Multiwinner Voting with Interval Preferences under Incomplete Information

Drew Springham, Edith Elkind, Bart de Keijzer, and Maria Polukarov

Abstract

In multiwinner approval elections with many candidates, voters may struggle to determine their preferences over the entire slate of candidates. It is therefore of interest to explore which (if any) fairness guarantees can be provided under reduced communication. In this paper, we consider voters with one-dimensional preferences: voters and candidates are associated with points in $\mathbb{R}$, and each voter's approval set forms an interval of $\mathbb{R}$. We put forward a probabilistic preference model, where the voter set consists of σ different groups; each group is associated with a distribution over an interval of $\mathbb{R}$, so that each voter draws the endpoints of her approval interval from the distribution associated with her group. We present an algorithm for computing committees that provide *Proportional Justified Representation + (PJR+)*, which proceeds by querying voters' preferences, and show that, in expectation, it suffices to make $\mathcal{O}(\log(\sigma \cdot k))$ queries per voter, where k is the desired committee size.

1 Introduction

The problem of selecting a set of winning candidates from a wider pool, given the preferences of many agents, arises in a wide variety of settings. This problem is formalised as *Multiwinner Voting (MWV)*. Real-world examples include, but are not limited to, committee elections, parliamentary elections, and choosing group recommendations [17]. Often, we wish to select a committee of size k in a way that is "fair"; this is captured by the concept of *Justified Representation (JR)* and its extensions, such as PJR, PJR+, EJR, EJR+, and FJR [11, 4, 2, 24]. Each of these extensions captures the idea that groups making up at least an ℓ/k fraction of the voter population deserve to be represented by at least ℓ of the k candidates in the winning set. E.g., for the winning set W to provide PJR+ —the axiom we focus on in this work—for every group of voters S that constitutes an ℓ/k fraction of the population and approves of a candidate not in W, the set W must contain ℓ candidates approved by some voter in S.

Importantly, some multiwinner elections have a large number of candidates in consideration. E.g., in the well-known *participatory budgeting* setting, in which residents of a city can propose projects to be completed, and then vote on the proposals, there may be dozens of projects to choose from: indeed, there are real-life examples of participatory budgeting elections with more than 150 candidate projects [10]. In such circumstances, one cannot realistically expect the voters to accurately evaluate every single candidate. Therefore, it is of interest to consider multiwinner voting mechanisms that do not require full information from each voter about their preferences, but still provide guarantees on the social desirability of the outcome.

In this paper, we introduce the general problem of multiwinner voting with incomplete information in a spatial preferences setting, where voters and candidates are associated with points in $\mathbb{R}^n$, and each voter approves candidates within a certain distance from her. We then focus on the one-dimensional case of this general problem, where each voter's approval set forms an interval of $\mathbb{R}$ (her *approval interval*). Such a model may be appropriate, e.g., in a political setting, where we can view candidates as positioned on the left-right political spectrum; then, each voter only disapproves of candidates they find too extreme (to the left or to the right).

We put forward a rich probabilistic model to represent the voters' preferences in this setting: the *Random Interval Voter model (RIV)*. In RIV (Definition 5), we assume that voters are divided into groups, so that each group is associated with a disjoint subinterval of $\mathbb{R}$ and a voter's approval interval is obtained by independently sampling two points from a given distribution over the interval associated with their group.

When the voter population is large, a probabilistic model can be thought of as providing aggregated preference information, based, e.g., on historical data. In particular, an attractive feature of RIV is that it can express both large-scale trends and small-scale behaviour. Indeed, by associating each voter with a disjoint subinterval of $\mathbb{R}$, we can capture settings where candidates are partitioned into well-defined groups (parties), so that voters' preferences are consistent with this partition; in this setting, each party naturally corresponds to a subinterval of $\mathbb{R}$. Thus, the RIV model can be thought of as a probabilistic variant of open party-list elections. By modelling voters as having random endpoints within each segment, we can capture voters' within-party preferences by specifying a segment-specific cumulative distribution function. Together, both factors give RIV sufficient depth to be a nuanced model of how agents might vote in practice.

Our contributions. We develop a preference elicitation framework for the RIV model. Our framework allows two types of queries: *point queries*, where we ask if a voter approves a specific candidate, and *interval queries*, where we ask if a voter's set of approved candidates is contained within a given interval. We then consider two well-studied notions of proportionality, namely, core [2] and PJR+ [4], and show that our preference elicitation framework enables us to construct proportional committees in the RIV model. Specifically, we first establish that one can construct a committee in the core using $\mathcal{O}(\log m)$ queries per voter in expectation, where m is the number of candidates. Our main technical contribution is an algorithm that outputs committees providing PJR+ and only requires $\mathcal{O}(\log(\sigma k))$ queries per voter in expectation, where k is the target committee size and σ the number of segments; notably, unlike for the core, this bound does not depend on the total number of candidates.

Outline. In Section 2, we review prior work on multiwinner voting with incomplete information and preference elicitation. Next, in Section 3 we introduce the preliminary definitions and results, which lay the foundation for the contributions of this paper. In Section 4, we describe our probabilistic model, and formulate the general problem of multiwinner voting with incomplete information in a spatial setting. Section 5 discusses our preferences elicitation framework; we also formulate accompanying lemmas, which we put to use in Section 6, where we prove that we can construct a committee that provides PJR+ using limited queries per voter in expectation. Finally, in Section 7 we conclude with ideas for future work and open questions.

2 Related Work

In this paper, we select committees that satisfy PJR+ in the case where the decision mechanism does not have full information about voters' preferences. There is substantial literature on approval-based multiwinner voting in general (see [17] for an overview). The preference elicitation aspect of voting has also been studied extensively, especially in the single winner context (e.g. [5, 21, 7, 6, 28]). Earlier work in multiwinner voting with incomplete information optimises minimax regret [19], whereas we are interested in guaranteeing forms of justified representation in this work.

Imber et al. [15] also investigate multiwinner voting with incomplete information. In their setting, the decision mechanism knows for each voter only a subset of their approved candidates and a subset of their disapproved candidates, with the remaining candidates being unknown. They consider the complexity of determining which candidates are possibly or necessarily in the winning set under different election rules that guarantee forms of justified representation. Later, Imber et al. [16] consider the same problems when voters approve candidates with respect to d-dimensional Euclidean space. In this paper, we consider a 1-dimensional space, however, [16] differs from our setting in two ways: we investigate the complexity of gaining the information from the voters in the first place, and we also take a probabilistic perspective rather than a nondeterministic perspective. Mandal et al. [20] study communication complexity of multiwinner voting, focussing on the trade-off between social welfare and information elicited from each voter, rather than aiming to achieve forms of justified representation. They also prove lower bounds on the communication complexity required.

Halpern et al. [13] motivate multiwinner voting with incomplete information in the context of users voting (approving or disapproving) on comments on an online surveying service. They do not ask their random sample of voters about all candidates (comments) and instead ask only about at most t candidates. They formalise an idea of approximate justified representation by allowing the required size of cohesive groups to have an additional $(1+\varepsilon)$ factor, and provide adaptive algorithms to achieve approximate EJR (a weaker notion than EJR+). They also provide a lower bound on the number of voters that need to be sampled to achieve EJR with probability of at least $1-\delta$. Brill and Peters [4] discuss a similar technique for achieving EJR+ with high probability in the general preference setting.

3 Preliminaries

Here, we present the relevant background to introduce our work. All our indexing of ordered sets (or lists) starts at 0, and we use "list slicing" notation of the form $X[a:b] := \{X[i] : a \leq i < b\}$.[1] We also use the notation $[t] := \{i \in \mathbb{N} : 1 \leq i \leq t\}$ to denote the set of natural numbers up to t.

Definition 1. *(Approval-based MWV election) An* MWV election *is a tuple $E = (V, C, k, A)$, where V is a set of n voters, C is a set of m candidates, $k \in \mathbb{N}$ is a total number of candidates to elect, and $A : V \to 2^C$ is a function that maps each voter to the set of candidates she approves. For each $G \subseteq V$, we write $A(G) = \bigcap_{v \in G} A(v)$. The primary task associated with an MWV election is to select a set,or committee, of* winners *$W \subseteq C$ with $|W| = k$.*

The literature on multiwinner voting with approval preferences defines a number of fairness axioms, ranging from Justified Representation (JR) to core stability [17]. All these axioms aim to capture the idea that large groups of voters with similar preferences should be represented in the committee in proportion to the group size; however, they differ in how they define which groups of voters are considered to have similar preferences and what it means to represent a group. Below, we define two axioms from this hierarchy, namely, PJR+ [4] and core stability [2]. In what follows, we will show that our method selects PJR+ committees with high probability while asking a small number of queries; obtaining a similar result for more demanding axioms, such as, e.g., EJR+ or core stability, remains a direction for future work.

Definition 2. *(PJR+) A committee W satisfies PJR+ [4] (a.k.a. IPSC [1]) if $|W| = k$ and for every $\ell \in [k]$ and every group $G \subseteq V$ that satisfies $|G| \geq n\ell/k$, $\bigcap_{v \in G} A(v) \setminus W \neq \emptyset$ it holds that $|W \cap \bigcup_{v \in G} A(v)| \geq \ell$.*

In words, a committee W satisfies PJR+ if it there is no group of voters who (i) deserve ℓ representatives, (ii) can all agree on a candidate not in W, but (iii) collectively support fewer than ℓ members of W. Another well-known related (but logically incomparable) concept is *the core.*

Definition 3. *(The core) A committee W satisfies* core stability *[2, 17] if $|W| = k$ and for every $\ell \in [k]$, every $T \subseteq C$ with $|T| = \ell$, and every group $G \subseteq V$ with $|G| \geq n\ell/k$ there is a voter $v \in G$ with $|W \cap A(v)| \geq |T \cap A(v)|$.*

Every MWV election has a committee that satisfies PJR+; in particular, every committee computed by the Method of Equal Shares (MES) [24] satisfies PJR+[2]. Later, we will use MES to find a PJR+ committee

[1]Similar to the Python programming language.
[2]MES satisfies a stronger notion of proportionality called EJR+, which we shall discuss in Section 7

(Algorithm 2). On the other hand, it is an open question whether every approval-based MWV election has a core stable committee [2, 17].

We will consider one-dimensional approval preferences.

Definition 4. *(Candidate-Interval (CI) preferences) Given a set of candidates C and a linear order $\lhd$ over C, we say that a subset of candidates $T \subseteq C$ is* consecutive *if for every $x, z \in T$ and $y \in C$ with $x \lhd y \lhd z$ it holds that $y \in T$. An election $E = (V, C, k, A)$ has* Candidate Interval (CI) preferences *[8] if there exists a linear order $\lhd$ over C such that for each voter v the set $A(v)$ is consecutive.*

If $C \subseteq \mathbb{R}$, each voter approves of an interval of candidates: $T \subseteq C$ is consecutive if there exists an interval $I = [a, b]$ such that $T = C \cap I$. One can interpret CI preferences as single-peaked preferences in the dichotomous setting; this class of preferences has been explored in a number of papers, e.g, [23, 12, 27, 25].

4 Random Interval Voter Model (RIV)

In this section, we introduce the probabilistic model for CI preferences that will be used throughout this paper; we will also briefly discuss an extension of this model to more than one dimension.

Definition 5. *An* RIV model *is a tuple $\mathcal{M} = (I_t, F_t, p_t)_{t\in[\sigma]}$, where $\sigma \in \mathbb{N}$, and for each $t \in [\sigma]$ it holds that $I_t = (z_t^-, z_t^+)$ is a subset of $\mathbb{R}$, with $I_x \cap I_y = \varnothing$ for $x \neq y$, F_t is an invertible cumulative distribution function over I_t, $p_t \in [0, 1]$, and $\sum_{t\in[\sigma]} p_t = 1$. Voters' approvals are sampled as follows: a segment I_t is chosen with probability p_t, two positions $X, Y \in I_t$ are drawn from distribution F_t, we set $a_v = \min(X, Y)$, $b_v = \max(X, Y)$, and the voter's approval ballot is defined as $A^*(v) = [a_v, b_v]$, $A(v) = A^*(v) \cap C$.*

An RIV model $\mathcal{M}$ is uniform *if $I_t = [t, t + 1]$ and F_t is the uniform distribution over $[t, t + 1]$ for each $t \in [\sigma]$. We assume that $C \subset \bigcup_{t\in[\sigma]} I_t$, let $C_t = C \cap I_t$, and for each $c \in C$ we define $t(c)$ to be the segment I_t with $c \in I_t$.*

Note that focusing on uniform RIVs is without loss of generality: If we are given a general RIV $\mathcal{M}$, we can transform it into a uniform RIV $\mathcal{M}'$, by mapping each point (e.g., locations of candidates, approval endpoints, query positions) $x \in I_t$ in $\mathcal{M}$ to $t + F_t(x)$ in $\mathcal{M}'$. Then a candidate is approved by a voter in $\mathcal{M}$ if and only if the transformed candidate is approved by the transformed voter in $\mathcal{M}'$, and it can be verified (see Appendix A) that this occurs with the same probability as in $\mathcal{M}$. Thus, from now on we will assume that all RIV instances are uniform.

We now define some notation that will be used throughout, and state a useful lemma.

Definition 6. *For each $T \subseteq C$ and $x \in \bigcup_{t=1}^{\sigma} I_t$, let*

$$x_T^{\rightleftarrows} = \min(\{c \in T : c \geq x\} \cup \{z_{t(x)}^+\}),$$
$$x_T^{\leftrightarrows} = \max(\{c \in T : c \leq x\} \cup \{z_{t(x)}^-\}).$$

In words, $x_T^{\leftrightarrows}$ (resp., $x_T^{\rightleftarrows}$) is the closest candidate in T to the left (resp., to the right) of x, or the endpoint of the segment $I_{t(x)}$ if no such candidate exists.

The following lemma gives a useful expression for the probability of approving and disapproving candidates.

Lemma 1. *For a (uniform) RIV instance, a subset of candidates $S \subseteq C$, and a candidate $c \in C \setminus S$ we have*

$$\Pr(c \in A(v) \wedge A(v) \cap S = \varnothing) = 2p_{t(c)} \, (c - c_S^{\leftrightarrows}) \, (c_S^{\rightleftarrows} - c) \, .$$

Proof. The event that both $c \in A(v)$ and $A(v) \cap S = \varnothing$, i.e., v approves c but disapproves each $c' \in S$, is equivalent to the event $c_S^{\leftarrow} < a_v \leq c \leq b_v < c_S^{\rightarrow}$, since otherwise either c would not be approved by v, or some candidate in S would be approved by v. Finally,

$$\Pr\left(c_S^{\leftarrow} < a_v \leq c \leq b_v < c_S^{\rightarrow} | a_v, b_v \in I_{t(c)}\right) = 2(F_i(c) - F_i(c_S^{\leftarrow}))(F_i(c_S^{\rightarrow}) - F_i(c)).$$

As we assume that the RIV instance is uniform, and $\Pr(a_v, b_v \in I_{t(c)}) = p_{t(c)}$, the result follows. □

4.1 Queries

We allow two types of queries: point queries and interval queries. A *point query* $\textsc{Point}(x, v)$ asks voter v whether they approve candidate position x; the answer is $\mathbb{I}[x \in A(v)]$. An *interval query* $\textsc{Int}(x, y, v)$ communicates two candidate or segment endpoint positions x, y to v, who responds with $\mathbb{1}[A^*(v) \subseteq [x, y]]$. We refer to a sequence of queries of both types, together with voters' responses, as a *dialogue*; e.g., a dialogue can take the form $\mathcal{I} = (c_1 \in A(v), c_2 \notin A(v'), A^*(v'') \nsubseteq [c_3, c_4])$.

Point queries allow us to obtain information about the approval of a single candidate, while interval queries allow us to effectively ask a voter "Would you only approve of some compromise between x and y?" Note that we only submit point queries at candidate positions: we may only ask the voters about actual candidates, and not about arbitrary positions on the real line. With interval queries, we additionally allow queries parametrised by the endpoints of segments.

Our aim is to minimise the number of queries that we pose to each voter, while still guaranteeing certain fairness properties. If we were to consider more expressive queries, we could ask a voter v to indicate which of the possible $\mathcal{O}\left(m^2\right)$ ballots she holds, and recover full information regarding each voter's preferences. We show in Section 5 that our query model still allows us to recover full information using $\mathcal{O}(\log m)$ bits per voter in expectation for RIV elections, matching the $\mathcal{O}(\log m)$ required in the more expressive model. However, we would like to determine what fairness guarantees can be provided with fewer bits per voter in expectation; in particular, the number of queries of our PJR+ algorithm (Section 6) is independent of m (in expectation).

We now introduce the d-dimensional Random Euclidean Voter Model, which is a generalisation of RIV. While we will not tackle this model in this work, we briefly discuss it to motivate our study of the RIV model.

Definition 7. *Every bounded interval I of $\mathbb{R}$ together with a finite set $C \subset I^d$, a probability distribution $\mathcal{D}$ over convex subsets of I^d, and $n \in \mathbb{N}$ define a d-*dimensional Random Euclidean Voter Model (d-REV) *as follows: we draw n samples $A_1^*, \ldots, A_n^* \sim \mathcal{D}$, let $V = [n]$, and define the approval set of voter $v \in V$ as $A(v) = A_v^* \cap C$. We define two types of queries that can be used to elicit voters' preferences:*

- *Point queries: Given a candidate $x \in C$ and a voter $v \in V$, the query $\textsc{Point}(x, v)$ returns $\mathbb{1}[x \in A^*(v)]$.*
- *Hull queries: Given a finite set of points $P \subseteq I^d$ and a voter $v \in V$, the query $\textsc{Hull}(P, v)$ computes H as the convex hull of points in P and returns $\mathbb{1}[A^*(v) \subseteq H]$.*

We note that point queries and hull queries in the d-REV model correspond exactly to point and interval queries in the RIV model. Spatial models are a common way to represent voters' preferences over alternatives [22, 12, 16]: each dimension corresponds to an issue, and a point in this space (a candidate) represents a specific stance on all of the issues simultaneously. In this model, each voter views some subset of this space as acceptable. We require acceptable sets to be convex: if a voter approves of positions x and y, she also approves of compromises between x and y. In general, voters' approval sets may have complex geometries, so that communicating (a description of) a voter's approval set may require a significant amount of information. However, the two types of queries above enable efficient

Algorithm 1: Resolving a voter for a candidate set P

Function query(v, C, P)**:**

- $T \leftarrow \{I_t : t \in [\sigma], P \cap I_t \neq \varnothing\}$, sorted;
- $I_t \leftarrow$ binary search on T using interval queries;
- **if** I_t *not found* **: return** $\varnothing$;
- **for** $i \leftarrow 1$ *to* $\lceil \log(|P|) \rceil$ **:**
  - **for** $j \leftarrow 1$ *to* 2^i **:**
    - Point$\left(\left(t + \frac{j}{2^i}\right)_C^{\leftarrowtail}, v\right)$;
    - Point$\left(\left(t + \frac{j}{2^i}\right)_C^{\rightarrowtail}, v\right)$;
    - **if** v *approves either point query* **:**
      - Perform binary search point queries on $P \cap (t + \frac{j}{2^i}, t + \frac{j+1}{2^i})$ and $P \cap (t + \frac{j-1}{2^i}, t + \frac{j}{2^i})$ to find $\alpha, \beta = \min, \max(P \cap A(v))$;
      - **return** $P \cap [\alpha, \beta]$;
    - Int$\left(\left(\left(t + \frac{j}{2^i}\right)_C^{\leftarrowtail}, \left(t + \frac{j}{2^i}\right)_C^{\rightarrowtail}\right), v\right)$;
    - **if** v *approves this query* **: return** $\varnothing$;
- `// Only occurs if not already returned`
- **for** $c \in P \cap I_t$ **:** Point(c, v) ;
- **return** all approved points in P;

communication; each query consists of a finite number of points in I^d, and the voter's response is always a single bit.

The spatial model can capture simple participatory budgeting scenarios; consider a 2-REV model representing the geographical layout of a city, where each project is associated with a particular location, and voters approve of all projects that lie within some area around their residence. We are able to query voters in two ways: (1) we can ask a voter whether they approve a specific project, and (2) we can ask a voter whether all of their approved projects lie within the convex hull of some set of points on the map. While the d-dimensional general model is more expressive than the RIV model, we hope that investigating the 1-dimensional model first may provide useful insights into how to begin tackling the more general model.

5 Obtaining Information in RIV Elections

We now return to the challenge of efficient information elicitation in the 1-dimensional RIV model. Suppose that we want to know with certainty which candidates in a set $P \subseteq C$ are approved by a given voter.

Definition 8. *We say that a voter v is* resolved for $P \subseteq C$ *given a dialogue $\mathcal{I}$ if* $\Pr(p \in A(v) \mid \mathcal{I}) \in \{0, 1\}$ *for all $p \in P$. We drop $\mathcal{I}$ from the notation when it is clear from the context.*

We now describe our algorithm for resolving a given voter v (Algorithm 1). First, it uses binary search to find the segment I_t containing v, implemented via interval queries. Specifically, we start with $T = \{I_t : P \cap I_t \neq \varnothing\}$, and set $x = \min \bigcup_{I \in T} I$, $y = \max \bigcup_{I \in T} I$. We perform an interval query Int(x, y, v); if the response is negative, we return $\varnothing$, since v cannot approve any candidates in P in this case, and otherwise we recurse. That is, we partition T into two (nearly-)equal cardinality sets L, U, where L is the $\lceil |T|/2 \rceil$ leftmost segments of T, and $Y = T \setminus L$, and perform an interval query with $x = \min \bigcup_{I \in L} I$, $y = \max \bigcup_{I \in L} I$. We set $T := L$ if the query response is positive and $T := U$ if it is negative, and recurse. We stop when $|T| = 1$; at that point, v lies in the unique interval $I_t \in T$.

Once the segment I_t containing v is determined, the algorithm performs a different type of binary search within I_t. It proceed in rounds; in each round i, it submits a point query on the candidates closest to position $t + j/2^i$ for $j \in [2^i]$. When some such query is answered positively, it yields a position in the voter's approval interval; the algorithm then performs two binary searches either side of this position to find the candidates $\min, \max(P \cap A(v))$. After $\log |P|$ rounds, if a voter does not approve of any query it receives, the algorithm "gives up" and submits point queries for all candidates in P (the final two lines of Algorithm 1). However, this only happens with low probability, so with high probability—and also in expectation—the algorithm uses $\mathcal{O}(\log |P|)$ queries.

Theorem 2. *for every set $P \subseteq C$, given a uniform RIV, Algorithm 1 resolves a voter $v \in V$ using $\mathcal{O}(\log(|P|)$ queries in expectation.*

Proof (sketch). The full proof can be found in Appendix B. First, we can find t, the segment that v lies in, using $\mathcal{O}(\log |P|)$ interval queries. Assume that we know t then; we perform $\mathcal{O}(\log |P|)$ rounds of queries with exponentially increasing "resolution". The probability that a voter does not approve of any point query in the first i rounds is $1/2^i$, and the number of queries we use in the first i rounds is $\mathcal{O}(2^i) + \mathcal{O}(\log |P|)$, where the $\mathcal{O}(\log |P|)$ term is required to perform the two binary searches. Suppose that, after $\mathcal{O}(\log |P|)$ rounds, the voter has not approved of any queries; call this event Z'. Then the algorithm will use $\mathcal{O}(|P|)$ point queries, but this only happens with probability $\Pr(Z') \leq 1/2^{\log |P|} = 1/|P|$. Therefore the expected number of queries is $\mathcal{O}(\log |P|)$. □

We have a fairness result as a consequence of this theorem.

Corollary 3. *In a RIV election, we can find a committee in the core using $\mathcal{O}(\log m)$ queries per voter in expectation.*

Proof. Running Algorithm 1 with $P = C$ for all $v \in V$, we get full information using $\mathcal{O}(\log m)$ queries per voter in expectation. Pierczyński and Skowron [25] give a polynomial time algorithm for constructing a core-stable committee given full information about a profile of CI preferences. □

6 Finding PJR+ Committees for RIV

We now present an algorithm for achieving PJR+ for RIV elections (Algorithms 2 to 4). The expected per-voter query complexity of our algorithm does not depend on m. We prove the following theorem.

Theorem 4. *For a uniform RIV, Algorithm 2 finds a PJR+ committee with $\mathcal{O}(\log(\sigma k))$ queries per voter in expectation and with high probability as long as the number of voters n satisfies $n = \Omega(k^2 \log m)$.*

Our algorithm first "guesses" a committee $\widehat{W}$; we will argue that $\widehat{W}$ provides PJR+ with high probability. For each $t \in [\sigma]$, we allocate $k_t = \lfloor p_t k \rfloor$ committee seats to I_t. Specifically, we mark points $t + \frac{2i-1}{2(k_t+2)}$ for $i \in [k_t + 1] \setminus \{1\}$, and form the set $\widehat{W}_t$ by selecting k_t candidates from $C_t = C \cap I_t$ that are closest to the marked points. We then set $\widehat{W} = \bigcup_{t=1}^{\sigma} \widehat{W}_t$. Intuitively, by selecting candidates closest to the uniformly spread marked points, we ensure that these candidates are approximately uniformly distributed on each segment; consequently, there are no large gaps between committee members on this segment.

Example 5. We provide an example of how we construct $\widehat{W}$. Let $\sigma = 1$ (an extension to multiple segments is straightforward), and consider candidate positions in Figure 1 with $k = k_1 = 6$. To select the first candidate, we order the candidates by distance to a marked point (excluding marked points 1 and 8) and select the candidate with the smallest distance; the candidate labelled 3 in our example, as it sits at a marked point. We then remove marked point 3 from consideration (in Algorithm 2, the set of marked points under consideration is denoted by J) and reorder the candidates. We then pick both 2

Algorithm 2: Finding a PJR+ committee

input : A uniform RIV with (V, C, k) and $(p_t)_{t\in[\sigma]}$

for $t \in [\sigma]$ **:**
- $k_t \leftarrow \lfloor p_t k \rfloor; \widehat{W}_t \leftarrow \varnothing; J \leftarrow [k_t + 1] \setminus \{1\};$
- **while** $J \neq \varnothing$ **do**
  - $(i, c) \leftarrow \arg\min_{i\in J, c\in C_t\setminus\widehat{W}_t} \left|c - t - \frac{2i-1}{2(k_t+2)}\right|;$
  - $\widehat{W}_t \leftarrow \widehat{W}_t \cup \{c\}$; $J \leftarrow J \setminus \{i\}$;

$\widehat{W} = \bigcup_{t=1}^{\sigma} \widehat{W}_t$, sorted;

for $v \in V$ **:**
- $t \leftarrow$ the index such that $v \in I_t$, found using binary search as in `query`;
- $P \leftarrow \widehat{W} \cup \left\{ \left(t + \frac{j}{15k}\right)_{C_t}^{+/-} : j \in [15k] \right\};$
- `query` (v, C, P);

if `validate` *($\widehat{W}, E$) (Algorithm 3)* **: output:** $\widehat{W}$

for $v \in V$ **:** `query` (v, C, C) ;

output: $W^* =$ `MES`(E) using full preference information

Algorithm 3: Validating if a committee $\widehat{W}$ provides PJR+

Function `validate`$(\widehat{W}, E)$**:**
- $u \leftarrow |\{v \in V : v$ not approved any query$\}|$;
- **for** $\ell = 1$ *to* k, $i = 0$ *to* $m - 1$ **:**
  - $c \leftarrow C[i]; \rho \leftarrow |\{x \in \widehat{W} : x < c\}|$;
  - **if** $c \in \widehat{W}$ **: continue** to next i;
  - **for** $j = 0$ *to* $\ell - 1$ **:**
    - $s_{\ell,i,j} \leftarrow 0; R \leftarrow \widehat{W}[\rho - j : \rho + \ell - 1 - j]$;
    - **for** $v \in V$ *that approved some query* **:**
      - **if** `poss` $(v, c, \widehat{W} \setminus R)$ *(Algorithm 4)* **:**
        - $s_{\ell,i,j} \leftarrow s_{\ell,i,j} + 1$;
        - **if** $s_{\ell,i,j} + u \geq \frac{n\ell}{k}$ **: return** *False* ;
- **return** *True*;

Algorithm 4: Determining if v has positive probability of approving t and disapproving S

Function `poss`(v, t, S)**:**
- **if** $t \in S$ **: return** *False*;
- $D, A \leftarrow$ set of point query positions v has (dis)approved resp. (Assert $A \neq \varnothing$);
- $\phi_2 \leftarrow \min A; \phi_3 \leftarrow \max A$;
- $\phi_1 \leftarrow (\phi_2)_D^{\leftarrow}; \phi_4 \leftarrow (\phi_3)_D^{\rightarrow}$;
- **return** $\phi_1 < t < \phi_4 \wedge t_S^{\leftarrow} < \phi_2 \wedge t_S^{\rightarrow} > \phi_3$;

and 6. We select the remaining candidates in this way, constructing a set $\widehat{W}$ of size k as numbered 2 through 7 in Figure 1. We see that the algorithm spreads k candidates across the segment as uniformly as possible.

Now, given our guessed committee, we verify whether it provides PJR+. For each voter v, we find which segment I_t contains v, using binary search similarly to Algorithm 1. We then pick points spaced $1/15k$ apart along the segment, along with $\widehat{W}$, and query the voter using Algorithm 1 to resolve each voter on the candidates closest to these points; this gives the algorithm an approximation for the voter's

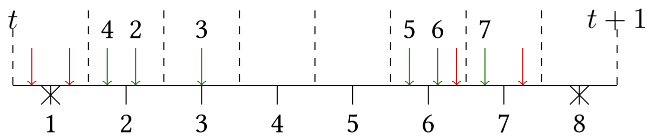

Figure 1: The algorithm selects candidates c for the committee $\widehat{W}_t$ in order of distance from marked points. We annotate the section between marked points containing each candidate. The candidates are selected in order $3, 2, 6, 7, 5, 4$.

approval interval. If the voter did not approve any query, we add them to the counter u, as we do not have much information about the location of the voter's approval interval, only that it must lie entirely between two adjacent point queries.

If the voter did approve some query, we check for every $\ell \in [k]$, every candidate $c \in C \setminus \widehat{W}$ and for every $0 \leq j \leq \ell - 1$ whether it could be consistent, given the information we have elicited from the voter, for the voter to approve of: (1) c, (2) at most j candidates in $\widehat{W}$ to the left of c, and (3) fewer than ℓ candidates in $\widehat{W}$. If this event has positive probability, captured by the $\texttt{poss}$ function, we add the voter to the counter $s_{\ell,i,j}$. Finally, if $s_{\ell,i,j} + u < n\ell/k$ for all ℓ, i, j we output $\widehat{W}$. Otherwise, we query all $v \in V$ on all $c \in C$ to get full information, and then run an algorithm (e.g., MES) that provides PJR+ to output a committee W^*.

We show in Appendix E that if $s_{\ell,i,j} + u < n\ell/k$ for all ℓ, i, j, then $\widehat{W}$ provides PJR+ with certainty given the query information. If this is not the case, we use $\mathcal{O}(m)$ queries per voter in order to obtain full information about $A(v)$ for each $v \in V$; however, we show that $\widehat{W}$ provides PJR+ with high probability, so we only require $\mathcal{O}(\log \sigma k)$ queries in expectation and with high probability. The time complexity of Algorithm 2 (ignoring MES, which runs in polynomial time [24]) is $\mathcal{O}(nmk^2)$.

To prove Theorem 4, we introduce intermediate lemmas. We use the definitions in Algorithm 2 for $\widehat{W}, \widehat{W}_t, s_{\ell,i,j}$ as they exist at the termination of the algorithm. We also define functions necessary for the proof. Before providing formal definitions, we give high-level intuition.

Assume all $\widehat{W}_t$ are ordered. The function $\rho(c)$ is used to help us index $\widehat{W}_{t(c)}$: for example, $\widehat{W}_{t(c)}[\rho(c) - r]$ is the r-th candidate in $\widehat{W}_t$ to the left of c. The set K_t contains the candidates in the subinterval of I_t bounded by the first and last marked points that we select candidates for. We prove probability bounds for candidates that appear within K_t, as we know that there exists at least one marked point to the left and right of c within this region. We deal with candidates $c \in C_t \setminus K_t$ separately. $d_r^{+/-}(c)$ is the distance between c and the $(r+1)$-th candidate in $\widehat{W}_{t(c)}$ to the right/left, respectively (if there exist $r+1$ candidates in $\widehat{W}$ to the right/left of c, and the corresponding endpoints of the segment otherwise); see Figure 2 for an illustration. Finally, $Y_{c,\ell,r}$ is the event that a voter approves candidate c, at most r candidates in $\widehat{W}$ to the left of c, and strictly fewer than ℓ candidates in $\widehat{W}$ in total. Now we provide formal definitions.

Definition 9. *Define function* $\rho(c) = |\{x \in \widehat{W}_{t(c)} : x < c\}|$. *Define* $K_t = C_t \cap \left(t + \frac{3}{2(k_t+2)}, t + 1 - \frac{3}{2(k_t+2)}\right)$. *For* $0 \leq r < \ell \leq k$ *and* $c \in C \setminus \widehat{W}$, *define* $Y_{c,\ell,r}$ *as the random event that:*

$$c \in A(v) \text{ and } A(v) \cap \widehat{W} \subseteq \widehat{W}_{t(c)}[\rho(c) - r : \rho(c) + \ell - 1 - r].$$

Define the functions $d_r^+(c) = \widehat{W}_{t(c)}[\rho(c) + r] - c$, $d_r^-(c) = c - \widehat{W}_{t(c)}[\rho(c) - r - 1]$, *and* $\pi(c, \ell, r) = \Pr(Y_{c,\ell,r})$.

We can write $\pi(c, \ell, r) = \Pr(Y_{c,\ell,r})$ as a product of $d_r^{+/-}(c)$, and bound $d_{r_1}^+(c) + d_{r_2}^-(c)$ to obtain a bound on π in terms of ℓ and k.

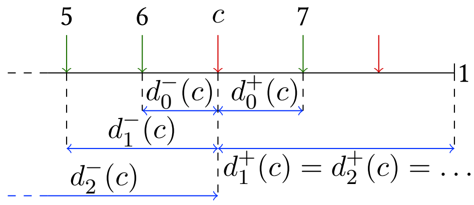

Figure 2: Example of $d_r^{+/-}(c)$. It is the distance from candidate $c \notin \widehat{W}$ to the $(r+1)$-th candidate in $\widehat{W}$ to the left/right of c, or the distance to the segment endpoints if there are not enough candidates in $\widehat{W}$.

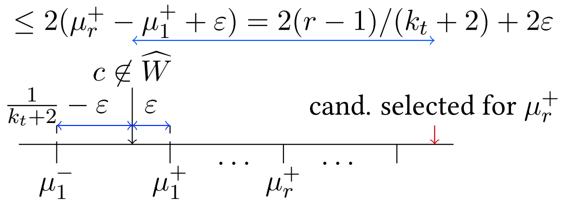

Figure 3: Intuition for Lemma 7. The candidate selected for a marked point μ_r^+ cannot be further from μ_r^+ than a candidate $c \notin W$, as otherwise c would have been chosen instead.

Lemma 6. $\pi(c, r_1 + r_2 + 1, r_1) = 2p_{t(c)} d_{r_2}^+(c) d_{r_1}^-(c)$.

Proof. $\pi(c, r_1 + r_2 + 1, r_1)$ is the probability that v approves of c and at most r_1 candidates in $\widehat{W}$ to the left of c and at most r_2 candidates to the right of c. We only need to consider $\widehat{W}_{t(c)}$, not all of $\widehat{W}$, since it is impossible for a voter to approve c and some candidate in $\widehat{W} \setminus \widehat{W}_{t(c)}$. We have $A(v) \cap \widehat{W} \subseteq \widehat{W}_{t(c)}[\rho(c) - r_1 : \rho(c) + r_2]$ if and only if $A(v) \cap (\widehat{W}_{t(c)} \setminus \widehat{W}_{t(c)}[\rho(c) - r_1 : \rho(c) + r_2]) = \varnothing$. Let $S = \widehat{W}_{t(c)} \setminus \widehat{W}_{t(c)}[\rho(c) - r_1 : \rho(c) + r_2]$. Then, with Lemma 1, we have

$$\Pr(c \in A(c) \wedge A(v) \cap S) = \varnothing) = 2p_{t(c)}(c - c_S^{\leftarrow})(c_S^{\rightarrow} - c) = 2p_{t(c)} d_{r_1}^-(c) d_{r_2}^+(c);$$

note that if $\rho(c) - r_2 < 1$, then we say $c_S^{\leftarrow} = t(c)$, and similar for $\rho(c) + r_2 \geq k_{t(c)}$, we have $c_S^{\rightarrow} = t(c) + 1$. $\square$

Given this result, we can transform a bound on $d_r^{+/-}$ into a bound on π. We provide a bound on $d_r^{+/-}(c)$.

Lemma 7. $d_{r_1}^+(c) + d_{r_2}^-(c) \leq \frac{2(r_1+r_2+1)}{k_{t(c)}+2}$ *for* $c \in K_t \setminus \widehat{W}_t$.

The full proof of this result can be found in Appendix C. The general intuition for this result can be seen in Figure 3 and is as follows: recall that in each segment I_t, we select k_t candidates, according to their distance to the marked points $t + \frac{2i-1}{2(k_t+1)}$. Given that $c \notin \widehat{W}_t$, we know that the candidate selected for the i-th midpoint p to the right of c must be closer to p than c is, as otherwise c would have been chosen instead. This allows us to upper bound $d_{r_1}^+(c) \leq \frac{2r_1}{k_t+2} + 2\varepsilon$. We can derive a similar bound on $d_{r_2}^-(c) \leq \frac{2r_2}{k_t+2} + 2\left(\frac{1}{k_t+2} - \varepsilon\right)$, giving us the desired result.

With Lemma 6, we can bound $\pi(c, \ell, r)$, the probability that a voter approves c, at most r candidates in $\widehat{W}_t$ to the left of c, and fewer than ℓ in total.

Corollary 8. *For* $c \in K_t \setminus \widehat{W}$, $\pi(c, \ell, r) \leq \frac{2p_t\ell^2}{(k_t+2)^2}$.

We can now upper bound π in terms of ℓ and k, and provide a bound for all $c \in C_t \setminus \widehat{W}$.

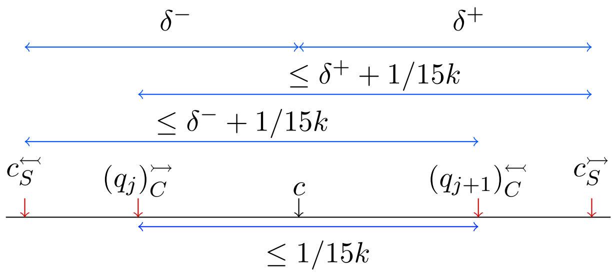

Figure 4: Intuition for Lemma 10. Query points are shown in red. Recall that we query at $c_S^{\leftarrow}$ and $c_S^{\rightarrow}$.

Lemma 9. *For $c \in C_t \setminus \widehat{W}$, $\pi(c, \ell, r) \leq \ell/k - 1/4k$.*

Proof. The case $c \in (C_t \setminus K_t) \setminus \widehat{W}_t$ is handled by Lemma 11 in Appendix D. Now, suppose $c \in K_t \setminus \widehat{W}_t$. We have $\Pr(c \in A(v)) \leq p_t/2$, and so when $p_t k + 1/2 \leq 2\ell$ we have the result. We also get the result if $2\ell \leq k_t + 1$; we have $\frac{2p_t\ell^2}{(k_t+2)^2} \leq \frac{2p_t\ell^2}{kp_t(2\ell+1)} = \frac{2\ell^2}{k(2\ell+1)} \leq \ell/k - 1/4k$ for $\ell \geq 1$. So then let us consider if we can have $k_t + 1 < 2\ell < p_t k + 1/2$. Consider that k_t and ℓ are integers, so $k_t + 2 \leq 2\ell < p_t k + 1/2$, so $p_t k > \lfloor p_t k \rfloor + 3/2$, a contradiction, so we are done. □

Now that we have upper bounds on the probability that a given voter may be a member of a group that has the potential to violate PJR+, we can bound the probability that $s_{\ell,i,j} \geq \ell n/k - n/12k$.

Lemma 10. *For a given ℓ, i, j and with $\alpha = 1800^{-1}$ (a constant), we have $\Pr(s_{\ell,i,j} \geq \frac{ln}{k} - \frac{n}{12k}) \leq \exp\left(-\frac{\alpha n}{k^2}\right)$.*

Proof of Lemma 10. Let us bound $s_{\ell,i,j}$. For ℓ, i, j, let $c = C[i]$ and say $c \in I_t$. Assume $c \notin \widehat{W}$. Let R be as defined in the algorithm. Recall the definition of $Y_{c,\ell,r}$ made in Definition 9. Let X be the event that $\texttt{poss}(v, c, \widehat{W} \setminus R))$ evaluates to true. This is the probabilistic event that the voters approval interval $[a_v, b_v]$ exists such that, given the information $\mathcal{I}$ obtained from querying v, $\Pr(Y_{c,\ell,r}|\mathcal{I}) > 0$. We want to find $\Pr(X)$. Let $S = \widehat{W} \setminus R$. We have $\delta^- := d_j^-(c) = c - c_S^{\leftarrow}$, and $\delta^+ := d_{\ell-1-j}^+(c) = c_S^{\rightarrow} - c$. We know $\delta^- + \delta^+ \leq 1$ and by Lemmas 1 and 9, we know that $2p_t\delta^+\delta^- = \pi(c, \ell, j) \leq \frac{\ell}{k} - \frac{1}{4k}$. First, if $c \in P$ (as defined in Algorithm 2), then $\Pr(X) = \pi(c, \ell, t)$, since we will perform queries at c and at all of S, and so $Y_{c,\ell,r} = X$.

Consider then if $c \notin P$. See Figure 4 for a diagram. Let $q_j = t + j/15k$, the j'th position in segment t around which we query. Then consider j where $q_j \leq c < q_{j+1}$. Since $c \notin P$, we know that $q_j \leq (q_j)_C^{\rightarrow} < c < (q_{j+1})_C^{\leftarrow} \leq q_{j+1}$, and so there exist queried positions either side of c with distance $\leq 1/15k$ between them. Recall that we query v with all candidates in $\widehat{W}$, and so we have no uncertainty as to whether v approves of $c_S^{\leftarrow}$ or $c_S^{\rightarrow}$. Assuming that the voter has approved of some query and a_v, b_v are the endpoints of their approval interval, if $a_v \in (c_S^{\leftarrow}, (q_{j+1})_C^{\leftarrow}]$ and $b_v \in [(q_j)_C^{\rightarrow}, c_S^{\rightarrow})$, then X must occur, and so we have $\Pr(X) \leq 2p_t\left(\delta^+ + \frac{1}{15k}\right)\left(\delta^- + \frac{1}{15k}\right) \leq \frac{\ell}{k} - \frac{1}{10k}$. $s_{\ell,i,j}$ behaves like a Binomial random variable, and so we can apply a Hoeffding bound [14]: $\Pr\left(\text{Bin}\left(n, \frac{\ell}{k} - \frac{1}{10k}\right) \geq \frac{\ell n}{k} - \frac{n}{12k}\right) \leq \exp\left(-\alpha n/k^2\right)$ to get our result. □

Recall the definition of $Y_{c,\ell,r}$ given in Definition 9. The idea is that for a given voter v, while we do not directly observe whether the event $Y_{c,\ell,r}$ has occurred, we perform queries sufficiently close to c such that the probability that $\texttt{poss}(v, c, \widehat{W} \setminus R))$ evaluates to true is within $3/20k$ of $\pi(c, \ell, r)$, the true probability of $Y_{c,\ell,r}$. Since $\pi(c, \ell, r) \leq \ell/k - 1/4k$, we find that the probability that $\texttt{poss}(v, c, \widehat{W} \setminus R))$ evaluates to true, and therefore that v gets counted towards $s_{\ell,i,j}$ is $\leq \pi(c, \ell, r) + 3/20k \leq \ell/k - 1/10k$. We can then use a Hoeffding bound [14] to show that $\Pr(s_{\ell,i,j} \geq \frac{\ell n}{k} - \frac{n}{12k}) \leq \exp\left(-\frac{\alpha n}{k^2}\right)$. We can now prove Theorem 4.

Proof of Theorem 4. First, the output of Algorithm 2 always provides PJR+:[3] W^* provides PJR+, and if $\widehat{W}$ is output by the algorithm, the voters in any large cohesive group would be counted by some $s_{\ell,i,j}$ or by u.

Let us bound the probability that $u + s_{\ell,i,j} \geq n\ell/k$. First, let us bound u. We want to find the probability a voter does not approve of any point or interval query made during $\texttt{query}(v, C, P)$. Recall the proof of Theorem 2, where Z' is the event that no query to v is approved; we have $\Pr(v \in U) \leq \Pr(Z') \leq 1/|P| \leq 1/15k$. We have a standard Hoeffding bound [14] with $\alpha = 1800^{-1}$: $\Pr\left(u \geq \frac{n}{12k}\right) \leq \Pr\left(\text{Bin}\left(n, \frac{1}{15k}\right) \geq \frac{n}{12k}\right) \leq \exp\left(-\frac{\alpha n}{k^2}\right)$. We can also bound $s_{\ell,i,j}$ with Lemma 10. By the union bound, the probability that $s_{\ell,i,j} \geq \frac{\ell n}{k} - \frac{1}{12k}$ for any $s_{\ell,i,j}$ is at most $k^2 m \exp\left(-\alpha n/k^2\right)$. Finally, combining that by union bound with the probability that $u \geq n/12k$, we get that the probability that $u + s_{\ell,i,j} \geq n\ell/k$ is at most $(k^2 m + 1) \exp\left(-\alpha n/k^2\right)$.

Hence, if $n = \Omega(k^2 \log(m^2 k^2)) = \Omega(k^2 \log m)$, we see that the expected number of queries per voter $E(Q) = \mathcal{O}(\log(\sigma k))$: we use $\mathcal{O}(\log \sigma)$ queries to determine which segment v lies in, and the expected number of queries is

$$E(Q) \leq (k^2 m + 1) m \exp\left(-\alpha n/k^2\right) + \mathcal{O}(\log \sigma) + \mathcal{O}(\log |P|) \leq \mathcal{O}(\log \sigma |P|).$$

With $|P| = \mathcal{O}(k)$, we have $E(Q) \leq \mathcal{O}(\log(\sigma k))$. □

7 Conclusion and Future Work

We introduced a new probabilistic voter model for multiwinner voting with one-dimensional approval preferences, which we call the Random Interval Voter model (RIV), as well as a query framework for this model. Given an RIV election, we can find a committee that satisfies PJR+ using $\mathcal{O}(\log \sigma k)$ queries per voter in expectation.

An immediate open question is whether our approach can be extended to EJR+ —a strictly stronger notion of fairness than PJR+, which requires each ℓ-cohesive group to contain some voter v with $|A(v) \cap W| \geq \ell$ [4]. To adapt our analysis to EJR+, we would have to bound the sum $\sum_{t=0}^{\ell-1} \left(d^-_{r_1}(c) - d^-_{r_1-1}(c)\right) d^+_{\ell-1-r_2}(c) \leq (\ell - \lambda)/2k_t$ for some constant $\lambda > 0$. Using techniques similar to those in Section 6, we can obtain a bound on $d^-_{r_1}(c) - d^-_{r_1-1}(c)$ that is tight for each individual value of r_1, in that there exists an election where the bound matches for a given r_1; however, this bound is loose in that we cannot construct an election such that the bound is tight for all r_1, r_2, and it is too loose to prove EJR+. Indeed, the elections that are tight for a given r_1 are often very loose for other values of r, so we would need a more sophisticated bound that simultaneously considers multiple values of r.

The query complexity, as defined in our work, can be seen as a crude heuristic for *cognitive burden*. We believe that $\mathcal{O}(\log \sigma k)$ is a reasonable measure of what a voter can evaluate; however, the constant in the analysis may be too large. We note that we can reduce $|P|$ used in Algorithm 2 by a constant factor, at the expense of requiring n to be larger by a constant factor.

One particular direction of interest is to consider lower bounds on the amount of information required in this model to guarantee forms of justified representation. The fact that we focus on *expected* rather than maximum number of queries makes it more challenging to establish lower bounds.

Alternatively, one can model voters as having an ideal position on the interval and a radius around that position they are willing to approve, as previously discussed by a number of authors [3, 9, 26]. In addition, we could also consider higher dimensions. E.g., Lewenberg et al. [18] suggest that political opinions in the UK lie in a 10-dimensional space, so expanding our model to higher dimensions may be necessary to accurately model real-world political elections.

[3] Details can be found in Appendix E.

References

[1] Haris Aziz and Barton E Lee. Proportionally representative participatory budgeting with ordinal preferences. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5110–5118, 2021.

[2] 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.

[3] Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, and Rolf Niedermeier. An experimental view on committees providing justified representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019*, pages 109–115, 2019.

[4] Markus Brill and Jannik Peters. Robust and verifiable proportionality axioms for multiwinner voting. In *Proceedings of the 24th ACM Conference on Economics and Computation, ACM EC 2023*, page 301, 2023.

[5] Vincent Conitzer and Tuomas Sandholm. Vote elicitation: Complexity and strategy-proofness. In *Proceedings of the 18th National Conference on Artificial Intelligence, AAAI 2002*, pages 392–397, 2002.

[6] Lihi Naamani Dery, Meir Kalech, Lior Rokach, and Bracha Shapira. Reducing preference elicitation in group decision making. *Expert Syst. Appl.*, 61:246–261, 2016.

[7] Palash Dey and Arnab Bhattacharyya. Sample complexity for winner prediction in elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015*, pages 1421–1430, 2015.

[8] Edith Elkind and Martin Lackner. Structure in dichotomous preferences. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 2019–2025, 2015.

[9] Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. Justifying groups in multiwinner approval voting. *Theor. Comput. Sci.*, 969: 114039, 2023.

[10] Piotr Faliszewski, Jarosław Flis, Dominik Peters, Grzegorz Pierczyński, Piotr Skowron, Dariusz Stolicki, Stanisław Szufa, and Nimrod Talmon. Participatory budgeting: Data, tools and analysis. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence, IJCAI 2023*, pages 2667–2674, 2023.

[11] Luis Sánchez Fernández, Edith Elkind, Martin Lackner, Norberto Fernández García, Jesús Arias-Fisteus, Pablo Basanta-Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pages 670–676, 2017.

[12] Michał Tomasz Godziszewski, Paweł Batko, Piotr Skowron, and Piotr Faliszewski. An analysis of approval-based committee rules for 2D-Euclidean elections. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5448–5455, 2021.

[13] Daniel Halpern, Gregory Kehne, Ariel D Procaccia, Jamie Tucker-Foltz, and Manuel Wüthrich. Representation with incomplete votes. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, pages 5657–5664, 2023.

[14] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[15] Aviram Imber, Jonas Israel, Markus Brill, and Benny Kimelfeld. Approval-Based Committee Voting under Incomplete Information. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, pages 5076–5083, 2022.

[16] Aviram Imber, Jonas Israel, Markus Brill, Hadas Shachnai, and Benny Kimelfeld. Spatial voting with incomplete voter information. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence, AAAI 2024*, pages 9790–9797, 2024.

[17] Martin Lackner and Piotr Skowron. *Multi-Winner Voting with Approval Preferences.* Springer International Publishing, Cham, 2023.

[18] Yoad Lewenberg, Yoram Bachrach, Lucas Bordeaux, and Pushmeet Kohli. Political dimensionality estimation using a probabilistic graphical model. In *Proceedings of the 32nd Conference on*

Uncertainty in Artificial Intelligence, UAI 2016, pages 447–456, 2016.

[19] Tyler Lu and Craig Boutilier. Multi-winner social choice with incomplete preferences. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence, IJCAI 2023*, pages 263–270, 2013.

[20] Debmalya Mandal, Nisarg Shah, and David P Woodruff. Optimal Communication-Distortion Tradeoff in Voting. In *Proceedings of the 21st ACM Conference on Economics and Computation, ACM EC 2020*, pages 795–813, 2020.

[21] Reshef Meir, Omer Lev, and Jeffrey S Rosenschein. A local-dominance theory of voting equilibria. In *Proceedings of the ACM Conference on Economics and Computation, ACM EC 2014*, pages 313–330. ACM, 2014.

[22] Nicholas R. Miller. The spatial model of social choice and voting. In *Handbook of Social Choice and Voting*, chapter 10, pages 163–181. Edward Elgar Publishing Limited, 2018.

[23] Dominik Peters and Martin Lackner. Preferences Single-Peaked on a Circle. *J. Artif. Intell. Res.*, 68: 463–502, 2020.

[24] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation, ACM EC 2020*, pages 793–794, 2020.

[25] Grzegorz Pierczyński and Piotr Skowron. Core-stable committees under restricted domains. In *Proceedings of the 18th International Conference on Web and Internet Economics, WINE 2022*, pages 311–329, 2022.

[26] Stanisław Szufa, Piotr Faliszewski, Łukasz Janeczko, Martin Lackner, Arkadii Slinko, Krzysztof Sornat, and Nimrod Talmon. How to sample approval elections? In *Proceedings of the 31st International Joint Conference on Artificial Intelligence, IJCAI 2022*, pages 496–502, 2022.

[27] Zoi Terzopoulou, Alexander Karpov, and Svetlana Obraztsova. Restricted domains of dichotomous preferences with possibly incomplete information. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5726–5733, 2021.

[28] Zhibing Zhao, Haoming Li, Junming Wang, Jeffrey O Kephart, Nicholas Mattei, Hui Su, and Lirong Xia. A cost-effective framework for preference elicitation and aggregation. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence, UAI 2018*, pages 446–456, 2018.

Drew Springham
Department of Informatics, King's College London
London, United Kingdom
Email: drew.springham@kcl.ac.uk

Edith Elkind
Department of Computer Science, Northwestern University
Evanston, IL, United States
Email: edith.elkind@northwestern.edu

Bart de Keijzer
Department of Informatics, King's College London
London, United Kingdom
Email: bart.de_keijzer@kcl.ac.uk

Maria Polukarov
Department of Informatics, King's College London
London, United Kingdom
Email: maria.polukarov@kcl.ac.uk

Appendix

A Uniformisation Is Equivalent

Fix a general RIV model $\mathcal{M} = (I_t, F_t, p_t)_{t\in[\sigma]}$ and a candidate set $C \subset \bigcup_{t\in[\sigma]} I_t$. Let $\mathcal{M}' = (I'_t, F'_t, p_t)_{t\in[\sigma]}$ be a uniform RIV model where for each $t \in [\sigma]$ it holds that $I_t = [t, t+1]$ and F_t is the uniform distribution over I_t. We define a mapping $\mu : \bigcup_{t\in[\sigma]} I_t \to [1, \sigma+1]$, where for each $t \in [\sigma]$ and $x \in I_t$ we set $\mu(x) = t + F_t(x)$.

Consider a voter v with endpoints a_v, b_v and $A^*(v) = [a_v, b_v]$, whose approvals are drawn from $\mathcal{M}$. Then $\mu(a_v), \mu(b_v)$ are distributed according to $\mathcal{M}'$, and, moreover, $x \in A^*(v)$ if and only if $\mu(x) \in \mu(A^*(v))$.

This transformation enables us to state all our results for uniform RIVs. Indeed, given a general RIV $\mathcal{M}$, we can first map the candidate set C to $\mu(C)$, and then run the algorithms on this new set of candidates assuming a uniform RIV. Whenever an algorithm needs to perform a point query $\textsc{Point}(x, v)$ with $x \in \mu(C)$ in $\mathcal{M}'$, we reverse the transformation, and perform a point query $\textsc{Point}(\mu^{-1}(x), v)$; we use a similar approach for interval queries.

Given an election E where voters' preferences are drawn from $\mathcal{M}$, we denote by $\mu(E)$ the election where each candidate's position and all voters' endpoints are replaced with their images under μ. Then an outcome $W \subseteq \mu(C)$ of $\mu(E)$ can be converted to an outcome $\mu^{-1}(W) \subseteq C$ for E. We claim that W satisfies PJR+/core stability for $\mu(E)$ if and only if $\mu^{-1}(W)$ satisfies PJR+/core stability for E. Indeed, consider a set of voters $G \subseteq V$ with $|G| \geq n\ell/k$ with $A(G) \setminus W \neq \varnothing$. Then

$$\mu(\mu^{-1}(W) \cap \bigcup_{v\in G} A(v)) = W \cap \bigcup_{v\in G} \mu \circ A(v)$$

and so

$$\begin{aligned} |\mu^{-1}(W) \cap \bigcup_{v\in G} A(v)| < \ell \text{ if and only if} \\ |W \cap \bigcup_{v\in G} \mu \circ A(v)| < \ell. \end{aligned}$$

Therefore W satisfies PJR+ in the election $\mu(E)$ if and only if $\mu^{-1}(W)$ satisfies PJR+ in E.

Similarly, let $v \in V, T \subseteq C, W \subseteq \mu(C)$. Then for every $v \in V$ it holds that

$$\begin{aligned} |A(v) \cap T| \leq& |A(v) \cap \mu^{-1}(W)| \text{ if and only if} \\ |\mu(A(v) \cap T)| \leq& |\mu(A(v) \cap \mu^{-1}(W))|, \end{aligned}$$

and $\mu(A(v) \cap T) = \mu \circ A(v) \cap \mu(T)$, $\mu(A(v) \cap \mu^{-1}(W)) = \mu \circ A(v) \cap W$, as desired.

B Proof of Theorem 2

Proof of Theorem 2. Consider a voter v, and some set $P \subseteq C$. Let $\ell = \lceil \log |P| \rceil$. The initial binary search stage to identify t such that $v \in I_t$ takes $\mathcal{O}(\log |P|) = \mathcal{O}(\ell)$ queries. If t is not found, then we use only $\mathcal{O}(\ell)$ queries. Consider then when t is found. Let I be the random variable indicating in which round (value of i) Algorithm 1 terminates ($I = \ell + 1$ if the algorithm has to query all of P). Let I' be the smallest value of $i' \in [\ell]$ such that $A^*(v) \cap \left\{t + j/2^{i'} : j \in [2^{i'}]\right\} \neq \varnothing$ (or $I' = \ell + 1$ otherwise), so I' is also a random variable. Consider some $x \in A^*(v) \cap \left\{t + j/2^{I'} : j \in [2^{I'}]\right\}$. We know that

$I \leq I'$; the algorithm in round I' would perform point queries at $x_C^{\leftarrow}$ and $x_C^{\rightarrow}$, and an interval query for $\left(x_C^{\leftarrow}, x_C^{\rightarrow}\right)$. The voter would have to approve of at least one of these 3 queries; since $x \in A^*(v)$, we have $a_v \leq x \leq b_v$, and so if the interval query was not approved of, then the voter must approve one of the point queries. Of course, the algorithm may have terminated before reaching round I', in which case $I < I'$, but it is certainly true that $I \leq I'$. Let Z be the event $I = \ell + 1$ and Z' the event that $I' = \ell + 1$. Now, by the same argument that showed $I \leq I'$, we know that $\Pr(Z) \leq \Pr(Z')$; if none of the queries in the ℓ rounds were approved, then it cannot be the case that any of the points $\left\{t + j/2^l : j \in [2^l]\right\}$ would be approved. We have that $\Pr(Z') = 2^{-\ell} \leq 1/|P|$.

Let Q be the number of queries used by the algorithm in total. We have that if $I \leq \ell$, then $Q \leq \mathcal{O}(l) + 3 \times 2^I + 2\log|P| \leq 3 \times 2^{I'} + \mathcal{O}(l)$, and if $I = \ell + 1$, then $Q \leq 3 \times 2^\ell + |P| \leq 3 \times 2^{I'} + |P|$. We have $E(2^{I'}) = \sum_{r=1}^{\ell} 2^r \Pr(I' = r) + 2^{\ell+1}\Pr(Z') \leq \ell + 4$ where $\Pr(I' = r) = \Pr(I' = r \mid I' > r-1)\Pr(I' > r-1) = \frac{1}{2}\frac{1}{2^{r-1}}$ and hence $\Pr(I' = r) = 2^{-r}$. Therefore

$$
\begin{aligned}
E(Q) &\leq \mathcal{O}(\ell) + 3E\left(2^{I'}\right) + E\left(\begin{cases} |P| & \text{if } I = \ell + 1 \\ \mathcal{O}(\ell) & \text{otherwise} \end{cases}\right) \\
&\leq \mathcal{O}(\ell) + \mathcal{O}(l) + |P|\Pr(Z') = \mathcal{O}(\log(|P|),
\end{aligned}
$$

and hence the result is proved. ☐

C Proof of Lemma 7

Proof. Clearly, we have that $d_{r_1}^+(c) + d_{r_2}^-(c) \leq 1$, and so if $r_1 + r_2 + 1 \geq \frac{k_t}{2} + 1$ we get the result. So let us assume $r_1 + r_2 + 1 < \frac{k_t}{2} + 1$. Then it cannot be true that both $\rho(c) + r_1 \geq k_t$ and $\rho(c) - r_2 < 1$; if that were true, then $r_1 + r_2 + 1 \geq k_t - \rho(c) + \rho(c) = k_t$ which contradicts $r_1 + r_2 + 1 < \frac{k_t}{2}$. We consider these cases.

Case where both $\rho(c) + r_1 < k_t$ and $\rho(c) - r_2 \geq 1$; Suppose that there exist $\geq r_1$ midpoints in I_t to the right of c. Then let $\mu_1^+ = t + \frac{2i-1}{2(k_t+2)}$ be the closest midpoint to the right of c, and $\mu_h^+ = t + \frac{2(i+h-1)-1}{2(k_t+2)}$ the h'th closest midpoint to the right of c. Then there exists a candidate chosen for $\widehat{W}_t$ that lies in $(c, \mu_h^+ + (\mu_h^+ - c)]$; otherwise c would have been chosen for midpoint μ_h^+. Therefore within $(c, 2\mu_{r_1+1}^+ - c]$ there exist at least $r_1 + 1$ members of $\widehat{W}_t$.

Hence $\widehat{W}_t[\rho(c) + r_1] \leq 2\mu_{r_1+1}^+ - c$ and so $d_{r_1}^+(c) \leq 2(\mu_{r_1+1}^+ - c) = 2(\mu_{r_1+1}^+ - \mu_1^+ + \mu_1^+ - c) = \frac{2r_1}{k_t+2} + 2(\mu_1^+ - c)$. By similar logic, we get that $d_{r_2}^-(c) \leq \frac{2r_2}{k_t+2} + 2(c - \mu_1^-)$ if there exist $\geq r_2$ midpoints to the left of c.

Now suppose there there are $< r_1$ midpoints to the right of c. We have that $r_1 > 1$: there always exists at least one midpoint to the right of c in this case since $c < t + 1 - \frac{3}{2(k_t+2)}$. Then $c \geq t + \frac{2(k_t+2-r_1)-1}{2(k_t+2)}$ and therefore $d_{r_1}^+(c) \leq 1 - \frac{2(k_t+2-r_1)-1}{2(k_t+2)} = \frac{2r_1+1}{2(k_t+2)} < \frac{2r_1}{k_t+2} + 2(\mu_1^+ - c)$ since $r_1 > 1$. Again by similar logic we get that $d_{r_2}^-(c) \leq \frac{2r_2}{k_t+2} + 2(c - \mu_1^-)$ if there exist $< r_2$ midpoints to left of c.

Hence we have that

$$
d_{r_1}^+(c) + d_{r_2}^-(c) \leq \frac{2r_1}{k_t+2} + 2(\mu_1^+ - c) + \frac{2r_2}{k_t+2} + 2(c - \mu_1^-) = \frac{2(r_1 + r_2 + 1)}{k_t + 2},
$$

where the last equality holds because $\mu_1^+ - \mu_1^- = \frac{1}{k_t+2}$.

We now have dealt with the first case in which both $\rho(c) + r_1 + 1 \leq k_t$ and $\rho(c) - r_2 \geq 1$. Now let us look at the case when one of these conditions does not hold.

Case where $\rho(c) + r_1 < k_t$ and $\rho(c) - r_2 < 1$ (and vice versa) Then $d^-_{r_2}(c) = c - t$ and $d^+_{r_1}(c) + d^-_{r_2}(c) = \widehat{W}_t[\rho(c) + r_1] - t$. Note that if $c \notin \widehat{W}_t$, and there exist r midpoints to the left of c, then $\rho(c) \geq r$; otherwise there exists some midpoint that has chosen a candidate to be in $\widehat{W}_t$ which lies to the right of c, further away from the midpoint than c, and so hence c should have been chosen instead, a contradiction. So then $\rho(c) \geq \lfloor (k_t + 2)(c - t) + \frac{1}{2} \rfloor - 1$ for $c \geq t + \frac{3}{2(k+2)}$.

Note also then that there must exist at least r_1 mid points to the right of c. If there are fewer than r_1 midpoints to the right of c, then there are at least $k_t - (r_1 - 1)$ mid points to the left of c. Since $c \notin \widehat{W}_t$, there must exist at least $k_t - r_1 + 1$ members of $\widehat{W}_t$ to the left of c, so $\rho(c) \geq k_t - r_1 + 1$, which contradicts $\rho(c) + r_1 + 1 \leq k_t$.

By similar argument in the previous case, we have that $\widehat{W}_t[\rho(c) + r_1] \leq 2\mu^+_{r_1+1} - c$. We have that $\mu^+_{r_1+1} - c \leq (r_1 + 1)/(k_t + 2)$ and so

$$d^+_{r_1}(c) + d^-_{r_2}(c) = \widehat{W}_t[\rho(c) + r_1] - t \leq 2\mu^+_{r_1+1} - c - t = c - t + 2(\mu^+_{r_1+1} - c) \leq c - t + 2\frac{r_1 + 1}{k_t + 2}.$$

Suppose to the contrary that $c - t > \frac{2r_2}{k_t+2}$. Then

$$r_2 \geq \rho(c) \geq \lfloor (k_t + 2)(c - t) + \frac{1}{2} \rfloor - 1 \geq \lfloor 2r_2 + \frac{1}{2} \rfloor - 1 = 2r_2 - 1,$$

which is a contradiction for $r_2 > 1$. $\rho(c) \neq 0$ since $c - t \geq \frac{3}{2(k_t+2)}$ and $c \notin \widehat{W}_r$, so therefore $1 = r_2 = \rho(c)$ is the only case we need to be concerned with. Again, since $c \notin \widehat{W}_r$, we know that $c - t \leq \frac{5}{2(k_t+2)}$ since $\rho(c) = 1$, and so $\mu^+_h = \frac{2h+3}{2(k_r+2)}$ and so

$$2\mu^+_{r_1+1} - c - t \leq \frac{2(r_1 + 1) + 3}{k_t + 2} - \frac{3}{2(k_t + 2)} = \frac{2r_1 + 3.5}{k_t + 2} < \frac{2(r_1 + r_2 + 1)}{k_t + 2}.$$

Therefore $d^+_{r_1}(c) + d^-_{r_2}(c) \leq \frac{2(r_1+r_2+1)}{k_t+2}$.

Using similar reasoning, we get the same result if $\rho(c) + r_1 + 1 > k_t$ and $\rho(c) - r_2 \geq 1$. $\square$

D Proof of Lemma 11

Now we consider when $c \in (C_t \setminus K_t) \setminus \widehat{W}_t$.

Lemma 11. *For $c \in (C_t \setminus K_t) \setminus \widehat{W}_t$, $\pi(c, \ell, r) \leq \ell/k - 1/4k$.*

Proof. $\Pr(c \in A(v)) \leq 2p_t \frac{3}{2(k_t+2)}$, and so when $\ell \geq 4$, we have $2p_t \frac{3}{2(k_t+2)} \leq \ell/k - 1/4k$. Also, if $k_t < \ell$, then $p_t k < \ell$, and $\Pr(c \in A(v)) \leq p_t/2 < \ell/2k \leq \ell/k - 1/4k$, so assume now that $\ell \leq k_t$ and $\ell \leq 3$. Say wlog $c \leq t + \frac{3}{2(k_t+2)}$ (otherwise we can consider mirroring the segment to get the result for $c \geq t + 1 - \frac{3}{2(k_t+2)}$), and so there exists $\geq \ell$ midpoints to the right of c. Consider the candidate $y \in \widehat{W}_t$ that was selected for the midpoint $p = t + \frac{2\ell+1}{2(k_t+1)}$. Then $|y - p| \leq |c - p|$, and $|\widehat{W}_t \cap [c, p]| \geq \ell$. We have $\Pr(c \in A(v) \wedge |A(v) \cap W_t| < \ell) \leq 2p_t(c - t)|c - y| \leq 4p_t(c - t)(p - c) = 4p_t(c - t)\left(t + \frac{2\ell+1}{2(k_t+2)} - c\right)$. We have 2 cases. If $\ell = 1$ or $\ell = 2$, then

$$4p_t(c - t)\left(t + \frac{2\ell + 1}{2(k_t + 2)} - c\right) \leq 4p_t\left(\frac{2\ell + 1}{4(k_t + 2)}\right)^2 \leq \frac{(4\ell^2 + 4\ell + 1)p_t}{4\,(k_t + 2)^2} \leq \ell/k - 1/4k.$$

If $\ell = 3$, then recalling that $c < \frac{3}{2(k_t+2)}$,

$$4p_t(c - t)\left(t + \frac{2\ell + 1}{2(k_t + 2)} - c\right) \leq 4p_t \frac{3}{2(k_t + 2)} \frac{4}{2(k_t + 2)} \leq \frac{12p_t}{(k_t + 2)^2} \leq \frac{\ell}{k} - \frac{1}{4k},$$

except when $p_t k < 1$. However in this case, we have $\Pr(c \in A(v)) \leq 2p_t \frac{3}{2(k_t+2)} \leq \frac{3}{2k} \leq \frac{3}{k} - \frac{1}{4k}$ and hence we are done. $\square$

E Proof that $\widehat{W}$ provides PJR+

Suppose that $\widehat{W}$ is ouput by Algorithm 2. $\widehat{W}$ is output if and only if $s_{l,i,j} + u < nl/k$ for all l, i, j. Suppose to the contrary that there exists $G \subseteq V$ with $|G| \geq nl/k$ where $|\widehat{W} \cap \bigcup_{v \in G} A(v)| < l$ for some l but there exists $c \in A(G) \setminus \widehat{W}$ (with $c = C[i]$ for some i). We will show that this implies $s_{l,i,j} + u \geq nl/k$ for some j, a contradiction. Let

$$H = \widehat{W} \cap \bigcup_{v \in G} A(v) \qquad S = \widehat{W} \setminus H$$
$$L = \{t \in H : t < c\} \qquad j = |L|$$

Now, for each $v \in G$, we have $\texttt{poss}(v, c, S)$ (as long as they approve of some query): $\phi_2 \leq c \leq \phi_3$, since all voters approve of c. $\phi_1 \geq \max(S_-)$ and $\phi_4 \leq \min(S_+)$, since $A(v) \subseteq \cup_{v' \in G} A(v')$, so no voter can approve a candidate in S. Since $\phi_1 < \phi_2 < \phi_3 < \phi_4$, we have $\texttt{poss}(v, c, S)$ for all $v \in G$ who approve of a query, and any voter that does not approve of any query gets counted in u and hence $u + s_{l,i,j} \geq |G| \geq nl/k$ since $G \in \mathcal{L}_l$. This contradicts $s_{l,i,j} + u < nl/k$, therefore if $\widehat{W}$ is outputted by Algorithm 2, then it provides PJR+ for the election.