Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Proportional Justified Representation

Luis Sánchez-Fernández
Universidad Carlos III de Madrid, Spain
luiss@it.uc3m.es

Edith Elkind
University of Oxford, United Kingdom
elkind@cs.ox.ac.uk

Martin Lackner
University of Oxford, United Kingdom
martin.lackner@cs.ox.ac.uk

Norberto Fernández
Escuela Naval Militar (CUD), Spain
norberto@cud.uvigo.es

Jesús A. Fisteus
Universidad Carlos III de Madrid, Spain
jaf@it.uc3m.es

Pablo Basanta Val
Universidad Carlos III de Madrid, Spain
pbasanta@it.uc3m.es

Piotr Skowron
University of Oxford, United Kingdom
p.k.skowron@gmail.com

Abstract

The goal of multi-winner elections is to choose a fixed-size committee based on voters' preferences. An important concern in this setting is *representation*: large groups of voters with cohesive preferences should be adequately represented by the election winners. Recently, Aziz et al. (2015a; 2017) proposed two axioms that aim to capture this idea: *justified representation (JR)* and its strengthening *extended justified representation (EJR)*. In this paper, we extend the work of Aziz et al. in several directions. First, we answer an open question of Aziz et al., by showing that Reweighted Approval Voting satisfies JR for $k = 3, 4, 5$, but fails it for $k \geq 6$. Second, we observe that EJR is incompatible with the *Perfect Representation* criterion, which is important for many applications of multi-winner voting, and propose a relaxation of EJR, which we call *Proportional Justified Representation (PJR)*. PJR is more demanding than JR, but, unlike EJR, it is compatible with perfect representation, and a committee that provides PJR can be computed in polynomial time if the committee size divides the number of voters. Moreover, just like EJR, PJR can be used to characterize the classic PAV rule in the class of weighted PAV rules. On the other hand, we show that EJR provides stronger guarantees with respect to average voter satisfaction than PJR does.

1 Introduction

Decision-making based on the aggregation of possibly conflicting preferences is a central problem in the field of social choice, which has received a considerable amount of attention from the artificial intelligence researchers (Conitzer 2010; Brandt et al. 2016). The most common preference aggregation scenario is the one where a single candidate has to be selected. However, there are also many applications where the goal is to select a fixed-size set of alternatives: example range from choosing a parliament or a committee to identifying a set of plans, allocating resources, shortlisting candidates for a job or an award, picking movies to be shown on a plane or creating a conference program (Barberà and Coelho 2008; Monroe 1995; Elkind, Lang, and Saffidine 2015; Skowron, Faliszewski, and Slinko 2015; Elkind et al. 2017). Recently, the complexity of multi-winner voting rules (Betzler, Slinko, and Uhlmann 2013; Aziz et al. 2015b) and their social choice properties (Elkind, Lang, and Saffidine 2015; Aziz et al. 2015a; 2017; Elkind et al. 2017) have been actively explored by the artificial intelligence research community.

Multi-winner voting rules are often applied in scenarios in which the set of winners needs to represent the different opinions or preferences of the agents involved in the election. Thus, it is important to formulate axioms that capture our intuition about what it means for a set of winners to provide a faithful representation of voters' preferences (Monroe 1995; Dummet 1984; Black 1958). Aziz et al. (2015a; 2017) have recently proposed two such axioms for approval-based multi-winner voting, namely *justified representation (JR)* and *extended justified representation (EJR)*. Intuitively, JR requires that a large enough group of agents with similar preferences is allocated at least one representative; EJR says that if this group is large enough and cohesive enough, it deserves not just one, but several representatives (see Section 2 for formal definitions). Similar axioms have been proposed for multi-winner voting rules with ranked ballots (Dummet 1984; Elkind et al. 2017). Aziz et al. show that for every collection of ballots there is a winning set that provides EJR; they then explore a number of popular multi-winner voting rules and show that several of them satisfy JR, but only one rule satisfies EJR.

Our first contribution in this paper is to answer a question left open by Aziz et al. Specifically, Aziz et al. prove that Reweighted Approval Voting (RAV)[1] satisfies JR if the desired number of winners k is 2 but it fails JR if $k \geq 10$. We close this gap and prove that RAV satisfies JR if $k \leq 5$ and fails JR if $k \geq 6$. Our proof proceeds by constructing and solving a linear program that establishes bounds on RAV scores.

We then formulate an axiom that we call *Perfect Representation (PR)*, which says that if a given instance admits a 'perfect solution' (all voters are represented, and each winner represents the same number of voters), then we expect a voting rule to output such a solution. This axiom is very ap-

[1] In the journal version of their paper, Aziz et al. (2017) refer to this rule as Sequential Proportional Approval Voting (SeqPAV).

pealing in parliamentary elections and similar applications of multi-winner voting. However, it turns out to be incompatible with EJR: there is an election where these two axioms correspond to disjoint sets of winning committees.

Motivated by this result, we propose a relaxation of EJR, which we call *Proportional Justified Representation (PJR)*. PJR is more demanding than JR, but, unlike EJR, it is compatible with perfect representation, and a committee that provides PJR can be computed in polynomial time; in particular, we show that a well-studied efficiently computable voting rule satisfies PJR if the committee size k divides the number of voters n. In contrast, it is conjectured that finding committees that provide EJR is computationally hard. Moreover, just like EJR, PJR can be used to characterize the classic Proportional Approval Voting (PAV) in the class of weighted PAV rules. However, we then show that the additional flexibility supplied by PJR comes at a cost: we define a measure of average voter satisfaction and show that EJR provides much stronger guarantees with respect to this measure than PJR does. We conclude the paper by discussing our results and indicating directions for future work.

2 Preliminaries

Given a positive integer s, we denote the set $\{1, \dots, s\}$ by $[s]$. We consider elections with a set of *voters* $N = \{1, \dots, n\}$ and a set of *candidates* $C = \{c_1, \dots, c_m\}$. Each voter $i \in N$ submits an *approval ballot* $A_i \subseteq C$, which represents the subset of candidates that she approves of. We refer to the list $\mathcal{A} = (A_1, \dots, A_n)$ as the *ballot profile*. An *approval-based multi-winner voting rule* takes as input a tuple $(N, C, \mathcal{A}, k)$, where k is a positive integer that satisfies $k \leq |C|$, and returns a subset $W \subseteq C$ of size k, which we call the *winning set*, or *committee*. We omit N and C from the notation when they are clear from the context.

The following voting rules have received a considerable amount of attention in the literature (Kilgour 2010; Elkind et al. 2017; Aziz et al. 2015a; 2017):

Proportional Approval Voting (PAV) Under PAV, an agent is assumed to derive an utility of $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{j}$ from a committee that contains exactly j of her approved candidates, and the goal is to maximize the sum of the agents' utilities. Formally, the PAV-score of a set $W \subseteq C$ is defined as $\sum_{i \in N} r(|W \cap A_i|)$, where $r(p) = \sum_{j \in [p]} \frac{1}{j}$, and PAV outputs a set $W \subseteq C$ of size k with the highest PAV-score.

We can generalize the definition of PAV by using an arbitrary non-increasing score vector in place of $(1, \frac{1}{2}, \frac{1}{3}, \cdots)$: for every vector $\mathbf{w} = (w_1, w_2, \dots)$, where $w_1, w_2, \dots$ are non-negative reals, $w_1 = 1$ and $w_1 \geq w_2 \geq \dots$, we define a voting rule $\mathbf{w}$-PAV that, given a ballot profile $(A_1, \dots, A_n)$ and a target number of winners k, returns a set W of size k with the highest $\mathbf{w}$-PAV score, defined as $\sum_{i \in N} r_{\mathbf{w}}(|W \cap A_i|)$, where $r_{\mathbf{w}}(p) = \sum_{j \in [p]} w_j$.

Reweighted Approval Voting (RAV) RAV is a multi-round rule that in each round selects a candidate and then reweighs the approvals for the subsequent rounds. Specifically, it starts by setting $W = \emptyset$. Then in round $j, j \in [k]$, it computes the *approval weight* of each candidate c as

$$\sum_{i : c \in A_i} \frac{1}{1 + |W \cap A_i|},$$

selects a candidate with the highest approval weight, and adds her to W. Just as for PAV, we can extend the definition of RAV to score vectors other than $(1, \frac{1}{2}, \frac{1}{3}, \cdots)$: every vector $\mathbf{w} = (w_1, w_2, \dots)$, where $w_1, w_2, \dots$ are non-negative reals, $w_1 = 1$ and $w_1 \geq w_2 \geq \dots$ defines a sequential voting rule $\mathbf{w}$-RAV, which proceeds as RAV, except that it computes the approval weight of a candidate c in round j as $\sum_{i : c \in A_i} w_{|W \cap A_i| + 1}$, where W is the winning set after the first $j - 1$ rounds.

The Monroe rule For each voter $i \in N$ and each candidate $c \in C$ we write $u_i(c) = 1$ if $c \in A_i$ and $u_i(c) = 0$ if $c \notin A_i$. Given a committee $W \subseteq C$ of size k, we say that a mapping $\pi : N \to W$ is *valid* if it satisfies $|\pi^{-1}(c)| \in \left\{ \lfloor \frac{n}{k} \rfloor, \lceil \frac{n}{k} \rceil \right\}$ for each $c \in W$. The Monroe score of a valid mapping π is given by $\sum_{i \in N} u_i(\pi(i))$, and the Monroe score of W is the maximum score of a valid mapping from N to W. The Monroe rule returns a committee of size k with the maximum Monroe score.

The Greedy Monroe rule Given a ballot profile $\mathcal{A} = (A_1, \dots, A_n)$ over a candidate set C and a target committee size k, the Greedy Monroe rule proceeds in k rounds. It maintains the set of available candidates C' and the set of unsatisfied voters N'; initially $C' = C$ and $N' = N$. It starts by setting $W = \emptyset$. In round t, $t = 1, \dots, k$, it selects a candidate c_t from C' and a group of voters N_t from N' of size approximately $\frac{n}{k}$ (specifically, $\lceil \frac{n}{k} \rceil$ if $t \leq n - k\lfloor \frac{n}{k} \rfloor$, and $\lfloor \frac{n}{k} \rfloor$ if $t > n - k\lfloor \frac{n}{k} \rfloor$) so as to maximize the quantity $|\{i \in N_t : c_t \in A_i\}|$ over all possible choices of (N_t, c_t). The candidate c_t is then added to W, and we set $C' = C' \setminus \{c_t\}$, $N' = N' \setminus N_t$. We say that the candidates in N_t are *assigned* to c_t. After k rounds, the rule outputs W.

PAV and RAV were defined by Thiele (1895). The Monroe rule was proposed by Monroe (1995) and Greedy Monroe is due to Skowron, Faliszewski, and Slinko (2015) (Skowron, Faliszewski, and Slinko define this rule for the setting where ballots are rankings of the candidates; we adapt their definition to approval ballots). For PAV and Monroe finding a winning committee is NP-hard (Aziz et al. 2015b; Procaccia, Rosenschein, and Zohar 2008), whereas for RAV and Greedy Monroe winning committees can be computed in polynomial time; in fact, RAV and Greedy Monroe were originally proposed as approximation algorithms for PAV and Monroe, respectively.

Under each of the rules we consider, there may be more than one winning committee. In what follows, we assume that all ties are broken in some deterministic way; none of our results depends on the tie-breaking rule.

We will now define the key concepts in the work of Aziz et al. (2015a; 2017): *justified representation* and *extended justified representation*.

Definition 1. (Extended) justified representation ((E)JR) *Consider a ballot profile $\mathcal{A} = (A_1, \dots, A_n)$ over a candidate set C and a target committee size k, $k \leq |C|$. Given a*

positive integer $\ell \in [k]$, we say that a set of voters $N^* \subseteq N$ is ℓ-cohesive *if $|N^*| \geq \ell\frac{n}{k}$ and $|\bigcap_{i \in N^*} A_i| \geq \ell$. A set of candidates W is said to provide* ℓ-justified representation (ℓ-JR) *for $(\mathcal{A}, k)$ if there does not exist an ℓ-cohesive set of voters N^* such that $|A_i \cap W| < \ell$ for each $i \in N^*$. We say that W provides* justified representation (JR) *for $(\mathcal{A}, k)$ if it provides 1-JR for $(\mathcal{A}, k)$; it provides* extended justified representation (EJR) *for $(\mathcal{A}, k)$ if it provides ℓ-JR for $(\mathcal{A}, k)$ for all $\ell \in [k]$. An approval-based voting rule* satisfies ℓ-JR *if for every ballot profile $\mathcal{A}$ and every target committee size k it outputs a committee that provides ℓ-JR for $(\mathcal{A}, k)$. A rule* satisfies *JR (respectively, EJR) if it satisfies ℓ-JR for $\ell = 1$ (respectively, for all $\ell \in [k]$).*

By definition, EJR implies JR. Aziz et al. (2015a; 2017) show that PAV satisfies EJR (and hence JR), Monroe satisfies JR, but fails EJR, and RAV fails JR for sufficiently large values of k; they do not consider Greedy Monroe in their work.

3 Justified Representation and RAV

Aziz et al. (2015a; 2017) prove that RAV satisfies JR for $k = 2$, but fails it for $k \geq 10$. Whether RAV satisfies JR for $k = 3, \ldots, 9$ was left an open problem. The following theorem provides a complete answer.

Theorem 1. *RAV satisfies JR for $k \leq 5$ but fails it for $k \geq 6$.*

Proof. For each $k \in \mathbb{N}$, we construct a linear program LP_k whose value is the maximum possible 'relative' approval weight of a yet unelected candidate (i.e., the ratio between her approval weight and the total number of voters) after $k-1$ steps of RAV. Fix a $k \in \mathbb{N}$. We can assume without loss of generality that RAV elects a committee $\{c_1, \ldots, c_k\}$, where for $i \in [k]$ candidate c_i is added to the committee in round i; our linear program includes constraints that impose this order. Moreover, as non-elected candidates do not have any influence on the approval weight under RAV, we may assume that $C = \{c_1, \ldots, c_k\}$. For $i \in [k]$, we write $C_i = \{c_1, \ldots, c_i\}$.

LP_k has a variable x_A for each nonempty candidate subset $A \subseteq C_k$; this variable corresponds to the fraction of voters that submit the approval ballot A. The objective function of LP_k is the ratio of the approval weight of candidate c_k and the total number of agents n after candidates $\{c_1, \ldots, c_{k-1}\}$ have already been elected. The constraints say that all variables should be non-negative and sum up to 1, and that RAV can select c_i in round i, for $i \in [k-1]$.

$$\text{maximize} \sum_{A: c_k \in A} \frac{x_A}{1 + |C_{k-1} \cap A|} \quad \text{subject to}$$

$$x_A \geq 0 \quad \text{for all } A \subseteq C_k; \tag{1}$$

$$\sum_{A \subseteq C} x_A = 1; \tag{2}$$

$$\sum_{A: c_i \in A} \frac{x_A}{1 + |C_{i-1} \cap A|} \geq \sum_{B: c_j \in B} \frac{x_B}{1 + |C_{i-1} \cap B|} \tag{3}$$

$$\text{for } i = 1, \ldots, k-1 \text{ and for } j = i+1, \ldots, k.$$

The number of variables in LP_k grows exponentially with k, but this is not an issue, because we only have to solve this linear program for small values of k. Solving LP_k for $k = 3, 4, 5, 6$, we obtain the following result.

Lemma 1. *For $k = 6$ the value of LP_k is $0.204 > \frac{1}{k-1}$. For $k = 3, 4, 5$ the value of LP_k is smaller than $\frac{1}{k-1}$.*

Consider an optimal solution $(x_A)_{A \subseteq C_6}$ of LP_6. We can find a positive integer n such all values $n_A = x_A \cdot n$ for $A \subseteq C_6$ are integer, and construct an n-voter ballot profile $\mathcal{A} = (A_1, \ldots, A_n)$ where each ballot $A \subseteq C_6$ occurs exactly n_A times; moreover, we can pick n so that $n/5$ is an integer. Lemma 1 implies that when we execute RAV on $(\mathcal{A}, k)$, in each round RAV selects a candidate whose approval weight is at least $0.204n$.

Now, consider the ballot $\mathcal{A}' = (A'_1, \ldots, A'_{6n/5})$ over $C_7 = C_6 \cup \{c_7\}$ where $A'_i = A_i$ for $i \in [n]$ and each of the the additional $n/5$ voters has $A_i = \{c_7\}$. Suppose that we run RAV on $\mathcal{A}'$ with $k = 6$. As we have $\frac{1}{6} \cdot (6n/5) = n/5$, the JR axiom requires that c_7 is elected. However, this does not happen: at each point the approval weight of c_7 is $n/5$, whereas by Lemma 1 in each round RAV can find a candidate whose approval weight is at least $0.204n > n/5$. In the full version of the paper (Sánchez-Fernández et al. 2016), we provide a concrete implementation of this idea: we describe an election with 5992 candidates on which RAV fails JR for $k = 6$. We also explain how to extend this example to $k > 6$.

We will now show that RAV satisfies JR for $k = 3, 4, 5$. Suppose for the sake of contradiction that RAV violates JR for some n-voter ballot profile $\mathcal{A}$ and some $k \in \{3, 4, 5\}$. That is, for some way of breaking ties RAV outputs a winning set W and there exists a set of voters N^* of size at least $\frac{n}{k}$ such that all voters in N^* approve some candidate c, yet no voter in N^* approves any candidate in W. Let w be the candidate that is added to W during the k-th round.

Consider a ballot profile $\mathcal{A}'$ obtained from $\mathcal{A}$ by removing all voters in N^*, and suppose that we execute RAV on $(\mathcal{A}', k)$. It is possible to break intermediate ties in the execution of RAV so that RAV outputs W on $(\mathcal{A}', k)$, and, moreover, candidates are added to W in the same order on both inputs. Indeed, as none of the candidates in W is approved by the voters in N^*, removing these voters does not change the approval weights of the candidates in W in each round of RAV, and can only lower the scores of the other candidates. Thus, we can assume that RAV selects W, and, moreover, adds w to W at the k-th step.

We can now apply Lemma 1 to $\mathcal{A}'$, which contains at most $n' = n - \frac{n}{k} = n\frac{k-1}{k}$ voters; by the lemma, when w is added to W, its approval weight is strictly less than $\frac{n'}{k-1} = \frac{n}{k}$. Since none of the removed voters approved w, when RAV is executed on $\mathcal{A}$, candidate w's approval weight in the k-th round is also strictly less than $\frac{n}{k}$. But this means that RAV should have favored c over w in the k-th round, a contradiction. □

4 Perfect Representation

A key application of multi-winner voting is parliamentary elections, where an important goal is to select a committee that reflects as fairly as possible the different opinions or preferences that are present in a society. Fairness in this context means that each committee member should represent approximately the same number of voters and as many voters as possible should be represented by a committee member that they approve. From this perspective, the best-case scenario is when each voter is represented by a candidate that she approves and each winning candidate represents exactly the same number of voters. Thus, we may want our voting rules to output committees with this property whenever they exist. This motivates the following definition.

Definition 2. Perfect representation (PR) *Consider a ballot profile $\mathcal{A} = (A_1, \ldots, A_n)$ over a candidate set C, and a target committee size k, $k \leq |C|$, such that k divides n. We say that a set of candidates W, $|W| = k$,* provides perfect representation (PR) for $(\mathcal{A}, k)$ *if it is possible to partition N into k pairwise disjoint subsets $N_1, \ldots, N_k$ of size $\frac{n}{k}$ each and assign a distinct candidate from W to each of these subsets in such a way that for each $\ell \in [k]$ all voters in N_ℓ approve their assigned member of W. An approval-based voting rule* satisfies PR *if for every profile $\mathcal{A}$ and every target committee size k the rule outputs a committee that provides PR for $(\mathcal{A}, k)$ whenever such a committee exists.*

An example of a voting rule that satisfies PR is the Monroe rule: a committee that provides perfect representation for an n-voter ballot profile has the Monroe score of n, i.e., the maximum possible score, whereas the Monroe score of any committee that does not provide perfect representation is at most $n - 1$.

We note that the PR axiom is quite demanding from a computational perspective: the problem of deciding whether there exists a committee that provides PR for a given pair $(\mathcal{A}, k)$ is NP-complete.

Theorem 2. *Given a ballot profile $\mathcal{A}$ and a target committee size k, it is* NP*-complete to decide whether there exists a committee that provides PR for $(\mathcal{A}, k)$.*

Proof sketch. To show containment in NP, we reduce the problem of checking whether a given committee W provides perfect representation for $(\mathcal{A}, k)$ to the problem of finding a b-matching in a bipartite graph that can be associated with $\mathcal{A}$ and W; the latter problem admits a polynomial-time algorithm (Anstee 1987). To prove NP-hardness, we adapt a reduction of Procaccia, Rosenschein, and Zohar (2008), which shows that finding a winning committee under the Monroe rule is NP-hard. The details can be found in the full version of the paper (Sánchez-Fernández et al. 2016). □

Remark 1. *Theorem 2 immediately implies that, unless P=NP, RAV and Greedy Monroe fail PR; it is also not hard to construct specific examples on which these rules fail PR. As for PAV, Theorem 3 below implies that it also fails PR.*

Viewed from a different perspective, PR is a rather weak axiom: it only constrains the behavior of a voting rule on inputs that admit a committee that provides PR. In particular, this axiom has no bite if k does not divide n. Also, unlike EJR, PR does not engage with the idea that a voter may benefit from being represented by more than one candidate. Thus, we may want a voting rule to satisfy both PR and another representation axiom, such as, e.g., EJR. However, this turns out to be impossible: PR and EJR are incompatible.

Theorem 3. *There exists a ballot profile $\mathcal{A}$ and a target committee size k such that the set of committees that provide PR for $(\mathcal{A}, k)$ is non-empty, but none of the committees in this set provides EJR.*

Proof. Let $C = \{c_1, \ldots, c_6\}$, and consider a ballot profile $\mathcal{A} = (A_1, \ldots, A_8)$ where $A_i = \{c_i\}$, $A_{i+4} = \{c_i, c_5, c_6\}$ for $i = 1, \ldots, 4$. Observe that $W = \{c_1, c_2, c_3, c_4\}$ is the unique committee of size 4 that provides PR for $(\mathcal{A}, 4)$. However, W fails to provide EJR: $\{A_5, A_6, A_7, A_8\}$ is a 2-cohesive set of voters, but each of these voters only approves one candidate in W. □

This motivates the following question: can we find a weakening of the EJR axiom that still provides meaningful guarantees to large cohesive groups of voters, yet is compatible with PR? We address this question in the next section.

5 Proportional Justified Representation

The EJR axiom provides the following guarantee: at least one member of an ℓ-cohesive group has at least ℓ representatives in the committee. This focus on a single group member does not quite reflect our intuition of what it means for a group to be well-represented. A weaker and perhaps more natural condition is to require that collectively the members of an ℓ-cohesive group are allocated at least ℓ representatives. This idea is captured by the following definition.

Definition 3. Proportional justified representation (PJR) *Given a ballot profile $\mathcal{A} = (A_1, \ldots, A_n)$ over a candidate set C and a target committee size k, $k \leq |C|$, we say that a set of candidates W, $|W| = k$,* provides proportional justified representation (PJR) *for $(\mathcal{A}, k)$ if for every $\ell \in [k]$ and every ℓ-cohesive set of voters $N^* \subseteq N$ it holds that $|W \cap (\bigcup_{i \in N^*} A_i)| \geq \ell$. We say that an approval-based voting rule* satisfies proportional justified representation (PJR) *if for every ballot profile $\mathcal{A}$ and every target committee size k it outputs a committee that provides PJR for $(\mathcal{A}, k)$.*

It is immediate that every committee that provides PJR also provides JR: the PJR condition for $\ell = 1$ is exactly JR. Also, it is easy to see that every committee that provides EJR also provides PJR: the condition "$|A_j \cap W| \geq \ell$ for some $i \in N^*$" in the definition of EJR implies the condition "$|W \cap (\bigcup_{i \in N^*} A_i)| \geq \ell$" in the definition of PJR. To summarize, we obtain the following proposition.

Proposition 1. *EJR implies PJR, and PJR implies JR.*

Moreover, unlike EJR, PJR is compatible with PR.

Theorem 4. *For every profile $\mathcal{A} = (A_1, \ldots, A_n)$ and every target committee size k, if a set of candidates W, $|W| = k$, provides PR, then W also provides PJR.*

Proof. Observe that because W provides PR, k divides n. Let $W = \{w_1, \dots, w_k\}$. As W provides PR, there exist k pairwise disjoint subsets $N_1, \dots, N_k$ of size $\frac{n}{k}$ each such that all voters in N_i approve w_i for each $i \in [k]$. Consider a set of agents $N^* \subseteq N$ and a positive integer ℓ such that $|N^*| \geq \ell\frac{n}{k}$. By the pigeonhole principle, N^* has a non-empty intersection with at least ℓ of the sets $N_1, \dots, N_k$. As each voter in $N^* \cap N_i$ approves w_i, it follows that the the number of candidates in W approved by some voter in N^* must be greater than or equal to ℓ. □

Another advantage of PJR is that a committee that provides PJR can be computed in polynomial time as long as the target committee size k divides the number of voters n; indeed, under this condition both the Monroe rule and Greedy Monroe (the latter of which is polynomial-time computable) provide PJR. We note that PAV satisfies EJR and hence PJR even if k does not divide n; however, computing the output of PAV is NP-hard.

Theorem 5. *Consider a ballot profile $\mathcal{A} = (A_1, \dots, A_n)$. If the target committee size k divides n then the outputs of Monroe and Greedy Monroe on $(\mathcal{A}, k)$ satisfy PJR.*

Proof. We provide a proof for Greedy Monroe; the proof for Monroe can be found in the full version of the paper (Sánchez-Fernández et al. 2016).

Let $s = \frac{n}{k}$; note that $s \in \mathbb{N}$. Suppose for the sake of contradiction that the set W output by Greedy Monroe fails PJR for some $\ell \in [k]$ and some ℓ-cohesive set of voters N^*; we can assume that $|N^*| = \ell \cdot s$. Consider a candidate $c \in C \setminus W$ that is approved by all voters in N^*.

By the pigeonhole principle N^* has a non-empty intersection with at least ℓ of the sets $N_1, \dots, N_k$ constructed by Greedy Monroe (the integrality of s is crucial here); let the first ℓ of these sets be $N_{i_1}, \dots, N_{i_\ell}$ with $i_1 < \dots < i_\ell$. For each $t = 1, \dots, \ell$, pick a voter in $N_{i_t} \cap N^*$; note that all these voters are assigned to different candidates in W. Now, if each of these ℓ voters approves the candidate she is assigned to, we are done, as we have identified ℓ distinct candidates in W each of which is approved by some voter in N^*. Otherwise, let $j = \min\{i_t :$ the voter we chose in $N_{i_t} \cap N^*$ does not approve of $c_{i_t}\}$. By our choice of j, not all voters in N_j approve c_j, yet the pair (N_j, c_j) was chosen at step j. Among the sets $N_1, \dots, N_{j-1}$ there are at most $\ell - 1$ sets of size s each that have a non-empty intersection with N^*, so at step j at least s voters in N^* are present in N' (N' is the set of unsatisfied voters; see the definition of Greedy Monroe in section 2). Now, candidate c, together with s voters from $N^* \cap N'$, would be a better choice for Greedy Monroe than (N_j, c_j), a contradiction. □

We remark that if $\frac{n}{k}$ is not an integer, the proof of Theorem 5 breaks down, because N^* may be covered by fewer than ℓ sets among $N_1, \dots, N_k$. The following example shows that both of these rules may fail PJR in this case.

Example 1. *Let $n = 10$, $k = 7$, $C = \{c_1, \dots, c_8\}$. Suppose that $A_i = \{c_i\}$ for $i = 1, \dots, 4$ and $A_i = \{c_5, c_6, c_7, c_8\}$ for $i = 5, \dots, 10$. Let $\ell = 4$. Then $\ell \cdot \frac{n}{k} = \frac{40}{7} < 6$, so the set of voters $\{5, 6, 7, 8, 9, 10\}$ "deserves" four representatives. However, under both Monroe and Greedy Monroe only three candidates from $\{c_5, c_6, c_7, c_8\}$ will be selected.*

It is then natural to ask if there is a polynomial-time computable voting rule that satisfies PJR for all values of n and k. Interestingly, it turns out that the answer to this question is 'yes': in a very recent paper, Brill et al. (2017) describe an approval-based multi-winner rule developed by the Swedish mathematician Lars Edvard Phragmén more than 100 years ago, and show that a sequential variant of this rule, which they refer to as seq-Phragmen, is polynomial-time computable and provides PJR. Another voting rule with this combination of properties is the ODH rule, which has been proposed by Sánchez-Fernández, Fernández, and Fisteus in a recent arXiv preprint (Sánchez-Fernández, Fernández, and Fisteus 2016). Interestingly, both rules are extensions of the D'Hondt seat allocation method (Farrell 2011) to approval-based multi-winner elections.

Moreover, PJR inherits a useful feature of EJR: it characterizes PAV within the class of $\mathbf{w}$-PAV rules.

Proposition 2. *The rule $\mathbf{w}$-PAV satisfies PJR if and only if $\mathbf{w} = (1, \frac{1}{2}, \frac{1}{3}, \dots)$.*

Proof sketch. If $\mathbf{w} = (1, \frac{1}{2}, \frac{1}{3}, \dots)$ then $\mathbf{w}$-PAV satisfies EJR (as shown by Aziz et al. (2015a; 2017)) and hence PJR. The converse statement is proved by reusing the key lemmas from the respective proof for EJR in the work of Aziz et al. (2015a; 2017); see the full version of our paper (Sánchez-Fernández et al. 2016) for details. □

In contrast, all $\mathbf{w}$-RAV rules fail PJR.

Proposition 3. *The rule $\mathbf{w}$-RAV fails PJR for each weight vector $\mathbf{w}$.*

Proof sketch. The proof is similar to the proof that $\mathbf{w}$-RAV fails EJR (Aziz et al. 2015a; 2017) and is relegated to the full version of the paper (Sánchez-Fernández et al. 2016). □

6 Average Satisfaction

A useful measure in the context of justified representation is that of *average satisfaction*: given a ballot profile $(A_1, \dots, A_n)$, a committee W, and a group of voters $N^* \subseteq N$, we define the *average satisfaction* of the voters in N^* as $\frac{1}{|N^*|}\sum_{i \in N^*} |A_i \cap W|$. While it is maybe impossible to ensure that every group of voters has high average satisfaction, it is natural to ask if we can provide some guarantees with respect to this measure to groups of voters that are large and cohesive.

Our first observation is that if a committee W provides JR, we can derive a lower bound on the average satisfaction of such groups.

Proposition 4. *Consider a ballot profile $(A_1, \dots, A_n)$, and suppose that the target committee size k divides n. Let W be a committee of size k that provides JR, and let ℓ be a positive integer. For every ℓ-cohesive group of voters N^* we have*

$$\frac{1}{|N^*|}\sum_{i \in N^*} |A_i \cap W| \geq 1 - \frac{1}{\ell} + \frac{1}{\ell n}.$$

Proof. Let $s = \frac{n}{k}$. Since W provides JR, we can find a voter $i_1 \in N^*$ with $|A_i \cap W| \geq 1$. If $|N^*| > s$ then the set $N^* \setminus \{i_1\}$ satisfies $|N^* \setminus \{i_1\}| \geq s, |\bigcap_{i \in N^* \setminus \{i_1\}} A_i| \geq \ell$, so, applying the JR condition to this set, we can conclude that there is another voter $i_2 \neq i_1$ that approves some candidate in W. By repeating this argument, we conclude that at most $s-1$ voters in N^* approve no candidate in W. Hence,

$$\frac{1}{|N^*|} \sum_{i \in N^*} |A_i \cap W| \geq \frac{|N^*| - s + 1}{|N^*|}$$
$$= 1 - \frac{s-1}{|N^*|} \geq 1 - \frac{s-1}{\ell s} \geq 1 - \frac{1}{\ell} + \frac{1}{\ell n}.$$

□

However, for voting rules that satisfy EJR we can obtain a much stronger guarantee.

Theorem 6. *Consider a ballot profile $(A_1, \ldots, A_n)$, and suppose that the target committee size k divides n. Let W be a committee of size k that provides EJR, and let ℓ be a positive integer. Then for every ℓ-cohesive group of voters $N^* \subseteq N$ it holds that*

$$\frac{1}{|N^*|} \sum_{i \in N^*} |A_i \cap W| \geq \frac{\ell - 1}{2}.$$

Proof. Let $s = \frac{n}{k}$ and let $|N^*| = n^*$. EJR implies that every subset of N^* of size $\ell \cdot s$ contains a voter who approves ℓ candidates in W. Let i_1 be some such voter. If $n^* > \ell \cdot s$ then the set $N^* \setminus \{i_1\}$ satisfies $|N^* \setminus \{i_1\}| \geq \ell \cdot s$, $|\bigcap_{i \in N^* \setminus \{i_1\}} A_i| \geq \ell$, so, applying the EJR condition to this set, we can conclude that there is another voter $i_2 \neq i_1$ that approves ℓ candidates in W. By repeating this argument, we can construct a subset $N_\ell \subseteq N^*$ of size $n^* - \ell \cdot s + 1$ such that each voter in N_ℓ approves at least ℓ candidates in W.

Now, consider the set $N^* \setminus N_\ell$. We have $|N^* \setminus N_\ell| = \ell \cdot s - 1 = (\ell - 1) \cdot s + (s - 1)$, $|\bigcap_{i \in N^* \setminus N_\ell} A_i| \geq \ell$. Thus, by the same argument, the set $N^* \setminus N_\ell$ contains at least s voters each of which approves at least $\ell - 1$ candidates in W; let $N_{\ell-1}$ be a set of exactly s such voters.

Continuing inductively, we partition N^* into $\ell + 1$ pairwise disjoint sets $N_\ell, \ldots, N_1, N_0$ so that we have $|N_\ell| = |N^*| - \ell s + 1$, $|N_{\ell-1}| = \cdots = |N_1| = s$, $|N_0| = s - 1$ and for $0 \leq j \leq \ell$ each voter in N_j approves at least j candidates in W.

The average satisfaction of voters in $N^* \setminus N_\ell$ is at least

$$\frac{1}{|N^* \setminus N_\ell|} \sum_{i \in N^* \setminus N_\ell} |A_i \cap W| \geq \frac{1}{\ell s - 1} \cdot \sum_{j=1}^{\ell-1} |N_j| \cdot j$$
$$\geq \frac{1}{\ell s} \cdot \frac{s(\ell-1)\ell}{2} = \frac{\ell - 1}{2},$$

whereas the average satisfaction of the voters in N_ℓ is at least ℓ. As the average satisfaction of voters in N^* is a convex combination of these two quantities, it is at least $\frac{\ell-1}{2}$. □

In contrast, the worst-case guarantee provided by PJR is not any stronger than the one provided by JR alone.

Example 2. *Consider a ballot profile $(A_1, \ldots, A_n)$ over a candidate set $C = \{c_1, \ldots, c_n, d_1, \ldots, d_n\}$ where $A_i = \{d_1, \ldots, d_n, c_i\}$ for $i \in [n]$. For $k = n$, the committee $\{c_1, \ldots, c_n\}$ provides PJR (and PR), but the average satisfaction of the voters in N (which form an n-cohesive group) is only* 1.

We remark, however, that when the group of voters is "very cohesive" (that is, when all the voters approve exactly the same set of candidates), the average satisfaction that is guaranteed by PJR and EJR is the same, and much higher than what is guaranteed by JR. In particular, consider an election in which there is a group of voters N^*, $|N^*| \geq \ell \cdot \frac{n}{k}$, who all approve precisely the same set of candidates S, $|S| \geq \ell$. Then every committee that provides PJR (or EJR) for this election elects at least ℓ members of S, thereby ensuring that the average satisfaction of voters in N^* is at least ℓ. In contrast, a committee that provides JR may select just one member of S, in which case the average satisfaction of voters in N^* will be just 1.

7 Discussion

We consider JR to be an important axiom in a variety of applications of multi-winner voting. Thus, the result of Section 3 shows that RAV should not be ruled out on these grounds if the target committee size is small. In some applications, such as, e.g., shortlisting candidates for a job, $k \leq 5$ may be a reasonable assumption (Barberà and Coelho 2008; Elkind et al. 2017). We find it surprising that the threshold value of k turns out to be 5 rather than 2 or 3; it would be interesting to see a purely combinatorial proof of this fact.

Our results also highlight a difficulty with the notion of EJR: this axiom is incompatible with perfect representation, which is a very desirable property in parliamentary elections and other settings where fairness is of paramount importance. We therefore propose an alternative to this axiom, Proportional Justified Representation, which is motivated by similar considerations (namely, ensuring that large cohesive groups of voters are allocated several representatives), but does not conflict with PR. PJR also has further attractive properties: it is satisfied by several well-known multi-winner rules (for some of these rules we have to additionally require that k divides n), some of which are efficiently computable, and, just like EJR, it provides a justification for using the harmonic weight vector $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ as the default weight vector for PAV.

However, the results of Section 6 can be viewed as an argument in favor of EJR: every committee that provides EJR guarantees high levels of average satisfaction to members of large cohesive groups, whereas the guarantee offered by committees that provide PJR is, in general, much weaker. Thus, one can think of EJR as a more pragmatic requirement: for every ballot profile a committee that provides EJR (and, as shown by Aziz et al. (2015a; 2017), such a committee is guaranteed to exist) ensures that members of large cohesive groups are happy on average, at the cost of possibly ignoring other agents. In some applications of multi-winner voting such a tradeoff may be acceptable. Consider, for instance, an academic department where

members of different research groups pool their funding to run a departmental seminar. Faculty members have preferences over potential speakers, with members of each research group agreeing on a few candidates from their field. Choosing speakers so as to please the members of large research groups may be a good strategy in this case, even if this means that some members of the department will not be interested in any of the talks. Indeed, if very few talks are of interest to members of a large group, this group may prefer to withdraw its contribution to the funding pool and run its own event series.

The relationship between PR and PJR is more subtle than it may seem at the first sight: while PR implies PJR at the level of committees (Theorem 4), this is not the case at the level of voting rules. Indeed, a voting rule that satisfies PR may behave arbitrarily when the committee size k does not divide the number of voters n, whereas the PJR axiom remains applicable in such scenarios. Of course, a voting rule can satisfy PJR, but not PR: Greedy Monroe, restricted to instances where k divides n, is a case in point. Thus, these two axioms are only loosely related. Indeed, none of the voting rules we consider always satisfies both PJR and PR: the Monroe rule comes closest, but even this rule only satisfies PJR when k divides n (see Example 1). We can, however, construct an artificial voting rule that satisfies both axioms: given $(\mathcal{A}, k)$, this rule checks if there is a committee that provides PR for $(\mathcal{A}, k)$ and, if so, outputs some such committee (which, by Theorem 4, also provides PJR), and otherwise it runs PAV. This voting rule can be seen as an analogue of the Black rule, which is a single-winner rule that outputs a Condorcet winner if one exists and a Borda winner otherwise. Very recently, Brill et al. (2017) identified a voting rule that provides both PR and PJR for all values of n and k, namely, a maximization version of the Phragmén's rule, which they refer to as max-Phragmen.

We conclude the paper by mentioning some open questions that are raised by our work. First, we do not know what is the complexity of checking whether a given committee provides PJR; we note that this problem is polynomial-time solvable for PR (the first part of Theorem 2) and coNP-complete for EJR (Theorem 11 of Aziz et al.). Also, it would be useful to derive bounds on average satisfaction provided by committees that are produced by the voting rules considered in this paper, both theoretically and empirically. In particular, it would be interesting to see whether PAV, which satisfies EJR, performs better in this regard than Monroe or Greedy Monroe, which only satisfy PJR.

Acknowledgments

This research was supported in part by the Spanish Ministerio de Economía y Competitividad (project HERMES-SMARTDRIVER TIN2013-46801-C4-2-R), by the Autonomous Community of Madrid (project e-Madrid S2013/ICE-2715), and by ERC Starting Grant 639945.

References

Anstee, R. P. 1987. A polynomial algorithm for b-matchings: an alternative approach. *Information Processing Letters* 24(3):153–157.

Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; and Walsh, T. 2015a. Justified representation in approval-based committee voting. In *29th AAAI Conference on Artificial Intelligence (AAAI)*, 784–790.

Aziz, H.; Gaspers, S.; Gudmundsson, J.; Mackenzie, S.; Mattei, N.; and Walsh, T. 2015b. Computational aspects of multi-winner approval voting. In *14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 107–115.

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*. To appear.

Barberà, S., and Coelho, D. 2008. How to choose a non-controversial list with k names. *Social Choice and Welfare* 31(1):79–96.

Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research* 47:475–519.

Black, D. 1958. *The theory of committees and elections*. Cambridge University Press.

Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds. 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

Brill, M.; Freeman, R.; Janson, S.; and Lackner, M. 2017. Phragmén's voting methods and justified representation. In *31st AAAI Conference on Artificial Intelligence (AAAI)*.

Conitzer, V. 2010. Making decisions based on the preferences of multiple agents. *Communications of the ACM* 53(3):84–94.

Dummet, M. 1984. *Voting Procedures*. Oxford University Press.

Elkind, E.; Faliszewski, P.; Skowron, P.; and Slinko, A. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare*. To appear.

Elkind, E.; Lang, J.; and Saffidine, A. 2015. Condorcet winning sets. *Social Choice and Welfare* 44(3):493–517.

Farrell, D. M. 2011. *Electoral Systems. A Comparative Introduction*. Palgrave Macmillan, second edition.

Kilgour, D. M. 2010. Approval balloting for multi-winner elections. In Laslier, J.-F., and Sanver, M. R., eds., *Handbook on Approval Voting*. Springer. 105–124.

Monroe, B. 1995. Fully proportional representation. *American Political Science Review* 89(4):925–940.

Procaccia, A. D.; Rosenschein, J. S.; and Zohar, A. 2008. On the complexity of achieving proportional representation. *Social Choice and Welfare* 30(3):353–362.

Sánchez-Fernández, L.; Elkind, E.; Lackner, M.; Fernández, N.; Fisteus, J. A.; Basanta Val, P.; and Skowron, P. 2016. Proportional Justified Representation. *ArXiv e-prints*. arXiv:1611.09928 [cs.GT].

Sánchez-Fernández, L.; Fernández, N.; and Fisteus, J. A. 2016. Fully Open Extensions to the D'Hondt Method. *ArXiv e-prints*. arXiv:1609.05370 [cs.GT].

Skowron, P.; Faliszewski, P.; and Slinko, A. 2015. Achieving fully proportional representation: Approximability results. *Artificial Intelligence* 222:67–103.

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