR) (Aziz et al. 2017), which requires that all voter groups that are large enough and cohesive enough are represented in the committee. We investigate the complexity of deciding whether a given set of candidates possibly or necessarily forms a committee that satisfies JR given an incomplete approval profile. We also briefly discuss stronger notions of proportional representation.

Related work Collective decision making under incomplete knowledge has been extensively studied; we refer to Lang (2020) for a recent survey. Most of that work focuses on single-winner elections (e.g., Xia and Conitzer 2011), with the paper by Lu and Boutilier (2013) representing a notable exception. To the best of our knowledge, there are only two papers that deal with incomplete information in the context of approval voting: First, Barrot et al. (2013) study single-winner approval voting and multiwinner approval voting (which we refer to as AV) given incomplete information. In the terminology of our paper, their incompleteness model corresponds to the special case of the linear model in which all candidates (except the most preferred one) belong to the unknown category. We discuss specific relationships between their work and ours in Section 4. Second, Terzopoulou, Karpov, and Obraztsova (2021) consider restricted domains for approval voting under uncertainty using a model of incompleteness corresponding to our 3VA model. Allowing voters to not only specify which candidates they approve and disapprove but also for which they are undecided can be seen as eliciting trichotomous preferences (without uncertainty). In this vein, our work relates to Felsenthal (1989), Alcantud and Laruelle (2014), Baumeister and Dennisen (2015), and Zhou, Yang, and Guo (2019).

2 Preliminaries

We consider a voting setting where a finite set V of *voters* has preferences over a finite set C of *candidates*. We usually use n for the number of voters and we enumerate voters as $v_1, v_2, \ldots, v_n$. The number of candidates is denoted by m. An *approval profile* $\mathbf{A} = (A(v_1), \ldots, A(v_n))$ lists the approval sets of the voters, where $A(v_i) \subseteq C$ denotes the set of candidates that are approved by voter v_i. The concatenation of two approval profiles $\mathbf{A}_1 = (A_1(v_1), \ldots, A_1(v_p))$ and $\mathbf{A}_2 = (A_2(v'_1), \ldots, A_2(v'_q))$ is denoted by $\mathbf{A}_1 \circ \mathbf{A}_2 := (A_1(v_1), \ldots, A_1(v_p), A_2(v'_1), \ldots, A_2(v'_q))$. An *approval-*

---

[1]The term is analogous to "Three-Valued Logic" (3VL) that is adapted, e.g., in SQL (Libkin 2016), where the three values are *true*, *false* and *unknown*.

(a) Poset approval

(b) Three-valued approval (3VA)
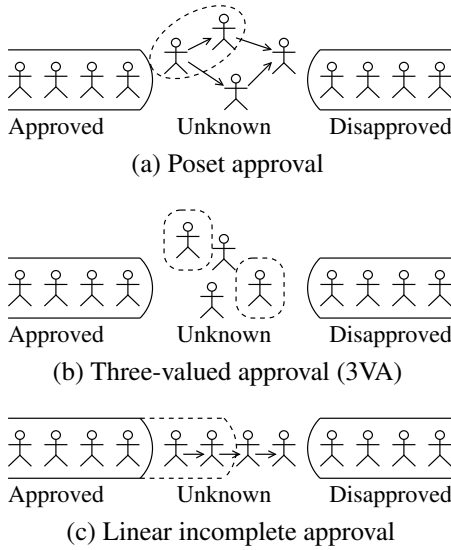

(c) Linear incomplete approval

Figure 1: Models of incomplete approval preferences: Poset approval and the special cases of 3VA and linear incomplete approval. Dashed frames depict candidates that are approved in a valid completion.

based committee (ABC) rule takes an approval profile as input and outputs one or more size-k subsets of candidates, so called *committees*, for a given parameter $k \in \mathbb{N}$.

Most of our work is focused on a class of ABC rules known as *Thiele rules* (Thiele 1895). These rules are parameterized by a *weight function* w, i.e., a non-decreasing function $w \colon \mathbb{N} \to \mathbb{Q}_{\geq 0}$ with $w(0) = 0$. The score that a voter v contributes to a subset $S \subseteq C$ is defined as $w(|S \cap A(v)|)$ and the score of S is $\sum_{v \in V} w(|S \cap A(v)|)$. The rule w-*Thiele* then outputs the subset(s) of size k with the highest score. Two of the most famous examples of Thiele rules are *Approval Voting (AV)*,[2] where $w(x) = x$, and *Proportional Approval Voting (PAV)*, where $w(x) = \sum_{i=1}^{x} 1/i$. We assume that $w(x)$ is computable in polynomial time in x, and that there exists an integer x with $w(x) \neq 0$ (otherwise, all subsets of voters have the same score). Observe that for every *fixed* number k, we can find the committee(s) under w-Thiele in polynomial time by computing the score of every subset $W \subseteq C$ of size k.

Without loss of generality, we assume that weight functions are normalized such that $w(x) = 1$ for the smallest x with $w(x) > 0$. We call a Thiele rule *binary* if $w(x) \in \{0, 1\}$ for all $x \in \mathbb{N}$. Since w is non-decreasing, binary Thiele rules are characterized by an integer t such that $w(x) = 0$ for every $x < t$ and $w(x) = 1$ for all $x \geq t$. The most prominent binary Thiele rule is *Chamberlin–Courant (CC)*, with $w(x) = 1$ for all $x \geq 1$.

---

[2]We can also define this rule using scores of individual candidates: For a candidate c, the *approval score* of c is defined as $|\{v \in V : c \in A(v)\}|$. AV selects committees consisting of the k candidates with highest approval scores.

3 Models of Incompleteness

In this section, we formally introduce the three models of incomplete information that we investigate. We first define *poset approval* as a general model, and then two models constituting special cases. In all models, a *partial profile* $\mathbf{P} = (P(v_1), \dots, P(v_n))$ consists of n *partial votes*.

Poset approval. For each voter v, the partial vote $P(v)$ consists of a partition $(\mathrm{Top}(v), \mathrm{Middle}(v), \mathrm{Bottom}(v))$ of C into three sets and a partial order[3] $\succ_v$ over the candidates in $\mathrm{Middle}(v)$. $\mathrm{Top}(v)$ represents the candidates that v approves in any case, $\mathrm{Bottom}(v)$ represents the candidates that v approves in no case. The partial order $\succ_v$ represents approval constraints on the candidates in $\mathrm{Middle}(v)$, whose approval is open. Specifically, if $c \succ_v c'$ and v approves c', then v also approves c. A *completion* of $\mathbf{P} = (P(v_1), \dots, P(v_n))$ is an approval profile $\mathbf{A} = (A(v_1), \dots, A(v_n))$ where each $A(v)$ "completes" $P(v)$. Namely, $\mathrm{Top}(v) \subseteq A(v)$, $\mathrm{Bottom}(v) \cap A(v) = \emptyset$, and for every pair $c, c' \in \mathrm{Middle}(v)$ with $c \succ_v c'$, it holds that $c \in A(v)$ whenever $c' \in A(v)$. Equivalently, we can describe $\mathbf{A}$ in the following way: For each voter v select a subset $\mathrm{Middle}_{\mathbf{A}}(v) \subseteq \mathrm{Middle}(v)$ which "respects" $\succ_v$ (i.e., $c \succ_v c'$ and $c' \in \mathrm{Middle}_{\mathbf{A}}(v)$ imply $c \in \mathrm{Middle}_{\mathbf{A}}(v)$) and v approves exactly $A(v) = \mathrm{Top}(v) \cup \mathrm{Middle}_{\mathbf{A}}(v)$.

Three-valued approval (3VA). This model is a special case of poset approval where for every voter v, the partial order $\succ_v$ over the candidates in $\mathrm{Middle}(v)$ is empty (except for the reflexive part). In other words, v might approve any subset of candidates in $\mathrm{Middle}(v)$. Hence, a completion of $\mathbf{P} = (P(v_1), \dots, P(v_n))$ is a complete approval profile $\mathbf{A} = (A(v_1), \dots, A(v_n))$ such that for every voter v we have $\mathrm{Top}(v) \subseteq A(v)$ and $\mathrm{Bottom}(v) \cap A(v) = \emptyset$.

Linear incomplete approval. This model is a special case of poset approval where for every voter v, $\succ_v$ is a complete linear order $c_{i_1} \succ_v \dots \succ_v c_{i_q}$ on the candidates in $\mathrm{Middle}(v)$. To construct a completion $\mathbf{A}$, for every voter v we select a $j \leq |\mathrm{Middle}(v)|$ such that $\mathrm{Middle}_{\mathbf{A}}(v) = \{c_{i_1}, \dots, c_{i_j}\}$.

In the linear model of incompleteness, every partial vote has at most m completions, as opposed to the previous two models where even a single partial vote can have 2^m completions. In all three models, the number of completions for a partial profile can be exponential in the number of voters n.

In order to define possible and necessary committees, fix an ABC voting rule r. Given a committee size k and a partial profile $\mathbf{P}$, a set $W \subseteq C$ of k candidates is a *possible committee* if there exists a completion $\mathbf{A}$ of $\mathbf{P}$ where W is a winning committee (under r), and a *necessary committee* if W is a winning committee for every completion $\mathbf{A}$ of $\mathbf{P}$. In the following sections, we investigate the computational complexity of deciding whether a given set of candidates is a possible or necessary committee, for different rules r.

4 Computing Possible Committees

The first decision problem we study concerns possible committees. For an ABC rule r and committee size k, con-

---

[3]Recall that a *partial order* is a reflexive, antisymmetric, and transitive binary relation.

sider the following decision problem that we denote by PosCom$\langle k \rangle$: Given a partial profile $\mathbf{P}$ and a subset $W \subseteq C$ of candidates of size k, decide whether W is a possible committee. As mentioned earlier, we parameterize the decision problem by k since otherwise winner determination even in the case of no uncertainty is known to be NP-hard for almost all Thiele rules (Lackner and Skowron 2021). Observe that PosCom$\langle k \rangle$ is in NP under all Thiele rules, for every fixed k, in all three models of uncertainty that we study. This is because, given a complete profile, we can find the winning committees of a fixed size k under Thiele rules in polynomial time.

Our results concerning the complexity of deciding PosCom$\langle k \rangle$ under different models of uncertainty are summarized in Table 1. In particular, we show that under the model of poset approval, PosCom$\langle k \rangle$ is NP-complete for all Thiele rules. To obtain this result, we study the complexity of PosCom$\langle k \rangle$ in the models of three-valued approval and linear incomplete approval in Sections 4.1 and 4.2, respectively. Then, we prove our main result in Section 4.3.

We first introduce a useful lemma that holds in all three models of uncertainty. The complete proofs of all results can be found in the full version of the paper (Imber et al. 2021).

Lemma 1. *Let w_1, w_2 be a pair of weight functions. Assume there are three integers $k, q > 0$ and $t \geq 0$ such that $w_1(x) = q \cdot (w_2(x+t) - w_2(t))$ for every $x \in \{0, 1, \ldots, k\}$. Then, in each of the three models of uncertainty, there is a polynomial-time reduction from PosCom$\langle k \rangle$ under w_1-Thiele to PosCom$\langle k + t \rangle$ under w_2-Thiele.*

Proof sketch. Let $\mathbf{P}_1 = (P_1(v_1), \ldots, P_1(v_n))$ and $W \subseteq C_1$ of size k be an instance of PosCom$\langle k \rangle$ under w_1-Thiele. Define an instance of PosCom$\langle k + t \rangle$ under w_2-Thiele where the candidates are $C_2 = C_1 \cup D$ for $D = \{d_1, \ldots, d_t\}$. The profile is $\mathbf{P}_2 = (P_2(v_1), \ldots, P_2(v_n))$, where for every voter v, $P_2(v)$ is the same as $P_1(v)$ except that v always approves the candidates of D. Formally, we add the candidates of D to $\text{Top}(v)$, and the rest is unchanged. It can be shown that W is a possible committee (of size k) for $\mathbf{P}_1$ under w_1-Thiele if and only if $W \cup D$ is a possible committee (of size $k + t$) for $\mathbf{P}_2$ under w_2-Thiele. $\square$

4.1 Three-Valued Approval

We now discuss the complexity of PosCom$\langle k \rangle$ in the 3VA model for different Thiele rules. We start with Approval Voting and Chamberlin–Courant.

Theorem 1. *In the 3VA model, PosCom$\langle k \rangle$ is solvable in polynomial time under Approval Voting, for all fixed $k \geq 1$.*

Proof sketch. Given $\mathbf{P} = (P(v_1), \ldots, P(v_n))$ and a subset $W \subseteq C$ of $|W| = k$ candidates, define a completion $\mathbf{A} = (A(v_1), \ldots, A(v_n))$ where for every voter v, $\text{Middle}_{\mathbf{A}}(v) = \text{Middle}(v) \cap W$; that is, except for $\text{Top}(v)$, voters only approve candidates in W. It can be shown that W is a possible committee for $\mathbf{P}$ if and only if W is a winning committee in $\mathbf{A}$. $\square$

Theorem 2. *In the 3VA model, PosCom$\langle 2 \rangle$ is NP-complete under Chamberlin–Courant.*

Proof sketch. We show NP-hardness by a reduction from *one-in-three positive 3SAT*, which is the following decision problem: Let $X = \{x_1, \ldots, x_n\}$ be a set of elements. Given sets $S_1, \ldots, S_m \subseteq X$, where $|S_i| = 3$ for every i, is there a subset $B \subseteq X$ such that for each i, $|B \cap S_i| = 1$? This problem is NP-complete (Garey and Johnson 1979).

Given X and $\mathcal{S} = \{S_1, \ldots, S_m\}$, we define an instance of PosCom$\langle 2 \rangle$ under Chamberlin–Courant. The candidates are $C = \mathcal{S} \cup \{w_1, w_2\}$. The partial profile $\mathbf{P} = \mathbf{P}_1 \circ \mathbf{A}_2$ consists of two parts that we describe next. The first part, $\mathbf{P}_1$, consists of a voter for every element in X. Let $x \in X$, and let $S(x) = \{S_i \in \mathcal{S} : x \in S_i\}$ be the sets of $\mathcal{S}$ that contain x. Define a voter v_x with $\text{Top}(v_x) = \mathcal{S} \setminus S(x)$, $\text{Middle}(v_x) = \{w_1, w_2\}$ and $\text{Bottom}(v_x) = S(x)$. The decision whether to approve w_1 or w_2 represents the decision whether to include x in the subset $B \subseteq X$. The second part, $\mathbf{A}_2$, consists of six voters without uncertainty (voters with $\text{Middle}(v) = \emptyset$). We add three voters approving $\mathcal{S}$, two voters approving $\{w_1\}$, and a single voter approving $\{w_2\}$. It can be shown that $\{w_1, w_2\}$ is a possible committee if and only if there is a solution to the instance of one-in-three positive 3SAT. $\square$

Next, we use Lemma 1 and Theorem 2 to obtain another hardness result.

Theorem 3. *Let w be a weight function such that $w(k-2) < w(k-1) = w(k)$ for some $k \geq 2$. Then, in the 3VA model, PosCom$\langle k \rangle$ is NP-complete under w-Thiele.*

Whereas the condition in Theorem 3 does not hold for PAV (for which the complexity of PosCom$\langle k \rangle$ in the 3VA model remains open), it clearly holds for binary Thiele rules.

Corollary 1. *For every binary Thiele rule, there exists $k \geq 2$ such that PosCom$\langle k \rangle$ is NP-complete in the 3VA model.*

4.2 Linear Incomplete Approval

Next, we discuss the complexity of PosCom$\langle k \rangle$ in the linear model. We start with binary Thiele rules.

Theorem 4. *In the linear model, PosCom$\langle k \rangle$ is solvable in polynomial time for binary Thiele rules, for all fixed $k \geq 1$.*

Proof sketch. Consider a binary Thiele rule and let t be such that $w(x) = 0$ for all $x < t$ and $w(x) = 1$ for all $x \geq t$. Without loss of generality, we can assume that $k \geq t$. Given $\mathbf{P} = (P(v_1), \ldots, P(v_n))$ and a subset $W \subseteq C$ of k candidates we construct a completion $\mathbf{A}$ as follows. Let $c_{i_1} \succ_v \cdots \succ_v c_{i_q}$ be the linear order of voter v over the candidates in $\text{Middle}(v)$. If $|\text{Top}(v) \cap W| \geq t$, then v always contributes a score of 1 to W. Since approving more candidates cannot increase the score of W, we select $\text{Middle}_{\mathbf{A}}(v) = \emptyset$. If $|(\text{Top}(v) \cup \text{Middle}(v)) \cap W| < t$, then v always contributes a score of 0 to W, and we select $\text{Middle}_{\mathbf{A}}(v) = \emptyset$ again. Finally, assume that $|\text{Top}(v) \cap W| < t$ and $|(\text{Top}(v) \cup \text{Middle}(v)) \cap W| \geq t$. Let j be minimal with $|(\text{Top}(v) \cup \{c_{i_1}, \ldots, c_{i_j}\}) \cap W| \geq t$ and $\text{Middle}_{\mathbf{A}}(v) = \{c_{i_1}, \ldots, c_{i_j}\}$. It can be shown that W is a possible committee if and only if W is a winning committee in $\mathbf{A}$. $\square$

We now turn to the complexity of PosCom$\langle k \rangle$ under non-binary w-Thiele rules. The next theorem states that

Model	AV	CC	PAV	w-Thiele
Three-Valued Approval	P [Thm. 1]	NP-c [Thm. 2]	?	NP-c for every binary w [Cor. 1]
Linear Incomplete Approval	NP-c [Cor. 2]* (for all $k \geq 2$)	P [Thm. 4]	NP-c [Cor. 2] (for all $k \geq 2$)	P for every binary w [Thm. 4]; NP-c for every strictly increasing w [Cor. 2] (for all $k \geq 2$)
Poset Approval	NP-c [Thm. 7]	NP-c [Thm. 7]	NP-c [Thm. 7]	NP-c for every w [Thm. 7]

Table 1: Overview of our complexity results for $\mathsf{PosCom}\langle k \rangle$. "NP-c" means that $\mathsf{PosCom}\langle k \rangle$ is NP-complete for at least one k, unless otherwise stated. "P" means that $\mathsf{PosCom}\langle k \rangle$ is solvable in polynomial time for all k. The result marked with * also follows from Barrot et al. (2013).

$\mathsf{PosCom}\langle 2 \rangle$ is NP-complete under w-Thiele for every function w whose first three values are pairwise distinct.

Theorem 5. *Let w be a weight function with $w(2) > w(1) > 0$. Then, in the linear model, $\mathsf{PosCom}\langle 2 \rangle$ is NP-complete under w-Thiele.*

Proof sketch. Without loss of generality we have $w(1) = 1$ and $w(2) = 1 + x$ where $x > 0$. We show two reductions, depending on whether $x \in (0, 1]$ or $x > 1$, from *exact cover by 3-sets* (X3C), which is the following decision problem: Given a set $U = \{u_1, \ldots, u_{3q}\}$ and a collection $E = \{e_1, \ldots, e_m\}$ of 3-element subsets of U, can we cover all the elements of U using q pairwise disjoint sets from E? This problem is NP-complete (Garey and Johnson 1979).

We start with the case where $x \in (0, 1]$. Given U and E, we construct an instance of $\mathsf{PosCom}\langle 2 \rangle$ under w-Thiele. The candidate set is $C = U \cup \{c, d, z\}$. The partial profile $\mathbf{P} = \mathbf{P}_1 \circ \mathbf{A}_2$ is the concatenation of the following two parts. The first part, $\mathbf{P}_1$, consists of a voter for every set in E. For every $e \in E$, voter v_e has $\mathrm{Top}(v_e) = \emptyset$, $\mathrm{Middle}(v_e) = e \cup \{c\}$, and $\mathrm{Bottom}(v_e) = (U \setminus e) \cup \{d, z\}$. The linear order on $\mathrm{Middle}(v_e)$ is an arbitrary order that ranks c last. This means that if v_e approves c, then it also approves the candidates of e. The idea is that approving c indicates that e is in the cover, and disapproving c indicates that e is not in the cover. The second part, $\mathbf{A}_2$, consists of $2q + 1$ voters without uncertainty (voters with $\mathrm{Middle}(v) = \emptyset$). Assume w.l.o.g. that q is even. We add $q/2$ voters approving $\{z\}$, $q/2$ voters approving $\{z\} \cup U$, $q/2$ voters approving $\{d\}$, $q/2$ voters approving $\{d\} \cup U$, and a single voter approving $\{c, d, z\}$. It can be shown that $\{c, d\}$ is a possible committee for $\mathbf{P}$ if and only if there exists an exact cover. For the case $x > 1$ a similar reduction is possible. $\square$

Next, we use Lemma 1 in order to generalize Theorem 5 to every function with three consecutive different values.

Theorem 6. *Let w be a weight function such that $w(k-2) < w(k-1) < w(k)$ for some $k \geq 2$. Then, in the linear model, $\mathsf{PosCom}\langle k \rangle$ is NP-complete under w-Thiele.*

Clearly, Theorem 6 applies to all Thiele rules with strictly increasing weight functions (such as AV and PAV).

Corollary 2. *Let w be a strictly increasing weight function. Then, in the linear model, $\mathsf{PosCom}\langle k \rangle$ is NP-complete under w-Thiele for every $k \geq 2$.*

As a special case, Corollary 2 reasserts the result of Barrot et al. (2013) that $\mathsf{PosCom}\langle k \rangle$ is NP-complete for the AV rule. (The hardness result by Barrot et al. (2013) even applies to the special case in which all candidates are unknown.)

4.3 Poset Approval

Combining the results from the previous two sections, we can now formulate the main result of this paper.

Theorem 7. *In the poset approval model, for every weight function w there exists an integer k such that $\mathsf{PosCom}\langle k \rangle$ is NP-complete under w-Thiele.*

Proof. Let w be a weight function, and let j be the minimal index such that $w(j) > 0$. Since w is non-decreasing, either $w(j+1) = w(j) > w(j-1)$ or $w(j+1) > w(j) > w(j-1)$. In the first case, by Theorem 3, $\mathsf{PosCom}\langle j+1 \rangle$ is NP-complete under w-Thiele in the 3VA model. In the second case, by Theorem 6, $\mathsf{PosCom}\langle j+1 \rangle$ is NP-complete under w-Thiele in the linear model. The result follows since 3VA and the linear model are special cases of poset approval. $\square$

4.4 Possible Committee Members

We now consider a different decision problem where we do not focus on whole committees but rather on individual candidates. We say candidate c is a *possible committee member* if there exists a completion $\mathbf{A}$ of $\mathbf{P}$ such that there exists a winning committee W with $c \in W$. For an ABC rule r and a committee size k, we consider the computational problem $\mathsf{PosMem}\langle k \rangle$, where the input consists of a partial profile $\mathbf{P}$ and a candidate c, and the goal is to determine whether c is a possible committee member under r.

Note that for every voting rule, if $\mathsf{PosCom}\langle k \rangle$ is solvable in polynomial time, then we can also solve $\mathsf{PosMem}\langle k \rangle$ efficiently: given a candidate c, test whether $\{c, d_1, \ldots, d_{k-1}\}$ is a possible committee for all other $k - 1$ candidates $d_1, \ldots, d_{k-1}$. Since k is fixed, this can be done efficiently. By Theorems 1 and 4 we can deduce the following.

Corollary 3. *For all fixed $k \geq 1$, $\mathsf{PosMem}\langle k \rangle$ is solvable in polynomial time:*

1. *In the 3VA model under approval voting;*
2. *In the linear model under every binary Thiele rule.*

We also obtain a tractability result in a setting for which the corresponding $\mathsf{PosCom}\langle k \rangle$ problem is intractable.

Theorem 8. *In the linear model,* PosMem$\langle k \rangle$ *is solvable in polynomial time under Approval Voting, for all fixed $k \geq 1$.*

Proof sketch. Given a candidate c and a partial profile $\mathbf{P} = (P(v_1), \ldots, P(v_n))$, we construct a completion $\mathbf{A}$ of $\mathbf{P}$ as follows. For all voters v with $c \in \text{Top}(v)$ or $c \in \text{Bottom}(v)$, we let $\text{Middle}_{\mathbf{A}}(v) = \emptyset$. Otherwise, $c \in \text{Middle}(v)$ and we define $\text{Middle}_{\mathbf{A}}(v)$ to be the smallest prefix of the linear order over $\text{Middle}(v)$ that includes c. It can be shown that c is a possible committee member if and only if c is a member of a committee in $\mathbf{A}$. $\square$

5 Computing Necessary Committees

The next decision problem we study concerns necessary winning committees. For an ABC rule r and a committee size k, we consider the following decision problem, which we refer to as NecCom$\langle k \rangle$: Given a partial profile $\mathbf{P}$ and a subset $W \subseteq C$ of candidates of size $|W| = k$, determine whether W is a necessary committee under r. As above, we parametrize the problem by the committee size k to evade hardness even in the case of complete information.

We show that NecCom$\langle k \rangle$ is solvable in polynomial time for a large family of ABC rules that includes all Thiele rules. Each member of the family of *ABC scoring rules* is associated with a *scoring function* $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ satisfying $f(x, y) \geq f(x', y)$ for $x \geq x'$. The score s of a subset $S \subseteq C$ given approval set $A(v)$ is

$$s(A(v), S) = f(|A(v) \cap S|, |A(v)|),$$

and the total score is $s(\mathbf{A}, S) = \sum_{v \in V} s(A(v), S)$. The rule defined by f outputs the subset(s) of size k with the highest score. Observe that Thiele rules are a special case of ABC scoring rules, where the score $s(A(v), S)$ that v assigns to S depends only on $|A(v) \cap S|$. An example for an ABC scoring rule which is not a Thiele rule is *Satisfaction Approval Voting* (Brams and Kilgour 2014), where the score a voter v contributes to a $S \subseteq C$ is $s(A(v), S) = |S \cap A(v)|/|A(v)|$. When we discuss scoring rules, we assume that $f(x, y)$ is computable in polynomial time given x and y.

The following positive result covers all ABC scoring rules and all considered models of incompleteness.

Theorem 9. *Let r be an ABC scoring rule. Then, in the model of poset approval, for all fixed $k \geq 1$,* NecCom$\langle k \rangle$ *is solvable in polynomial time under r.*

In order to prove Theorem 9, observe that a subset $W \subseteq C$ of size k is *not* a necessary committee if and only if there exists a completion where another size-k subset has a score strictly greater than the score of W. Formally, let $\mathbf{P}$ be a partial profile and let $\mathcal{W}_k = \{W' \subseteq C : |W'| = k\}$ be the set of subsets of size k. For a completion $\mathbf{A}$ of $\mathbf{P}$ and $W, W' \in \mathcal{W}_k$, define the *score difference* as

$$\Delta_{\mathbf{A}}(W, W') = s(\mathbf{A}, W') - s(\mathbf{A}, W).$$

When $\mathbf{A}$ consists of a single set $A(v)$ we write $\Delta_{A(v)}(W, W')$ instead of $\Delta_{\mathbf{A}}(W, W')$.

Let $\mathcal{C}(\mathbf{P})$ be the set of completions of $\mathbf{P}$. Again, when the profile consists of a single voter v, we write $\mathcal{C}(P(v))$ instead

of $\mathcal{C}(\mathbf{P})$. Define the *maximal score difference* as

$$\overline{\Delta}_{\mathbf{P}}(W) := \max_{\mathbf{A} \in \mathcal{C}(\mathbf{P}), W' \in \mathcal{W}_k} \Delta_{\mathbf{A}}(W, W').$$

Observe that W is *not* a necessary committee if and only if $\overline{\Delta}_{\mathbf{P}}(W) > 0$. To compute $\overline{\Delta}_{\mathbf{P}}(W)$, we iterate over all sets $W' \in \mathcal{W}_k$, and for every W' we compute $\overline{\Delta}_{\mathbf{P}}(W, W') := \max_{\mathbf{A} \in \mathcal{C}(\mathbf{P})} \Delta_{\mathbf{A}}(W, W')$. Once we compute $\overline{\Delta}_{\mathbf{P}}(W, W')$ for every $W' \in \mathcal{W}_k$ we are done, since $\overline{\Delta}_{\mathbf{P}}(W) = \max_{W' \in \mathcal{W}_k} \overline{\Delta}_{\mathbf{P}}(W, W')$.

Now, let $W' \in \mathcal{W}_k$. Since r is an ABC scoring rule, the score of W' in a completion $\mathbf{A}$ is $s(\mathbf{A}, W') = \sum_{v \in V} s(A(v), W')$, therefore

$$\overline{\Delta}_{\mathbf{P}}(W, W') = \max_{\mathbf{A} \in \mathcal{C}(\mathbf{P})} \sum_{v \in V} \Delta_{A(v)}(W, W')$$

$$= \sum_{v \in V} \max_{A(v) \in \mathcal{C}(P(v))} \Delta_{A(v)}(W, W') = \sum_{v \in V} \overline{\Delta}_{P(v)}(W, W').$$

The following lemma completes the proof of Theorem 9 by showing that $\overline{\Delta}_{P(v)}(W, W')$ can be computed in polynomial time for each voter $v \in V$.

Lemma 2. *Let r be an ABC scoring rule and k a fixed natural number. Computing $\overline{\Delta}_{P(v)}(W, W')$ is possible in polynomial time given a partial vote $P(v)$ over a set C of candidates, and a pair of sets $W, W' \in \mathcal{W}_k$.*

Proof sketch. Set $S = \text{Middle}(v) \cap (W \cup W')$. We call a subset $R \subseteq S$ *feasible* if there exists a completion $A(v)$ of $P(v)$ where $\text{Middle}_{\mathbf{A}}(v) \cap S = R$. For feasible $R \subseteq S$, set

$$\overline{\Delta}_{P(v)}^R(W, W') = \max_{\substack{A(v) \in \mathcal{C}(P(v)): \\ \text{Middle}_{\mathbf{A}}(v) \cap S = R}} \Delta(W, W', A(v)).$$

Otherwise, define $\overline{\Delta}_{P(v)}^R(W, W') = -\infty$. We get that $\overline{\Delta}_{P(v)}(W, W') = \max_{R \subseteq S} \overline{\Delta}_{P(v)}^R(W, W')$. Since $|S| \leq 2k$, there is a constant number of subsets of S. It can be shown that we can compute $\overline{\Delta}_{P(v)}^R(W, W')$ in polynomial time for every $R \subseteq S$. $\square$

Necessary Committee Members. A candidate c is a *necessary committee member* if for every completion $\mathbf{A}$ of $\mathbf{P}$ there exists a winning committee W for which $c \in W$. For an ABC rule r and a committee size k, we consider the computational problem NecMem$\langle k \rangle$, where the input consists of a partial profile $\mathbf{P}$ and a candidate c, and the goal is to determine whether c is a necessary committee member.

Note that a necessary committee member can be a member of different winning committees in different completions. Thus, in contrast to PosCom and PosMem, being able to solve NecCom does not help deciding NecMem, because a candidate could be a necessary committee member without being a member of a necessary committee.

We show that in the 3VA model we can find the necessary members in polynomial time under approval voting.

Theorem 10. *In the 3VA model,* NecMem$\langle k \rangle$ *is solvable in polynomial time under Approval Voting, for all fixed $k \geq 1$.*

In the linear model, NecMem$\langle k \rangle$ is solvable in polynomial time under AV and under binary Thiele rules.

Theorem 11. *In the linear model,* NecMem$\langle k \rangle$ *is solvable in polynomial time under Approval Voting and under every binary Thiele rule, for all fixed $k \geq 1$.*

6 Representation under Incompleteness

An important goal in committee voting concerns the representation of voters. Sufficiently large and cohesive groups of voters should be adequately represented in the committee. In this section, we investigate the problem of representation under incomplete preference information: Given a committee W and partial profile $\mathbf{P}$, does W possibly or necessarily provide some form of representation? We mainly focus on *justified representation* (Aziz et al. 2017), and discuss stronger representation axioms at the end of this section.

For a (complete) approval profile $\mathbf{A}$ and committee size k, a committee $W \subseteq C$ of size $|W| = k$ satisfies *justified representation (JR) w.r.t.* $\mathbf{A}$ if for all $G \subseteq V$ with $|G| \geq \frac{n}{k}$ and $|\bigcap_{v \in G} A(v)| \neq \emptyset$, it holds that $|W \cap \bigcup_{v \in G} A(v)| \neq \emptyset$.

Aziz et al. (2017) show that it can be tested in polynomial time whether a given committee provides JR w.r.t. a given approval profile (without incompleteness). Given a partial profile $\mathbf{P}$ and a committee W, we consider two decision problems. In PosJR, we ask whether there exists a completion $\mathbf{A}$ of $\mathbf{P}$ such that W satisfies JR w.r.t. $\mathbf{A}$. In NecJR we ask whether W satisfies JR w.r.t. every completion of $\mathbf{P}$. In contrast to the decision problems we considered earlier, the committee size k is part of the input for both problems.

We first prove that a committee satisfying JR w.r.t. a given approval profile $\mathbf{A}$ also satisfies JR w.r.t. a new approval profile $\mathbf{A}'$ that results from modifying $\mathbf{A}$ in certain ways.

Lemma 3. *Let $W \subseteq C$ be a committee satisfying JR with respect to a given approval profile $\mathbf{A}$. Then, W satisfies JR with respect to a modified approval profile $\mathbf{A}'$ if $\mathbf{A}'$ is constructed in one of the following two ways:*

(i) a voter $v \in V$ stops to approve a candidate not in W, i.e., $A'(v) = A(v) \setminus \{c\}$ for some $c \notin W$ and $A'(u) = A(u)$ for all $u \neq v$; or

(ii) a voter $v \in V$ changes her approval set such that the modified approval set contains a candidate in W, i.e., $A'(v) \cap W \neq \emptyset$ and $A'(u) = A(u)$ for all $u \neq v$.

Using this, we show tractability of PosJR and NecJR by proving that for each problem it is sufficient to check whether W satisfies JR for one carefully chosen completion.

Theorem 12. *In the poset approval model,* PosJR *and* NecJR *are both solvable in polynomial time.*

Proof sketch. We prove only NecJR (and omit PosJR for lack of space). Assume we are given a partial profile $\mathbf{P}$ and a committee W of size k. For each voter $v \in V$, let S_v be the inclusion-maximal subset of candidates in Middle(v) satisfying $S_v \cap W = \emptyset$ while respecting the partial order $\succ_v$ over Middle(v). We set $A(v) = \text{Top}(v) \cup S_v$ for each voter $v \in V$. It can be shown that if W satisfies JR in $\mathbf{A}$, then it also satisfies JR in any other completion $\mathbf{A}'$ of $\mathbf{P}$. $\square$

Note that if we treat k as a constant, we can use Theorem 12 to also check whether a given candidate is a possible member of a JR committee. This is because, for fixed k, we can enumerate all size-k subsets containing c.

We now turn our attention to two stronger representation axioms: *proportional justified representation (PJR)* and *extended justified representation (EJR)* require that cohesive groups of voters are represented proportionally to their size (Aziz et al. 2017; Sánchez-Fernández et al. 2017). Informally, if a group $G \subseteq V$ of voters satisfies both $|G| \geq \ell \cdot \frac{n}{k}$ and $|\bigcap_{v \in G} A(v)| \geq \ell$ for some integer ℓ, then PJR and EJR require that G is represented ℓ times by the committee. The axioms differ in the way that they define representation. (For formal definitions, refer to the full version of the paper.)

Even for a complete approval profile $\mathbf{A}$, it is coNP-complete to decide whether a given committee provides PJR or EJR w.r.t. $\mathbf{A}$ (Aziz et al. 2017, 2018). However, we show that this problem becomes tractable if we assume the committee size k to be fixed. Our proof is based on a proof by Aziz et al. (2018), who show an analogous result for a fixed number of candidates (instead of a fixed committee size).

Theorem 13. *Given a committee W and a (complete) approval profile $\mathbf{A}$, testing whether W satisfies PJR or EJR is possible in time $\mathcal{O}(m^{k+2} \cdot n \cdot k)$.*

If the committee size is assumed to be fixed, Theorem 13 implies that PJR and EJR can be verified efficiently in the case of complete information. This observation gives rise to computational questions regarding PJR and EJR under incompleteness: Given a committee W and a partial approval profile $\mathbf{P}$ and assuming k is fixed, can we decide in polynomial time whether W satisfies PJR or EJR for any or all of the completions of $\mathbf{P}$? We leave this open for future work.

7 Concluding Remarks

We studied computational aspects of approval-based committee voting under incomplete approval information. We adopted poset approval as a general model of incompleteness, along with the 3VA and linear special cases, and we focused on committees of a fixed size. We established a quite broad picture of complexity for the problems of determining whether a given set of candidates is a possible or necessary committee for large classes of ABC rules. We also proved several results on the problems of possible and necessary committee members, albeit leaving some cases open for future work. Finally, we investigated the question of whether a given committee satisfies, possibly or necessarily, the representation axiom *justified representation*. The problem remains open for the stronger axioms of *proportional* and *extended justified representation*; we established that if the committee size is fixed the potential source of hardness can only be the incompleteness. It seems promising to study the problem of possible and necessary committees also for other ABC rules such as sequential Phragmén or Rule X (Peters and Skowron 2020). Additional directions for future research include other models of incompleteness, and uncertainty in general, such as a model where voter attendance to the ballot is uncertain.

Acknowledgments

References

Alcantud, J. C. R.; and Laruelle, A. 2014. Dis&Approval Voting: A Characterization. *Social Choice and Welfare*, 43(1): 1–10.

Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; and Walsh, T. 2017. Justified Representation in Approval-Based Committee Voting. *Social Choice and Welfare*, 48(2): 461–485.

Aziz, H.; Elkind, E.; Huang, S.; Lackner, M.; Sánchez-Fernández, L.; and Skowron, P. 2018. On the Complexity of Extended and Proportional Justified Representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 902–909. AAAI Press.

Barrot, N.; Gourvès, L.; Lang, J.; Monnot, J.; and Ries, B. 2013. Possible Winners in Approval Voting. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, 57–70. Springer-Verlag.

Baumeister, D.; and Dennisen, S. 2015. Voter Dissatisfaction in Committee Elections. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1707–1708. IFAAMAS.

Brams, S. J.; and Kilgour, D. M. 2014. Satisfaction Approval Voting. In *Voting Power and Procedures*, Studies in Choice and Welfare, 323–346. Springer.

Cao, Z.; Qin, T.; Liu, T.; Tsai, M.; and Li, H. 2007. Learning to Rank: From Pairwise Approach to Listwise Approach. In *ICML*, volume 227 of *ACM International Conference Proceeding Series*, 129–136. ACM.

Chevaleyre, Y.; Lang, J.; Maudet, N.; Monnot, J.; and Xia, L. 2012. New Candidates Welcome! Possible Winners with Eespect to the Addition of new Candidates. *Mathematical Social Sciences*, 64(1): 74–88.

Felsenthal, D. S. 1989. On Combining Approval with Disapproval Voting. *Behavioral Science*, 34: 53–60.

Garey, M. R.; and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.

Imber, A.; Israel, J.; Brill, M.; and Kimelfeld, B. 2021. Committee Voting with Incomplete Approvals. Technical report, arxiv:2103.14847 [cs.GT].

Konczak, K.; and Lang, J. 2005. Voting procedures with incomplete preferences. In *Proceedings of the Multidisciplinary Workshop on Advances in Preference Handling (MPREF)*, 124–129.

Lackner, M.; and Skowron, P. 2021. Multi-Winner Voting with Approval Preferences. Technical report, arXiv:2007.01795 [cs.GT].

Lang, J. 2020. Collective Decision Making under Incomplete Knowledge: Possible and Necessary Solutions. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI) Survey Track*, 4885–4891. IJCAI.

Libkin, L. 2016. SQL's Three-Valued Logic and Certain Answers. *ACM Transactions on Database Systems*, 41(1): 1:1–1:28.

Lu, T.; and Boutilier, C. 2013. Multiwinner Social Choice with Incomplete Preferences. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, 263–270. AAAI Press.

Peters, D.; and Skowron, P. 2020. Proportionality and the Limits of Welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation (ACM-EC)*, 793–794. ACM Press.

Sánchez-Fernández, L.; Elkind, E.; Lackner, M.; Fernández, N.; Fisteus, J. A.; Basanta Val, P.; and Skowron, P. 2017. Proportional Justified Representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 670–676. AAAI Press.

Terzopoulou, Z.; Karpov, A.; and Obraztsova, S. 2021. Restricted Domains of Dichotomous Preferences with Possibly Incomplete Information. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 5726–5733. AAAI Press.

Thiele, T. N. 1895. Om Flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, 415–441.

Xia, L.; and Conitzer, V. 2011. Determining Possible and Necessary Winners under Common Voting Rules Given Partial Orders. *Journal of Artificial Intelligence Research*, 41: 25–67.

Zhou, A.; Yang, Y.; and Guo, J. 2019. Parameterized Complexity of Committee Elections with Dichotomous and Trichotomous Votes. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 503–510. IFAAMAS.