Approval-Based Shortlisting

Martin Lackner
DBAI, TU Wien
Vienna
lackner@dbai.tuwien.ac.at

Jan Maly
DBAI, TU Wien
Vienna
jmaly@dbai.tuwien.ac.at

ABSTRACT

Shortlisting is the task of reducing a long list of alternatives to a (smaller) set of best or most suitable alternatives from which a final winner will be chosen. Shortlisting is often used in the nomination process of awards or in recommender systems to display featured objects. In this paper, we analyze shortlisting methods that are based on approval data, a common type of preferences. Furthermore, we assume that the size of the shortlist, i.e., the number of best or most suitable alternatives, is not fixed but determined by the shortlisting method. We axiomatically analyze established and new shortlisting methods and complement this analysis with an experimental evaluation based on imperfect quality estimates. Our results lead to recommendations which shortlisting methods to use, depending on the desired properties.

KEYWORDS

Social Choice; Shortlisting; Approval Voting; Multiwinner Voting

ACM Reference Format:
Martin Lackner and Jan Maly. 2021. Approval-Based Shortlisting. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 9 pages.

1 INTRODUCTION

Shortlisting is a task that arises in many scenarios and applications: given a large set of alternatives, identify a smaller subset that consists of the best or most suitable alternatives. Prototypical examples of shortlisting are awards, where we often find a two-stage process. In a first shortlisting step, the large number of contestants (books, films, individuals, etc.) is reduced to a smaller number. In a second step, the remaining contestants can be evaluated more closely and one contestant in the smaller set is chosen to receive the award. Both steps may involve a form of group decision making (voting), but can also consist of a one-person or even automatic decision. For example, the shortlist of the Booker Prize is selected by a small jury [33], whereas the shortlists of the Hugo Awards are compiled based on thousands of ballots [32]. Another very common application of shortlisting is the selection of a number of most promising applicants for a position who will be invited for an interview [4, 31]. Apart from these prototypical examples, shortlisting is also useful in many less obvious applications like the aggregation of expert opinions for example in the medical domain [19] or in risk management and assessment [34]. Shortlisting can even be used in scenarios without agents in the traditional sense, for example if we consider features as voters to perform an initial screening of objects, i.e., a feature approves all objects that exhibit this feature [17].

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

In this paper, we consider shortlisting as a form of collective decision making. We assume that a group of voters announce their preferences by specifying which alternatives they individually view worthy of being shortlisted, i.e., they file approval ballots. In practice, approval ballots are commonly used for shortlisting, because the high number of alternatives that necessitates shortlisting in the first place precludes the use of ranked ballots. Furthermore, we assume that the number of alternatives to be shortlisted is not fixed (but there might be a preferred number), as there are very few shortlisting scenarios where there is a strong justification for an exact size of the shortlist. Due to this assumption, we are not in the classical setting of multiwinner voting [16, 22, 24], where a fixed-size committee is selected but in the more general setting of multiwinner voting with a variable number of winners [17, 20, 21]. One can also view shortlisting rules as a particular type of *social dichotomy functions* [7, 11], i.e., voting rules which partition alternatives into two groups.

In real-world shortlisting tasks, there are two prevalent methods in use: *Multiwinner Approval Voting* (selecting the k alternatives with the highest approval score) and threshold rules (selecting all alternatives approved by more than a fixed percentage of voters). Further shortlisting methods have been proposed in the literature [6, 17, 21]. Despite the prevalence of shortlisting applications, there does not exist work on systematically choosing a suitable shortlisting method. Such a recommendation would have to consider both expected (average-case) behavior and guaranteed axiomatic properties, neither have been studied previously specifically for shortlisting applications (cf. related work below). Our goal is to answer this need and provide principled recommendations for shortlisting rules, depending on the properties that are desirable in the specific shortlisting process.

In more detail, the contributions of this paper are the following:

- We define shortlisting as a voting scenario and specify minimal requirements for shortlisting methods (Section 2). Furthermore, we introduce three new shortlisting methods: *First k-Gap*, *Largest Gap*, and *Size Priority* (Section 3).
- We conduct an axiomatic analysis of seven shortlisting methods and by that identify essential differences between them. Furthermore, we axiomatically characterize *Approval Voting*, *f-Threshold*, and the new *First k-Gap* rule (Section 4).
- We present a connection between shortlisting and clustering algorithms, as used in machine learning. We show that *First k-Gap* and *Largest Gap* can be viewed as instantiations of linkage-based clustering algorithms (Section 5).
- In numerical simulations, we approach an essential difficulty of shortlisting processes: voters with imperfect (noisy) perception of the alternatives. These simulations complement our axiomatic

analysis by highlighting further properties of shortlisting methods and provide further data points for recommending shortlisting methods (Section 6).

- The recommendations based on our findings are summarized in Section 7. In brief, our analysis leads to a recommendation of *First k-Gap*, *f-Threshold*, and *Size Priority*, depending on the general shortlisting goal and desired behavior.

Related work. There are two recent papers that are particularly relevant for our work. Both in [17] and [28], the authors investigate multiwinner voting with a variable number of winners. In contrast to our paper, the main focus of [17] lies on computational complexity, which is less of a concern for our shortlisting setting (as discussed later). The paper also contains some numerical simulations related to the number of winners (which is one of three metrics we consider in our paper). In the few cases where shortlisting rules are considered[1], their simulation concerning winner set sizes agrees with ours (cf. Fig. 1c, $\lambda = 0$). In [28], the authors study proportionality. Proportionality is incompatible with Efficiency, which we require for shortlisting rules. Thus, the rules and properties considered in [28] do not intersect with ours.

More generally, there is a substantial literature on multiwinner voting with a *fixed* number of winners (i.e., committee size), as witnessed by recent surveys [16, 22, 24]. Multiwinner voting rules are much better understood, both from an axiomatic [2, 14, 18, 25, 29] and experimental [8, 13] point of view, also in the context of shortlisting [3, 9]. Results for multiwinner rules, however, typically do not translate to the setting with a variable number of winners.

2 THE FORMAL MODEL

In this section we describe our formal model that embeds score-based shortlisting in a voting framework. The model consists of two parts: a general framework for approval-based elections with a variable number of winners [17, 20, 21] on the one hand and, on the other hand, four basic axioms that we consider essential prerequisites for shortlisting rules.

An approval-based election $E = (C, V)$ consists of a non-empty set of alternatives $C = \{c_1, \dots, c_m\}$ and an n-tuple of approval ballots $V = (v_1, \dots, v_n)$ where $v_i \subseteq C$ and $c_j \in v_i$ if voter i approves alternative c_j and $c_j \notin v_i$ otherwise. In the following we will always write n_E for the number of voters and m_E for the number alternatives in an election E. If no ambiguity arises, we will omit the subscript. The *approval score* $sc_E(c_j)$ of alternative c_j in election E is the number of approvals of c_j in V, i.e., $sc_E(c_j) = |\{i : 1 \leq i \leq n \text{ and } c_j \in v_i\}|$. We write $sc(E)$ for the vector $(sc_E(c_1), \dots, sc_E(c_m))$. To avoid unnecessary case distinctions, we only consider *non-degenerate* elections: these are elections where not all alternatives have the same approval score. An *approval-based variable multiwinner rule* (which we refer to just as "voting rule") is a function mapping an election $E = (C, V)$ to a subset of C. Given a rule $\mathcal{R}$ and an election E, $\mathcal{R}(E) \subseteq C$ is the *winner set* according to voting rule $\mathcal{R}$, i.e., $\mathcal{R}(E)$ is the set of alternatives which have been shortlisted. Note that $\mathcal{R}(E)$ may be empty or contain all candidates.

[1] *NAV* and *NCAS* in [17] are equivalent to *f-Threshold* and *Approval Voting* in our paper (subject to tiebreaking). Also *First Majority* is considered in [17].

Now we introduce the basic axioms that we require every shortlisting rule to satisfy. Anonymity and Neutrality are two basic fairness axioms that are considered essential for voting rules [35].

Axiom 1 (Anonymity)**.** All voters are treated equal, i.e., for every permutation $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ and election $E = (C, V)$, if $E^* = (C, V^*)$ with $V^* = (v_{\pi(1)}, \dots, v_{\pi(n)})$, then $\mathcal{R}(E) = \mathcal{R}(E^*)$.

Axiom 2 (Neutrality)**.** All alternatives are treated equally, i.e., for every election $E = (C, V)$ and permutation $\pi : C \to C$, if $E^* = (C, V^*)$ where $V^* = (v_1^*, \dots, v_n^*)$ with $v_i^* = \{\pi(c) \mid c \in v_i\}$, then $\pi(c) \in \mathcal{R}(E^*)$ iff $c \in \mathcal{R}(E)$ for all $c \in C$.

Shortlisting differs from other multiwinner scenarios in that we are not interested in representative or proportional committees. Instead, the goal is to select the most excellent alternatives. This goal is formalized in the following axiom.

Axiom 3 (Efficiency)**.** No winner can be (strictly) less approved than a non-winner, i.e., for all elections $E = (C, V)$ and all candidates c_i and c_j if $sc_E(c_i) > sc_E(c_j)$ and $c_j \in \mathcal{R}(E)$ then also $c_i \in \mathcal{R}(E)$.

The assumption that approval scores are approximate measures of the general quality of alternatives can also be argued in a probabilistic framework: under reasonable assumptions a set of alternatives with the highest approval scores coincides with the maximum likelihood estimate of the truly best alternatives [26]. Thus, we impose Efficiency to guarantee the inclusion of the most-likely best alternatives.

Since the number of winners is variable in our setting, there is generally no need to break ties. Because tiebreaking is usually an arbitrary and unfair process, voting rules should not introduce unnecessary tiebreaking.

Axiom 4 (Non-tiebreaking)**.** If two alternatives have the same approval score, either both or neither should be winners i.e., for all elections $E = (C, V)$ and all candidates c_i and c_j if $sc_E(c_i) = sc_E(c_j)$ then either $c_i, c_j \in \mathcal{R}(E)$ or $c_i, c_j \notin \mathcal{R}(E)$.

We set these four axioms as the minimal requirements for a voting rule to be considered a shortlisting rule in our sense.

Definition 1. An approval-based variable multiwinner rule is a shortlisting rule if it satisfies Anonymity, Neutrality, Efficiency and is non-tiebreaking.

Observe that Non-tiebreaking and Efficiency are axioms that are only interesting if we consider voting with a variable number of winners. Clearly, no voting rule for voting with a fixed number of winners can be Non-tiebreaking. Furthermore, except for the issue of how to break ties, there is exactly one voting rule for approval voting with a fixed number k of winners that satisfies Efficiency, namely picking the k alternatives with maximum approval score (*Multiwinner Approval Voting*).

A consequence of Efficiency and Non-tiebreaking is that a shortlisting rule only has to decide how many winners there should be. This reduces the complexity of finding the winner set drastically as there are only linearly many possible winner sets, in contrast to the exponentially many subsets of C.

Observation 1. For every election there are at most $m + 1$ sets that can be winner sets under a shortlisting rule.

3 SHORTLISTING RULES

In the following, we define the shortlisting rules that we study in this paper. We define these rules by specifying which properties an alternative has to satisfy to be contained in the winner sets. As before, let $E = (C, V)$ be an election. We assume additionally that $c_1, \dots, c_m$ is an enumeration of the alternatives in descending order of approval score, i.e., such that $sc_E(c_{i-1}) \geq sc_E(c_i)$ for all $2 \leq i \leq m$. Some of the following rules are parameterized by a positive integer k. An example illustrating the rules follows at the end of the section.

3.1 Established Rules

A natural idea is to select all most-approved alternatives. The corresponding winner set equals the set of co-winners of classical Approval Voting [5].

Approval Voting. An alternative c is a winner iff c's approval score is maximal, i.e., $c \in \mathcal{R}(E)$ iff $sc_E(c) = \max(sc(E))$.

Another natural way to determine the winner set is to fix some percentage threshold and declaring all alternatives to be winners that surpass this approval threshold [20]. For example, for a baseball player to be entered into the Hall of Fame, more than 75% of the members of the Baseball Writers' Association of America have to approve this nomination [10]. Such rules are known as quota rules in judgment aggregation [15].

f-Threshold. Let $f : \mathbb{N} \to \mathbb{N}$ be a function such that $0 < f(|V|) < |V|$. Then, $c \in \mathcal{R}(E)$ for an alternative $c \in C$ if and only if $sc_E(c) > f(|V|)$. We write α-Threshold for a constant $0 < \alpha < 1$ to denote the f-Threshold rule with $f(n) = \lfloor \alpha \cdot n \rfloor$.

The next two rules are further shortlisting methods that have been proposed in the literature. *First Majority* [21] includes as many alternatives as necessary to comprise more than half of all approvals. The following definition deviates slightly from the original definition in that it is non-tiebreaking.

First Majority. Let i be the smallest index such that $\sum_{j \leq i} sc_E(c_j) > \sum_{j>i} sc_E(c_j)$. Then $c \in \mathcal{R}(E)$ if and only if $sc_E(c) \geq sc_E(c_i)$.

Next-k [6] is a rule that includes alternatives starting with the highest approval score, until a major drop in the approval scores is encountered, more precisely, if the total approval score of the next k alternatives is less than the score of the previous alternative.

Next-k. We have $c_i \in \mathcal{R}(E)$ if for all $i' < i$ it holds that $sc_E(c_{i'}) \leq \sum_{j=1}^{k} sc_E(c_{i'+j})$, where $sc_E(c_{i'+j}) = 0$ if $i' + j > m$.

Observe that for both *Next-k* and *First Majority* the winner set does not depend on the chosen enumeration of alternatives. This will also hold for the new voting rules introduced in the following.

3.2 New Shortlisting Rules

Similarly to *Next-k*, the next two rules are based on the idea that one wants to make the cut between winners and non-winners in a place where there is a large gap in the approval scores. This can either be the overall largest gap or the first sufficiently large gap.

Largest Gap. Let i be the smallest index such that $sc_E(c_i) - sc_E(c_{i+1}) = \max_{j<m}(sc_E(c_j) - sc_E(c_{j+1}))$. Then $c \in \mathcal{R}(E)$ if and only if $sc_E(c) \geq sc_E(c_i)$.

Note that in this definition a smallest index is guaranteed to exist due to our assumption that profiles are non-degenerate.

First k-Gap. Let i be the smallest index such that $sc_E(c_i) - sc_E(c_{i+1}) \geq k$. Then $c \in \mathcal{R}(E)$ if and only if $sc_E(c) \geq sc_E(c_i)$. If no such index exists, then $\mathcal{R}(E) = C$.

The parameter k has to capture what it means in a given shortlisting scenario that there is a sufficiently large gap between alternatives, which in particular depends on $|V|$. If no further information is available, one can choose k by a simple probabilistic argument. Assume, for example, alternative c's approval score is binomially distributed $sc_E(c) \sim B(n, q_c)$, where n is the number of voters and q_c can be seen as c's quality. We choose k such that the probability of events of the following type are smaller than a selected threshold α: two alternatives a and b have the same objective quality ($q_a = q_b$) but have a difference in their approval scores of k or more. In such a case, the *First k-Gap* rule might choose one alternative and not the other even though they are equally qualified, which is an undesirable outcome. For example, if $n = 100$ and we want $\alpha = 0.5$, we have to choose $k \geq 5$ and if we want $\alpha = 0.1$ we need $k \geq 12$. Note that this argument leads to rather large k-values; if further assumptions about the distribution of voters can be made, smaller k-values are feasible.

The voting rules above output winner sets of very different sizes (as we will see in the experimental evaluation, Section 6). It is a common case, however, that there is a preferred size for the winner set, but this size can be varied in order to avoid tiebreaking. This flexibility is especially crucial if the electorate is small and ties are more frequent. Based on real-world shortlisting processes, we propose a rule that deals with this scenario by accepting a preference order over set sizes as parameter and selecting a winner set with the most preferred size that does not require tiebreaking.

Size Priority. Let $\rhd$ be a strict total order on $\{0, \dots, m\}$, the *priority order*. Then $\mathcal{R}(E) = \{c_i \in C \mid 1 \leq i \leq k\}$ if and only if

- either $sc_E(c_k) \neq sc_E(c_{k+1})$ or $k = 0$ or $k = m$,
- and $sc_E(c_\ell) = sc_E(c_{\ell+1})$ for all $\ell \rhd k$.

Size Priority is a non-tiebreaking analogue of *Multiwinner Approval Voting*, which selects the k alternatives with the highest approval score. A specific instance of *Size Priority* is used by the Hugo Award with the priority order $5 \rhd 6 \rhd 7 \dots$ [32]. Generally, the choice of a priority order depends on the situation at hand. For award-shortlisting, typically a small number of alternatives is selected (the Booker Prize, e.g., has a shortlist of size 6). In a much more principled fashion, Amegashie [1] argues that the optimal size of the winner set for shortlisting should be proportional to $\sqrt{m}$, i.e., the square root of the number of alternatives.

In practice, the most common priority order is $k \rhd k+1 \rhd \dots \rhd m$ for some $k < m$, i.e., the smallest non-tiebreaking committee that contains at least k alternatives is selected. Another important special case are instances of *Size Priority* that rank 0 and m the lowest, i.e., that are decisive whenever possible. Therefore, we give *Size Priority* rules with based on such priority orders a special name.

Definition 2. Let $\rhd$ be a strict total order on $0, \dots, m$ and let k be a positive integer with $k \leq m$ such that $k \rhd k+1 \rhd \dots \rhd m$

and $m \rhd \ell$ for all $\ell < k$. Then, the *Size Priority* rule defined by the priority order $\rhd$ is an *Increasing Size Priority* rule.

Let $\rhd$ be a strict total order on $0, \ldots, m$ such that $k \rhd m$ and $k \rhd 0$ holds for all $0 < k < m$. Then, the *Size Priority* rule defined by the priority order $\rhd$ is an *Decisive Size Priority* rule.

Other special cases of *Size Priority* could be defined in a similar way, for example *Decreasing Size Priority*. However, *Increasing Size Priority* and *Decisive Size Priority* are the most natural and common types of *Size Priority* and additionally satisfies better axiomatic properties than, e.g., *Decreasing Size Priority*.

Example. Let $E = (C, V)$ be an election with 10 voters and 8 alternatives $c_1, \ldots, c_8$. Furthermore, let $sc(E) = (10, 10, 9, 8, 6, 3, 3, 0)$. Then the set of *Approval Voting* winners is $\{c_1, c_2\}$ and the set of *0.5-Threshold* winners is $\{c_1, \ldots, c_5\}$. The set of *First Majority* winners is $\{c_1, c_2, c_3\}$, since $\sum_{c \in C} sc_E(c) = 49$ and $sc_E(c_1) + sc_E(c_2) + sc_E(c_3) = 29$. For every $i \leq 7$ we have $sc_E(c_{i-1}) \leq sc_E(c_i) + sc_E(c_{i+1})$. Therefore, the set of *Next-k* winners is $\{c_1, \ldots, c_7\}$ for every $k \geq 2$. There are two 3-gaps, between c_5 and c_6 and between c_7 and c_8 and there are no larger gaps, hence $\{c_1, \ldots, c_5\}$ is the set of winners under *Largest Gap*. The first 2-gap is between c_4 and c_5, hence the winner set according to *First* 2*-Gap* is $\{c_1, \ldots, c_4\}$. Now let $\rhd$ be a strict total order such that the top elements are $1 \rhd 6 \rhd 0 \rhd \ldots$. Then the set of *Size Priority* winners under $\rhd$ is the empty set, because $\{c_1\}$ and $\{c_1, \ldots, c_6\}$ break ties, as $sc_E(c_1) = sc_E(c_2)$ and $sc_E(c_6) = sc_E(c_7)$.

We note that, due to Observation 1, all of the above rules can be computed in polynomial time. Finally, we observe that *Approval Voting* is a special case of *First k-Gap*, *Next-k* and *Increasing Size Priority*. This is because *First k-Gap* and *Next-k* equal *Approval Voting* if we set $k = 1$ and *Increasing Size Priority* equals *Approval Voting* with priority order $1 \rhd 2 \rhd \cdots \rhd m \rhd 0$.

4 AXIOMATIC ANALYSIS

In this section, we axiomatically analyze shortlisting rules with the goal to discern their defining properties. First, we consider axioms that are motivated by the specific requirements of shortlisting, then we study well-known axioms that describe more generally desirable properties of voting rules. For an overview, see Table 1.

When shortlisting is used for the initial screening of a set of alternatives, for example for an award or a job interview, then we do not assume that the voters have perfect judgment. Otherwise, there would be no need for a second round of deliberation, as we could just choose the highest-scoring alternative as a winner. Therefore, small differences in approval may not correctly reflect which alternative is more deserving of a spot on the shortlist. Thus, out of fairness, we want our voting rule to treat alternatives differently only if there is a significant difference in approval between them.

Axiom 5 (ℓ-Stability)**.** If the approval scores of two alternatives differ by less than ℓ, either both or neither should be a winner, i.e., for every election $E = (C, V)$ and candidates c_i and c_j if $|sc_E(c_i) - sc_E(c_j)| < \ell$ then either $c_i, c_j \in \mathcal{R}(E)$ or $c_i, c_j \notin \mathcal{R}(E)$.

Here, the parameter ℓ has to capture what a significant difference is in a given election. This will depend, for example, on the number and trustworthiness of the voters.

Observe that 1-Stability equals non-tiebreaking. Furthermore, as the approval scores approximate the underlying quality of alternatives[2], at least we want to include alternatives that are approved by everyone and exclude alternatives that are approved by no one.

Axiom 6 (Unanimity)**.** If an alternative is approved by everyone, it must be a winner, i.e., for every election $E = (C, V)$, if $sc_E(c) = n$ then $c \in \mathcal{R}(E)$.

Axiom 7 (Anti-Unanimity)**.** If an alternative is approved by no one, it cannot win, i.e., for every election $E = (C, V)$ if $sc_E(c) = 0$ then $c \notin \mathcal{R}(E)$.

Unfortunately, it turns out that these three axioms are incompatible unless there are many more voters than alternatives. Indeed Unanimity, Anti-Unanimity and ℓ-Stability can be jointly satisfied if and only if $|V| \geq l \cdot |C| + 1$.

Theorem 1. *For every ℓ there is a rule that satisfies Unanimity, Anti-Unanimity and ℓ-Stability for every election E such that $n_E > (\ell - 1) \cdot (m_E - 1)$. This is a tight bound in the following sense: If $\ell > 1$, no voting rule satisfies Unanimity, Anti-Unanimity and ℓ-Stability for all elections E with $n_E \leq (\ell - 1) \cdot (m_E - 1)$.*

Proof. For the one direction, we claim that a slightly modified version of *First k-Gap* satisfies all three axioms for elections E with $n_E > (\ell - 1) \cdot (m_E - 1)$. We define *Modified First ℓ-Gap* as follows: Let $c_1, \ldots, c_m$ be an enumeration of C such that $sc_E(c_{i-1}) \geq sc_E(c_i)$. Let i be the smallest index such that $sc_E(c_i) - sc_E(c_{i+1}) \geq \ell$. Then $c \in \mathcal{R}(E)$ if and only if $sc_E(c) \geq sc_E(c_i)$. If no such index exists, then $\mathcal{R}(E) = \emptyset$ if there is an alternative c with $sc_E(c) = 0$, and $\mathcal{R}(E) = C$ otherwise. Clearly, this rule still satisfies ℓ-Stability.

Now, let E be an election such that there is an alternative c with $sc_E(c) = n$. Assume first that there is no alternative c' with $sc_E(c') = 0$. In that case, *Modified First ℓ-Gap* vacuously satisfies Anti-Unanimity and, by definition, also Unanimity. Now assume that there is an alternative c with $sc_E(c) = 0$. We claim that there is an index i such that $sc_E(c_i) - sc_E(c_{i+1}) \geq \ell$ and hence only alternatives c such that $sc_E(c) \geq sc_E(c_i) > \ell - 1$ are winners. Otherwise, we have $sc_E(c_{i+1}) \geq sc_E(c_i) - (\ell - 1)$ for all $i < m$ and hence $sc_E(c_m) \geq sc_E(c_1) - (\ell - 1) \cdot (m - 1)$. However, as $sc_E(c_1) = n > (\ell - 1) \cdot (m - 1)$ this contradicts the assumption that there is an alternative c with $sc_E(c) = 0$, i.e., $sc_E(c_m) = 0$.

Finally, let E be an election such that there is no alternative c with $sc_E(c) = n$. Then, *Modified First ℓ-Gap* vacuously satisfies Unanimity. Now, if there is an alternative c' with $sc_E(c') = 0$ then we have to distinguish two cases. If there is no ℓ-gap, then $\mathcal{R}(E) = \emptyset$ by definition and hence *Modified First ℓ-Gap* satisfies Anti-Unanimity. On the other hand, if there is a ℓ-gap, then only alternatives above the ℓ-gap are selected, which must have a score of ℓ or larger. Hence, Anti-Unanimity is also satisfied.

Now we assume towards a contradiction that there is a rule $\mathcal{R}$ that satisfies Unanimity, Anti-Unanimity and ℓ-Stability ($\ell > 1$) for all elections $E = (C, V)$ with $n_E = (\ell - 1) \cdot (m_E - 1)$. Let E be an election with 2 alternatives and $\ell - 1$ voters such that $sc(E) = (\ell - 1, 0)$. We observe $n_E = \ell - 1 \geq (\ell - 1) \cdot (2 - 1)$. Therefore, $\mathcal{R}$

[2]The relation between approval voting and maximum likelihood estimation is analyzed in detail by Procaccia and Shah [26], in particular, under which conditions approval voting selects the most likely "best" alternatives.

	Unanimity	Anti-Unanimity	Independence	Ind. of Losing Alt.	ℓ-Stability	Determined	Set Monot.	Superset Monot.
Approval Voting	✓	✓	×	✓	×	✓	✓	✓
f-Threshold	✓	✓	✓	✓	×	×	✓	×
First Majority	✓	✓	×	×	×	✓	×	×
Next-k	✓	✓	×	×	×	✓	✓	×
Largest Gap	✓	✓	×	×	×	✓	✓	×
First k-Gap	✓	×	×	✓	$\ell \leq k$	✓	✓	✓
Decis. Size Priority	✓	✓	×	×	×	✓	✓	×
Incr. Size Priority	✓	×	×	✓	×	✓	✓	✓

Table 1: Results of the axiomatic analysis.

must satisfy Unanimity, Anti-Unanimity and ℓ-Stability on E. Hence, $c_1 \in \mathcal{R}(E)$ must hold by Unanimity. Then $sc_E(c_1) - sc_E(c_2) < \ell$ implies $c_2 \in \mathcal{R}(E)$ by ℓ-Stability, contradicting Anti-Unanimity. □

First k-Gap satisfies Unanimity and ℓ-Stability for $k \geq \ell$ for all elections. Therefore, it cannot satisfy Anti-Unanimity. Furthermore, we observe that *First k-Gap* is the only voting rule considered in this paper that satisfies ℓ-Stability for $\ell > 1$, as *Approval Voting*, *f-Threshold*, *Next-k*, *First Majority* and *Largest Gap* satisfy Unanimity and Anti-Unanimity for all non-degenerate profiles. *Size Priority* always satisfies either Unanimity or Anti-Unanimity. In particular it satisfies Unanimity if $m \rhd 0$ holds and Anti-Unanimity if $0 \rhd m$ holds. It satisfies both axioms (for non-degenerate profiles) if and only if it is decisive. Therefore *Increasing Size Priority* satisfies Unanimity but not Anti-Unanimity. Finally, *Size Priority* satisfies ℓ-Stability for $\ell > 1$ if and only if 0 or m is the most preferred size. It is worth noting, however, that *Largest Gap* satisfies ℓ-Stability whenever there is an ℓ-gap.

Another requirement for a shortlisting rule is that it produces short shortlists. To find voting rules that produce small sets of winners without compromising on quality, we define the concept of a minimal voting rule that satisfies a set of axioms.

Definition 3. Let $\mathcal{A}$ be a set of axioms and let $S(\mathcal{A})$ be the set of all voting rules satisfying all axioms in $\mathcal{A}$. Then, we say a voting rule is a minimal voting rule $\mathcal{R}$ for $\mathcal{A}$ if for all elections E it holds that $\mathcal{R}(E) = \bigcap_{\mathcal{R}^* \in S(\mathcal{A})} \mathcal{R}^*(E)$.

In general that a minimal voting rule $\mathcal{R}$ for a set of axioms $\mathcal{A}$ satisfies all axioms in $\mathcal{A}$. Consider, e.g., the following axiom:

Axiom 8 (Determined)**.** Every election must have at least one winner, i.e., for all elections E we have $\mathcal{R}(E) \neq \emptyset$.

Besides *f-Threshold* and *Size Priority* all voting rules considered in this paper are determined. *Size Priority* is determined if and only if it is either decisive or $m \rhd 0$. We observe the minimal determined voting rule always returns the empty set and is hence not determined. However, the following holds:

Proposition 2. *Let $\mathcal{A}$ be a set of axioms that contains the four basic shortlisting axioms (Axioms 1–4). Then the minimal voting rule for $\mathcal{A}$ is again a shortlisting rule, i.e., it satisfies Axioms 1–4.*

Proof. We show that the minimal voting rule $\mathcal{R}$ satisfies efficiency and is non-tiebreaking. Let E be an election. As every rule in $S(\mathcal{A})$ is a shortlisting rule, there is a $k_{\mathcal{R}^*} \in \{0, \ldots, m\}$ for every rule $\mathcal{R}^* \in S(\mathcal{A})$ such that $\mathcal{R}^*(E) = \{c_1, \ldots, c_{k_{\mathcal{R}^*}}\}$. Now let k_m be the smallest k such that there is a rule $\mathcal{R}^* \in S(\mathcal{A})$ with $\mathcal{R}^*(E) = \{c_1, \ldots, c_k\}$. Then, by definition $\mathcal{R}(E) = \mathcal{R}^*(E)$. As $\mathcal{R}^*(E)$ does not violate efficiency and non-tiebreaking for E, neither does $\mathcal{R}$. As this argument holds for arbitrary elections, $\mathcal{R}$ satisfies efficiency and is non-tiebreaking. It is easy to see that neutrality and anonymity hold as well. □

The minimal shortlisting rule (without additional axioms) is the voting rule that always outputs the empty set, *Approval Voting* is the minimal determined shortlisting rule and the minimal shortlisting rule that is k-stable is *First k-Gap*.

Theorem 3. *Approval Voting is the minimal voting rule that is efficient, non-tiebreaking and determined. Furthermore, for every positive integer k, First k-Gap is the minimal voting rule that is efficient, k-stable and determined.*

Proof. Let $\mathcal{A}$ be the set $\{\text{Efficiency}, k - \text{Stability}, \text{Determined}\}$ and $\mathcal{R}$ be *First k-Gap*. We know that *First k-Gap* is efficient, k-stable and determined, therefore we know $\bigcap_{\mathcal{R}^* \in S(\mathcal{A})} \mathcal{R}^*(E) \subseteq \mathcal{R}(E)$.

Now, every determined voting rule must have a non-empty set of winners. If the voting rule is efficient, the set of winners must contain at least one top ranked alternative. Now, consider an enumeration of the alternatives $c_1, \ldots, c_m$ such that $sc_E(c_j) \geq sc_E(c_{j+1})$ holds for all j. If a voting rule is k-stable, a winner set containing one top ranked alternative must contain all alternatives c_i for which $sc_E(c_j) < sc_E(c_{j+1}) + k$ holds for all $j < i$. By the definition of *First k-Gap* this implies $\mathcal{R}(E) \subseteq \bigcap_{\mathcal{R}^* \in S(\mathcal{A})} \mathcal{R}^*(E)$.

The minimality of *Approval Voting* is a special case of the minimality of *First k-Gap*, as 1-Stability equals non-tiebreaking. □

This result is another strong indication that *First k-Gap* is promising from an axiomatic standpoint. It produces shortlists that are as short as possible without violating k-Stability, an axiom that is very desirable in many shortlisting scenarios.

ℓ-Stability formalizes the idea that the winner determination should take the magnitude of difference between approval scores into account. This contradicts an idea that is often considered in judgment aggregation, namely that all alternatives should be treated independently [15].

Axiom 9 (Independence)**.** If an alternative is approved by exactly the same voters in two elections then it must be a winner either in both or in neither. That is, for an alternative c, and two elections $E = (C, V)$ and $E^* = (C, V^*)$ with $|V| = |V^*|$ and $c \in v_i$ if and only if $c \in v_i^*$ for all $i \leq n$, it holds that $c \in \mathcal{R}(E)$ if and only if $c \in \mathcal{R}(E^*)$.

f-*Threshold* rules are the only rules in our paper satisfying Independence. Indeed, Independence characterizes f-*Threshold* rules.

THEOREM 4. *Given a fixed set of alternatives C, every shortlisting rule that satisfies Independence is an f-Threshold rule for some function f.*

PROOF. Let $\mathcal{R}$ be a voting rule that satisfies Anonymity and Independence. Then we claim that for two elections $E = (C, V)$ and $E^* = (C, V^*)$ with $|V| = |V^*|$ and an alternative $c_i \in C$ we have that $sc_E(c_i) = sc_{E^*}(c_i)$ implies that either $c_i \in \mathcal{R}(E), \mathcal{R}(E^*)$ or $c_i \notin \mathcal{R}(E), \mathcal{R}(E^*)$. If $sc_E(c_i) = sc_{E^*}(c_i)$, then there is a permutation $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ such that $c_i \in v_i$ if and only if $c_i \in v^*_{\pi(i)}$. Now, let $E' = (C, \pi(V))$. Then, by Anonymity, $c_i \in \mathcal{R}(E)$ if and only if $c_i \in \mathcal{R}(E')$. Now, as c_i is approved by the same voters in E' and E^*, Independence implies $c_i \in \mathcal{R}(E')$ if and only if $c_i \in \mathcal{R}(E^*)$.

Now let $\mathcal{R}$ additionally satisfy Efficiency. Let $E = (C, V)$ and $E^* = (C, V^*)$ be two elections with $|V| = |V^*|$. Furthermore, assume $c_i \in \mathcal{R}(E)$ and $sc_E(c_i) < sc_{E^*}(c_i)$. We claim that this implies $c_i \in \mathcal{R}(E^*)$. By Independence, we can assume w.l.o.g. that there is an alternative c_j such that $sc_E(c_j) = sc_{E^*}(c_i)$. Then, by efficiency, $c_j \in \mathcal{R}(E^*)$. Now, let E' be the same election as E but with c_i and c_j switched. Then by Neutrality we have $c_i \in \mathcal{R}(E')$. As by definition $sc_{E'}(c_i) = sc_{E^*}(c_i)$ this implies $c_i \in \mathcal{R}(E^*)$ by Anonymity and Independence. This means that for every alternative c_i and $n \in \mathbb{N}$ there is a k such that for all elections $E = (C, V)$ with $|V| = n$ we know $c_i \in \mathcal{R}(E)$ if and only if $sc_E(c_i) \geq k$. If $\mathcal{R}$ also satisfies Neutrality, then k must be the same for every $c_i \in C$ and hence $\mathcal{R}$ must be a Threshold rule. □

A sensible modification of f-*Threshold* would be to select all alternatives with an above-average approval score, i.e., the set of winners consists of all alternatives c with $sc_E(c) > \frac{1}{m} \cdot \sum_{c' \in C} sc_E(c')$. Duddy et al. [12] analyzed this rule and concluded that it is the best rule for partitioning alternatives into homogeneous groups (see also the axiomatic characterization of this rule in [7]). This rule is not a f-*Threshold* rule (by definition).

Let us now consider classic axioms of social choice theory, adapted to the shortlisting setting. The first one states that removing a non-winning alternative cannot change the outcome of an election.

Axiom 10. (Independence of Losing Alternatives) Let $E = (C, V)$ with $V = (v_1, \dots, v_n)$ and $E^* = (C^*, V^*)$ where $C^* = C \setminus \{c^*\}$ and $V^* = (v^*_1, \dots, v^*_n)$ be two elections such that $c^* \notin \mathcal{R}(E)$ and $v^*_i = v_i \setminus \{c^*\}$ for all $i \leq n$. Then $\mathcal{R}(E) = \mathcal{R}(E^*)$.

Clearly, f-*Threshold* satisfies this axiom as it also satisfies Independence. Furthermore, as the removal of a losing alternative can only widen the gap between the winners and the non-winners, *First k-Gap* satisfies the axiom, and so does *Approval Voting*, which is a special case of *First k-Gap*. None of the other rules satisfy Independence of Losing Alternatives. *First Majority* does not satisfy the axiom: Assume E is an election such that $sc(E) = (3, 2, 1, 0)$. Then the winning set under *First Majority* is $\{c_1, c_2\}$ but removing c_3 changes the winning set to $\{c_1\}$. For the same election, the winner set under *Largest Gap* is $\{c_1\}$ but removing c_3 changes this to $\{c_1, c_2\}$. For *Next-k*, consider an election E with $sc(E) = (4, 3, 2, 0)$. Then, for every $k > 1$, we have $\mathcal{R}(E) = \{c_1, c_2\}$ under *Next-k*, but after deleting c_3 we have $\mathcal{R}(E) = \{c_1\}$.

For *Size Priority* we encounter a difficulty: Independence of Losing Alternatives cannot be applied to *Size Priority* because each instance of *Size Priority* is defined by a linear order on $0, \dots, m$ and decreasing the number of alternatives necessitates a different order. However, we can say that a linear order $\rhd$ on $\mathbb{N}$ defines a class of *Size Priority* instances as follows: for every number of alternatives m, we define a *Size Priority* instance by restriction $\rhd$ to $\{0, 1, \dots, m\}$. This allows us to precisely say what it means that a class of *Size Priority* instances (defined by $\rhd$) satisfies Independence of Losing Alternatives. Consider, e.g., the class of *Size Priority* instances defined by any order of the form $2 \rhd 1 \rhd \dots$ and an election E with $sc(E) = (2, 1, 1)$. Then $\mathcal{R}(E) = \{c_1\}$ but the removal of c_3 leads to $\mathcal{R}(E) = \{c_1, c_2\}$. Thus, *Size Priority* fails Independence of Losing Alternatives in general. However, removing a losing alternative cannot change the outcome of a *Increasing Size Priority* rule, as the rule selects the smallest non-tiebreaking winner set with at least k alternatives; if $k \leq m$ then it selects all m alternatives.

Finally, we study two versions of Monotonicity, an axiom that is very common for example in judgment aggregation. Monotonicity intuitively demands that increasing the support for an alternative cannot hurt this alternative. We first consider a variation that is tailored to the shortlisting setting, stating that if a voter that did not approve the winning alternatives changes his mind and approves all winners, then this cannot change the outcome of an election.

Axiom 11 (Set Monotonicity)**.** Let $E = (C, V = (v_1, \dots, v_n))$ be an election. If $E^* = (C, V^*)$ is another election with $V^* = (v^*_1, \dots, v^*_n)$ such that for some $j \leq n$ we have $v_j \cap \mathcal{R}(E) = \emptyset$, $v^*_j = v_j \cup \mathcal{R}(E)$ and $v^*_l = v_l$ for all $l \neq j$, then $\mathcal{R}(E^*) = \mathcal{R}(E)$.

All of our rules except *First Majority* satisfy Set Monotonicity. Let E be an election with $sc(E) = (2, 2, 1, 1, 1, 1)$. Then under *First Majority* we have $\mathcal{R}(E) = \{c_1, c_2, c_3\}$. Now if a voter who did not approve $\{c_1, c_2, c_3\}$ before approves it, then we get $sc(E) = (3, 3, 2, 1, 1, 1)$ and hence $\mathcal{R}(E) = \{c_1, c_2\}$. Set Monotonicity is a very natural axiom for many applications, so the fact that *First Majority* does not satisfy it makes it hard to recommend the rule in most situations.

We can strengthen this axiom as follows: a voter that previously disapproved all winning alternatives changes her mind and now approves a superset of all (previously) winning alternatives; this should not change the set of winning alternatives. This is a useful property as it guarantees that if an additional voter enters the election, who agrees with the set of currently winning alternatives but might approve additional alternatives, then the set of winning alternatives remains the same and, in particular, does not expand.

Axiom 12 (Superset Monotonicity)**.** Let $E = (C, V = (v_1, \dots, v_n))$ be an election. If $E^* = (C, V^* = (v^*_1, \dots, v^*_n))$ is another election such that for some $j \leq n$ we have $v_j \cap \mathcal{R}(E) = \emptyset$, $\mathcal{R}(E) \subseteq v^*_j$ and $v^*_l = v_l$ for all $l \neq j$, then $\mathcal{R}(E) = \mathcal{R}(E^*)$.

Clearly, Superset Monotonicity implies Set Monotonicity, hence *First Majority* cannot satisfy Superset Monotonicity. Furthermore f-*Threshold*, *Next-k*, *Largest Gap* and *Size Priority* do not satisfy Superset Monotonicity. This is easy to check for f-*Threshold.* For *Next-k*, consider an election E such that $sc(E) = (3, 1, 1)$. Then the winner under *Next-k* is c_1. Now, if a voter changes his mind and additionally approves all three alternatives, then all three alternatives become winners under *Next-k* (for every $k > 1$). Next, consider an election E such that $sc(E) = (2, 1, 0)$. For *Largest Gap*, $\mathcal{R}(E) = \{c_1\}$. If one voter additionally approves $\{c_1, c_2\}$, then $sc(E) = (3, 2, 0)$ and $\mathcal{R}(E) = \{c_1, c_2\}$. For *Size Priority*, consider an election E with $sc(E) = (2, 1, 1)$ and $2 \triangleright 1 \triangleright 3 \triangleright 0$. Then $\mathcal{R}(E) = \{c_1\}$. Now, if one voter additionally approves $\{c_1, c_2\}$, then $\mathcal{R}(E) = \{c_1, c_2\}$. In contrast, *Increasing Size Priority* satisfies Superset Monotonicity as any ties between winners remain. This is essentially the only case for which *Size Priority* satisfies Superset Monotonicity. Finally, as the size of the gap between winners and non-winners cannot decrease and gaps within the winner set remain, *First k-Gap* satisfies Superset Monotonicity for all k (which includes *Approval Voting*).

5 CLUSTERING ALGORITHMS AS SHORTLISTING METHODS

Essentially, the goal of shortlisting is to classify some alternatives as most suitable based on their approval score. The machine learning literature offers a wide variety of clustering algorithms that can perform such a classification. We can turn these algorithms into shortlisting rules in the following way: Let $E = (C, V)$. We use $sc(E)$ as input for a clustering algorithm. This algorithm produces a partition $S_1, \dots, S_k$ of $sc(E)$. The winner set is (the set of candidates corresponding to) the partition that contains the highest score. Under the assumption that the algorithm outputs clusters that are non-intersecting intervals (a condition that any reasonable clustering algorithm satisfies), it is straight-forward to verify that this procedure indeed defines a shortlisting rule, i.e., this rule satisfies Anonymity, Neutrality, Efficiency and is non-tiebreaking. In the following, we focus on linkage-based algorithms [30].

Linkage-based algorithms work in rounds and start with the partition of $sc(E)$ into singletons. Then, in each round, two sets (clusters) are merged until a stopping criterion is satisfied. We consider linkage-based algorithms where always the two clusters with minimum distance are merged. Thus, such algorithms are specified by two features: a distance metric for sets (to select the next sets to be merged) and a stopping criterion. We assume that if two or more pairs of sets have the same distance, then the pair containing the smallest element are merged. Following Shalev-Shwartz and Ben-David [30], we consider three distance measures: the minimum distance between sets (Single Linkage), the average distance between sets (Average Linkage), and the maximum distance between sets (Max Linkage). These three methods can be combined with arbitrary stopping criteria; we consider two: (A) stopping as soon as only two clusters remain, and (B) stopping as soon as every pair of clusters has a distance of $\geq \alpha$. Interestingly, two of our previously proposed methods correspond to linkage-based algorithms: First, if we combine the minimum distance with stopping criterion (A) we obtain the *Largest Gap* rule. Secondly, if we use the minimum distance and impose a distance upper-bound of k (stopping criterion B), we obtain the *First k-Gap* rule. Since these two rules exhibited favorable properties in our axiomatic analysis, it stands to reason that other linkage-based rules may be of interest as well. Our analysis reveals that this is not the case; rules based on other distance measures do not exhibit a particularly interesting set of properties. Due to space constraints we have to omit the specifics.

To conclude, we see that stopping criteria (A) and (B) appear to be only sensible when combined with the minimum distance. However, the average and maximum distance may be more beneficial with different stopping criteria. In particular, the stopping criteria of $\beta > 2$ remaining clusters could be advantageous if winning sets of size roughly m/β are sought. We leave this for future investigation.

6 EXPERIMENTS

In numerical experiments, we want to evaluate the characteristics of the considered shortlisting rules (Python code: [23]).

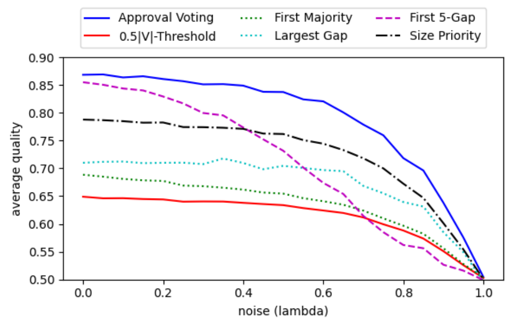
Basic setup. We consider a shortlisting scenario with 100 voters and 30 alternatives. Each alternative c has an objective quality q_c, which is a real number in $[0, 1]$. For each alternative, we generate q_c from a truncated normal (Gauss) distribution with mean 0.5 and standard deviation 0.2, restricted to values in $[0, 1]$. That is, most alternatives are of average quality (around 0.5) and only few have especially high or low quality. (Sampling from a uniform distribution yielded comparable results.) Our base assumption is that voters approve an alternative with likelihood q_c. Thus, the approval score of alternatives are binomially distributed, specifically $sc_E(c) \sim B(100, q_c)$. We then modify this assumption to study a complication for shortlisting: imperfect quality estimates (noise).

The noise model. This model is controlled by a variable $\lambda \in [0, 1]$. We assume that voters do not perfectly perceive the quality of alternatives, but with increasing λ fail to differentiate between alternatives. Instead of our base assumption that each voter approves an alternative c with likelihood q_c, we change this likelihood to $(1 - \lambda)q_c + 0.5\lambda$. Thus, for $\lambda = 0$ this model coincides with our base assumption; for $\lambda = 1$ we have complete noise, i.e., all alternatives are approved with likelihood 0.5. As λ increases from 0 to 1, the amount of noise increases, or, in other words, the voters become less able to judge the quality of alternatives. Additionally, we studied a bias model. The results are largely similar and thus omitted.

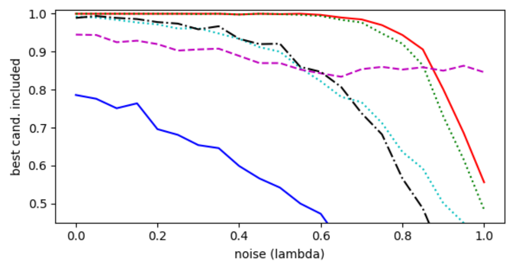
Considered voting rules. We ran our experiments on all rules defined in Section 3. However, we do not mention Next-k, as it returns very large winner sets (average > 25 for any $k \geq 2$); such large winner sets are undesirable for shortlisting. We instantiate *First k-Gap* with $k = 5$ (this corresponds to 5% of the voters). For the *Size Priority* rule we use the priority order $4 \triangleright 5 \triangleright 6 \triangleright \dots$, i.e., an *Increasing Size Priority* rule. Finally, we chose *0.5-Threshold* as representative for threshold rules.

Our comparison of shortlisting rules is visualized in Figure 1 for the noise model (with varying λ). Each data point (corresponding to a specific λ) is based on $N = 1000$ instances $E_1, \dots, E_N$. The figures visualize the behavior of each considered shortlisting rule $\mathcal{R}$ via three metrics: (1) average quality of $\mathcal{R}$'s winner sets, (2) frequency of the objectively best alternative(s) being contained in $\mathcal{R}$'s winner sets, and (3) average size of $\mathcal{R}$'s winner sets.

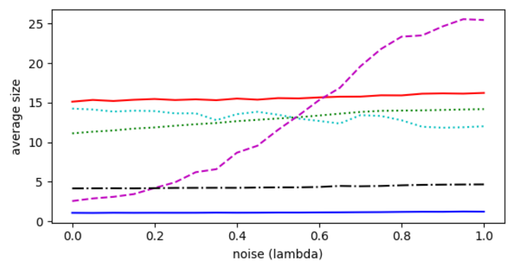
To have a high average quality and to always include the highest-quality alternative can be viewed as somewhat orthogonal objectives. The first objective is easiest to achieve by returning small winner sets, the second by returning large winner sets (so that the objectively best alternative is guaranteed to be included, independently of the voters' noisy perception). This contrast can be seen clearly when comparing *Approval Voting* and *0.5-Threshold*: *Approval Voting* returns rather small winner sets (as seen in Fig. 1c) and thus has a high average quality (Fig. 1a), however if λ increases, the objectively best alternative is often not contained in the winner



(a) Average quality of winner sets



(b) Inclusion of objectively best alternative in winner sets



(c) Average size of winner sets

Figure 1: Numerical simulations for the noise model

set (Fig. 1b). *0.5-Threshold* has a low average quality (Fig. 1a) due to large winner sets (Fig. 1c), but is likely to contain the objectively best alternative even for large λ (Fig. 1b).

Size Priority (with the considered priority order) is a noteworthy alternative to *Approval Voting*. It achieves a very similar average quality, while having a significantly larger chance to include the objectively best alternative. As in shortlisting processes it is generally not necessary to have very small winner sets, we view *Size Priority* (with a sensibly chosen priority order) as superior to *Approval Voting*. *First Majority* and *Largest Gap* produce rather large winner sets and thus the graphs resemble that of *0.5-Threshold*.

We see a very interesting property of *First* 5*-Gap*: it is the only rule where the size of winner sets significantly adjusts to increasing noise. If λ increases, the differences between the approval scores vanishes and thus fewer 5-gaps exist. As a consequence, the winner sets increase in size. This is a highly desirable behavior, as it allows *First* 5*-Gap* to maintain a high likelihood of containing the objectively best alternative without sacrificing average quality for low-noise instances. Also *First Majority* reacts to an increase in noise, albeit to only a very small degree.

To sum up, our experiments show the behavior of shortlisting rules with accurate and inaccurate voters, and the trade-off between large and small winner set sizes. In our opinion, two shortlisting rules have particularly favorable characteristics: 1) *Size Priority* produces small, high-quality winner sets but includes more than just the highest-scoring alternative (as *Approval Voting* does). Thus, it shows a certain robustness to a noisy selection process, as is desirable in shortlisting settings. 2) *First k-Gap* manages to adapt in high-noise settings by increasing the winner set size, the only rule with this distinct feature. This makes it particularly recommendable in settings with unclear outcomes (few or many best alternatives), where a flexible shortlisting method is required.

7 DISCUSSION

Based on our analysis, we recommend three shortlisting methods: *Size Priority*, *First k-Gap*, and *f-Threshold*. *Size Priority*, in particular *Increasing Size Priority*, is recommendable if the size of the winner set is of particular importance, e.g., in highly structured shortlisting processes such as the nomination for awards. Typically, in such processes, the *Multiwinner Approval Voting* rule is used. This rule requires, however, a tiebreaking mechanism. *Size Priority* does not break ties (a requirement we impose on shortlisting rules) and thus removes arbitrariness from the shortlisting process. *Increasing Size Priority* exhibits a very solid behavior in our numerical experiments as well as good axiomatic properties (cf. Table 1).

Our axiomatic analysis reveals *First k-Gap* as a particularly strong rule in that it is the minimal rule satisfying ℓ-Stability. Furthermore, it is the only rule that adapts to increasing noise in our simulations. This behavior is particularly desirable if including the best candidates in the shortlist is more important than the size of the winner set, as is often the case when deciding which applicants should be invited for an interview. A potential disadvantage is that the parameter k has to be chosen according to the given scenario (number of voters, reliability of voters), which requires in-depth knowledge about the shortlisting process.

Finally, Theorem 4 shows that *f-Threshold* rules are the only rules satisfying the Independence axiom. Therefore, if the selection of alternatives should be independent from each other, then clearly a *f-Threshold* rule should be chosen. For example, the inclusion in the Baseball Hall of Fame should depend on the quality of a player and not on the quality of the other candidates.

These recommendations are applicable to most shortlisting scenarios. There are, however, possible variations of our shortlisting framework that require further analysis in the future. For example, while strategyproofness is usually not important with independent experts, there are some shortlisting applications with a more open electorate where this may become an issue [9, 27]. Further, *q-NCSA*, a recently proposed voting rule [17], is worth being investigated in a shortlisting setting. In general, the class of variable multiwinner rules (and social dichotomy functions) deserves further attention as many fundamental questions (concerning proportionality, axiomatic classifications, algorithmic questions, etc.) are unexplored.

Acknowledgments. This work was supported by the Austrian Science Fund (FWF): P31890.

REFERENCES

[1] J. Atsu Amegashie. 1999. The design of rent-seeking competitions: Committees, preliminary and final contests. *Public Choice* 99, 1-2 (1999), 63–76.

[2] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* 48, 2 (2017), 461–485.

[3] Haris Aziz, Edith Elkind, Piotr Faliszewski, Martin Lackner, and Piotr Skowron. 2017. The Condorcet Principle for Multiwinner Elections: From Shortlisting to Proportionality. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*. ijcai.org, 84–90.

[4] Luc Bovens. 2016. Selection under uncertainty: Affirmative action at shortlisting stage. *Mind* 125, 498 (2016), 421–437.

[5] Steven J. Brams and Peter C. Fishburn. 1978. Approval Voting. *The American Political Science Review* 72, 3 (1978), 831–847.

[6] Steven J. Brams and Marc Kilgour. 2012. Narrowing the field in elections: The next-two rule. *Journal of Theoretical Politics* 24, 4 (2012), 507–525.

[7] Florian Brandl and Dominik Peters. 2019. An axiomatic characterization of the Borda mean rule. *Social choice and welfare* 52, 4 (2019), 685–707.

[8] Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, and Rolf Niedermeier. 2019. An Experimental View on Committees Providing Justified Representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019)*. ijcai.org, 109–115.

[9] Robert Bredereck, Andrzej Kaczmarczyk, and Rolf Niedermeier. 2017. On coalitional manipulation for multiwinner elections: Shortlisting. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, Carles Sierra (Ed.). ijcai.org, 887–893.

[10] Baseball Writers' Association of America BWAA. 2019. BBWAA Election Rules. https://baseballhall.org/hall-of-famers/rules/bbwaa-rules-for-election. Accessed: 2019-11-12.

[11] Conal Duddy, Nicolas Houy, Jérôme Lang, Ashley Piggins, and William S Zwicker. 2014. Social dichotomy functions. (2014). Extended abstract for presentation at the 2014 meeting of the Society for Social Choice and Welfare.

[12] Conal Duddy, Ashley Piggins, and William S. Zwicker. 2016. Aggregation of binary evaluations: A Borda-like approach. *Social Choice and Welfare* 46, 2 (2016), 301–333.

[13] Edith Elkind, Piotr Faliszewski, Jean-François Laslier, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2017. What do multiwinner voting rules do? An experiment over the two-dimensional Euclidean domain. In *Thirty-First AAAI Conference on Artificial Intelligence*. AAAI Press, 494–501.

[14] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare* 48, 3 (2017), 599–632.

[15] Ulle Endriss. 2016. Judgment Aggregation. In *Handbook of Computational Social Choice* (1st ed.), Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press, New York, NY, USA, 399–426.

[16] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2017. Multiwinner Voting: A New Challenge for Social Choice Theory. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access, Chapter 2, 27–47.

[17] Piotr Faliszewski, Arkadii Slinko, and Nimrod Talmon. 2020. The complexity of multiwinner voting rules with variable number of winners. In *Proceedings of 24th European Conference on Artificial Intelligence (ECAI 2020)*.

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

[19] Clemens Gangl, Martin Lackner, Jan Maly, and Stefan Woltran. 2019. Aggregating expert opinions in support of medical diagnostic decision-making. In *Knowledge Representation for Health Care/ProHealth (KR4HC)* (Poznan, Poland). 56–62.

[20] Marc Kilgour. 2010. Approval balloting for multi-winner elections. In *Handbook on Approval Voting*, J.-F. Laslier and R. Sanver (Eds.). Springer, 105–124.

[21] Marc Kilgour. 2016. Approval elections with a variable number of winners. *Theory and Decision* 81 (02 2016).

[22] Marc Kilgour and Erica Marshall. 2012. Approval balloting for fixed-size committees. In *Electoral Systems*. Springer, 305–326.

[23] Martin Lackner and Jan Maly. 2020. Python code for "Approval-Based Shortlisting". https://doi.org/10.5281/zenodo.3821983

[24] Martin Lackner and Piotr Skowron. 2020. Approval-Based Committee Voting: Axioms, Algorithms, and Applications. *arXiv preprint arXiv:2007.01795* (2020).

[25] Martin Lackner and Piotr Skowron. 2021. Consistent Approval-Based Multi-Winner Rules. *Journal of Economic Theory* 192 (2021), 105173.

[26] Ariel D. Procaccia and Nisarg Shah. 2015. Is approval voting optimal given approval votes?. In *Advances in Neural Information Processing Systems*. 1801–1809.

[27] Jameson Quinn and Bruce Schneier. 2016. A proportional voting system for awards nominations resistant to voting blocs. (2016). Preprint per https://www.schneier.com/academic/archives/2016/05/a_proportional_votin.html, Accessed: 2019-11-14.

[28] David M. Pennock Rupert Freeman, Anson Kahng. 2020. Proportionality in Approval-Based Elections With a Variable Number of Winners. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-2020)*. ijcai.org, 132–138.

[29] Luis Sánchez-Fernández and Jesús A Fisteus. 2019. Monotonicity axioms in approval-based multi-winner voting rules. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2019)*. International Foundation for Autonomous Agents and Multiagent Systems, 485–493.

[30] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge University Press, New York, NY, USA.

[31] Amit Singh, Catherine Rose, Karthik Visweswariah, Vijil Chenthamarakshan, and Nandakishore Kambhatla. 2010. PROSPECT: a system for screening candidates for recruitment. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. ACM, 659–668.

[32] The Hugo Awards. 2019. The Voting System. http://www.thehugoawards.org/the-voting-system. Accessed: 2019-11-12.

[33] The Man Booker Prize. 2018. Rules & Entry Form. https://thebookerprizes.com/sites/manbosamjo/files/uploadedfiles/files/ManBookerPrize2018RulesAndEntryForm.pdf. Accessed: 2019-11-13.

[34] Hamish Mark Tweeddale, Ronf F. Cameron, and Steven S. Sylvester. 1992. Some experiences in hazard identification and risk shortlisting. *Journal of Loss Prevention in the Process Industries* 5, 5 (1992), 279–288.

[35] William S. Zwicker and Hervé Moulin. 2016. Introduction to the theory of voting. In *Handbook of Computational Social Choice* (1st ed.), Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press, New York, NY, USA, 23–56.