Mathematical Social Sciences 135 (2025) 102417

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/mss

Election methods and political polarization

Salvatore Barbaro[1]

Johannes-Gutenberg University, Interdisciplinary Public Policy, Mainz, 55099, RP, Germany

ARTICLE INFO

JEL classification:
D71

Keywords:
Elections
Voting rules
Political polarization
Scoring rules

ABSTRACT

Political polarization poses a significant challenge to democratic societies. While much of the scholarly focus has been on the socio-demographic factors that drive polarization, this paper focuses on voting rules and their effects on the incentives for candidates to be moderate or polarizing. It addresses the question: Which voting methods most impede the success of a polarizing candidate? Through a comparative analysis of plurality voting, the Condorcet method, and Borda's rule, we find the Borda rule to be the most effective at discouraging extreme platforms. In a generalization, we show that a scoring rule is more effective at hindering the success of polarizing candidates the more weight that it places on voters' second preferences in the tallying process.

1. Introduction

The escalation of political polarization endangers democratic principles and peaceful coexistence in modern societies (Sen, 2020). While prevailing research attributes this trend primarily to socio-economic factors and the role of the media (Bail et al., 2018; Levy, 2021), the present paper focuses on a less explored but equally critical aspect: the role of election methods. By shaping how voter preferences are amalgamated, voting rules influence polarization and affect candidates' incentives to adopt polarizing strategies.

This study examines how well various voting rules impede the electoral success of polarizing candidates. This research is motivated by the observation that the commonly used plurality rule creates strong incentives for polarization. It encourages candidates to focus their campaigns solely on their supporters, disregarding how non-supporters perceive them. In contrast, voting methods such as the simple-majority rule (hereafter referred to as the Condorcet method) and the Borda rule consider voters' entire preference orderings, albeit through distinct mechanisms. To illustrate this point, consider the stylized example in Table 1.

The electorate consists of three distinct groups of voters, each with identical and strict preferences within the group. At time t_0, candidate a emerges as the plurality winner despite being the Condorcet loser. A closer examination reveals that candidate d is both the Condorcet winner and the winner under the Borda rule.

Because of the election outcome at t_0, candidate d adopts a polarizing stance for the subsequent voting round at t_1—one that appeals strongly to a minority but is ranked last by the majority. Consequently, while d rises in the preferences of Group 1, they fall in the rankings of all other voters. Once a Condorcet winner, candidate d now becomes the Condorcet loser. This strategy proves effective under the plurality rule, where d emerges as the winner despite promoting further division among the electorate.

The question of how different voting systems influence the success of polarizing candidates has been examined in various empirical studies. Using data from Italian municipal elections, Bordignon et al. (2016) showed that a runoff system, where the top two candidates from the first round advance to a second round, makes it harder for extreme preferences to win and favors moderate preferences compared to a single-stage plurality rule. Similarly, Laslier and Van der Straeten (2008), in a comprehensive in-situ experiment during the 2002 French presidential elections, showed that the far-right candidate, who secured the second-highest number of votes (earning him a place in the runoff), would have performed significantly worse under Approval Voting. Focusing on the 2012 presidential elections in France, Baujard et al. (2014) found that consensus-oriented candidates ('inclusive candidates') performed better under Approval Voting and Range Voting than under the actual runoff system. Two in-situ experiments conducted in Germany by Alós-Ferrer and Granić (2012) also concluded that Approval Voting tends to favor moderate and consensus-oriented candidates compared to the plurality rule.

These studies typically do not compare a broad range of voting systems but rather an individual voting rule against the commonly

E-mail address: sbarbaro@uni-mainz.de.

[1] The author is indebted to the editor and the associate editor for their insightful and professional handling of the submission, as well as for their valuable comments. The paper has greatly benefited from the reports of three anonymous reviewers, to whom the author is deeply grateful. Additionally, the author wishes to thank Anna-Sophie Kurella, Martin Kolmar, Giuliano Oggioni, Maike Roth, Philip Sauré, Juliane Steffens, and Jens Südekum for their valuable feedback. The author is also thankful for the fruitful discussions following the presentation of an earlier version of this paper at the Social Choice and Welfare Conference in Paris in 2024.

https://doi.org/10.1016/j.mathsocsci.2025.102417
Received 11 November 2023; Received in revised form 16 April 2025; Accepted 28 April 2025
Available online 12 May 2025

Table 1
Voters' preferences and electoral outcomes.

Election at time t_0			Election at time t_1		
Group 1 34%	Group 2 33.5%	Group 3 32.5%	Group 1 34%	Group 2 33.5%	Group 3 32.5%
a	b	c	d	b	c
d	d	d	a	c	b
c	c	b	c	a	a
b	a	a	b	d	d

Table 2
Polarized profiles $\mathscr{R}$.

$\mathscr{R}$	n_1	n_2	n_3	n_4	n_5	n_6	k	r
①	γ_1	0	γ_3	0	γ_2	0	0	γ_3
②	γ_1	γ_3	0	0	γ_2	0	γ_3	0
③	0	γ_1	γ_3	0	γ_2	0	γ_3	0
④	γ_3	γ_1	0	0	γ_2	0	γ_3	0
⑤	γ_1	0	0	γ_3	γ_2	0	γ_3	γ_3
⑥	γ_1	0	0	0	γ_2	γ_3	0	0
⑦	0	γ_1	0	γ_3	γ_2	0	0	0
⑧	0	γ_1	0	0	γ_2	γ_3	0	0

applied plurality rule. As we illustrated in the example above and will show analytically below, the plurality rule strongly favors polarizing candidates. It is, therefore, unsurprising that other voting methods reward polarizing candidates less. This paper aims to provide a comprehensive theoretical analysis of the effect of voting rules on the success of polarizing candidates.

In its methodological aspect, our contribution is most closely related to Dasgupta and Maskin (2008) and Dasgupta and Maskin (2020). They proved the comparative robustness of the Condorcet method in terms of domain requirements. Whenever the plurality rule works well (i.e., complies with a standard set of axioms) on a particular domain, so does the Condorcet method, but not vice versa. They show that the same applies to the comparison between the Borda rule and the Condorcet method.

We adopt existing models of polarized electorates for our analysis, in particular drawing on models that explicitly take into account the roles of consensual and polarizing candidates (Gehrlein, 2005; Miller, 2013; Campante and Hojman, 2013), to analyze our specific research question using established frameworks.

This paper is organized as follows. We begin by introducing the analytical framework in Section 2. The performance of polarizing candidates under the plurality rule, the Condorcet method, and the Borda rule is examined in Section 3. Section 4 broadens the perspective to include additional voting methods and demonstrates that the anti-plurality rule is the most robust against polarization. The paper concludes with a summary and discussion in Section 5.

2. The model

2.1. Framework

We assume a continuum of voters, indexed by points on the unit interval $[0,1]$, that are sorted into three groups. Because the Lebesgue measure is not defined for all subsets of the interval, we will restrict attention to *Borel profiles*. In doing so, we follow the framework developed by Dasgupta and Maskin (2008). In particular, the choice domain is a set of three candidates $\{A, B, C\}$. The voters have anti-symmetric (strict) orderings over the candidates, reflecting their true (honest) preferences.

We use the framework of Gehrlein (2005) and Campante and Hojman (2013) to model a polarized electorate. The framework is characterized by one candidate (without loss of generality candidate A) who is perceived as the most preferred by a significantly large group and simultaneously viewed as the least favorable candidate by another large group. In this sense, candidate A is polarizing.

We call the first group the 'supporters' and the second group the 'despisers'. A third, 'moderate', group ranks the candidates arbitrarily. The groups constitute fractions $\gamma_1, \gamma_2, \gamma_3$ of the electorate with $\gamma_1, \gamma_2 \in \left(\gamma_3, \frac{1}{2}\right)$.

Dividing the electorate into three groups is senseless if multiple groups can have the same preference orderings. Thus, we impose the assumption that distinct groups have distinct preferences. Due to the restriction to anti-symmetric preferences, voters can have six different preference orderings $n_1, \ldots, n_6$ that represent the specific preference ordering and the corresponding population share of individuals holding that preference.

$n_1 : A \succ B \succ C,$

$n_2 : A \succ C \succ B,$

$n_3 : B \succ A \succ C,$

$n_4 : C \succ A \succ B,$

$n_5 : B \succ C \succ A,$

$n_6 : C \succ B \succ A.$

Given the structure of polarized profiles, sixteen possible profiles can appear. Eight are isomorphic to the other eight profiles, which differ only in their relative ranking of B and C. Thus, without loss of generality, we denote by $\mathscr{R}$ the domain of polarized profiles. Table 2 summarizes the setup.

Columns k and r present two parameters introduced by Gehrlein (2005) to characterize polarizing profiles based on the degree to which they reflect consensus or polarization. The parameter k is defined as

$$k \equiv \min\{(n_1+n_3),(n_2+n_4),(n_5+n_6)\} \in \left[0, \tfrac{1}{3}\right). \tag{1}$$

A profile with a low k-value has some candidate that only few voters think is the worst candidate. A k-value of zero indicates that at least one candidate is never ranked last and at least one candidate is polarizing (see Profiles ①, ⑥–⑧). Conversely,

$$r \equiv \min\{(n_3+n_4),(n_1+n_6),(n_2+n_5)\} \in \left[0, \tfrac{1}{3}\right) \tag{2}$$

measures the number of candidates that are ranked second. The parameter r is zero if some candidate is seen as either top or bottom (thus, never medium) by all voters. We will claim that such a candidate polarizes. In our model, both parameters, k and r, can be described solely by the share of the moderate group, γ_3. This, in turn, explains why both k and r cannot exceed one-third: γ_3 is the smallest group, and we have constrained the size of the other groups to the interval $\left(\gamma_3, \frac{1}{2}\right)$.[2]

For clarity, we define what we mean by a polarizing candidate and, in contrast, a consensus candidate.

Definition 1 (*Polarizing Candidate*). A polarizing candidate (PC) is either

1. ranked best by their supporters and ranked least by their despisers or

[2] The range of values for k and r on $\left(0, \frac{1}{3}\right)$ applies to a general three-candidate model (Gehrlein, 2005, p. 323). This can be understood as follows in the case of k: k measures how frequently individual candidates are ranked last by voters and picks the smallest of these values. If candidates A and B are each perceived as the least preferred option by more than one-third of the electorate, then fewer than one-third of the voters must rank C in last place. Conversely, if exactly one-third of the electorate ranks both A and B as their least preferred candidate, then C will also attain a value of one-third. A similar argument applies to r.

2. ranked best by their despisers and ranked least by their supporters.

Note that there is always one, and in one-half of the eight profiles, even two, polarizing candidates (A is always a PC, and B is a PC whenever ranked last by the *supporters*). Obviously, $r = 0$ implies that there is a polarizing candidate in the respective profile, but not vice versa (when a PC is middle-ranked by the moderates).

Definition 2 (*Consensus Candidate*)**.** A consensus candidate is a candidate who is never ranked last by any voter.

This definition corresponds to that of a 'centrist candidate' by Miller (2013) and strengthens the concept of a 'unifying candidate' by Gehrlein (2005), defined as a candidate who is ranked lowest by few voters (or, in the special case, by none). The k-value of a profile with a consensus candidate is zero.

Lemma 1 (*Consensus Candidate in Polarized Profiles*)**.** *On the domain $\mathscr{R}$ of polarized profiles, every consensus candidate is a Condorcet winner and cannot be defeated under the Borda rule.*

To prove Lemma 1, we must bear in mind that in a three-candidate competition, a candidate can only defeat a Condorcet winner under the Borda rule if the Condorcet winner is bottom ranked by at least one group. This follows immediately from the property of Quasi-Agreement (Dasgupta and Maskin, 2008).[3]

A distinctive feature of profile ⑤ is that it displays a Condorcet cycle (i.e., cyclic majorities) as well as positive values for k and r. When γ_3 (hence k and r) is close to zero, candidate B is last-ranked by only a few, while candidate A stands out as a strongly polarizing candidate. As $\gamma_3 \to 1/3$, this changes, and all candidates can be 'regarded'[4] as polarizing, with each being top-ranked and bottom-ranked by approximately one-third of the electorate.

A similar pattern emerges in profile ②. In situations where the group of moderates is small, candidate B is rejected by only a few but is viewed as the best alternative by many. Under these circumstances, B takes on the role of the primary challenger to the polarizing candidate A. When γ_3 reaches higher values, though, B can be 'regarded' as a polarizing candidate.

2.2. Voting rules

We consider the Condorcet method and scoring rules.

The Condorcet method is a voting system in which the winner is the candidate who defeats every other candidate in head-to-head majority comparisons. However, due to the Condorcet paradox, a Condorcet winner need not exist. In such cases, no winner is declared.[5]

Table 3
Electoral outcomes on profile $R \in \mathscr{R}$.

$\mathscr{R}$	PC	CC	Condorcet	Borda	Plurality
①	A	B	**B**	**B**	**B**
②	A	∅	A	$A\|$**B**	A
③	A, B	∅	B	$A\|B$	B
④	A, B	∅	A	$A\|B$	A
⑤	A	∅	∅	$A\|$**B**	$A\|$**B**
⑥	A	B	**B**	**B**	$A\|$**B**
⑦	A, B	C	**C**	**C**	$A\|B$
⑧	A, B	C	**C**	**C**	$A\|B$

Scoring rules constitute a class of electoral systems where points are allocated to candidates based on the order of preference expressed by the voters, with the winner being the candidate accumulating the highest total score. In a contest involving κ candidates and assuming anti-symmetric preferences, a scoring vector is a κ-tuple $(s_1, s_2, \ldots, s_{\kappa-1}, s_\kappa)$, made up of real numbers that satisfy $s_1 \le s_2 \le \cdots \le s_\kappa$, and $s_1 < s_\kappa$. For elections featuring three candidates, a normalized vector $(1, \lambda, 0)$ can describe the diversity of scoring rules.

For instance, the plurality rule corresponds to the $\lambda = 0$, the Dowdall rule (Grofman et al., 2017) to $\lambda = \frac{1}{4}$, the Borda rule to $\lambda = \frac{1}{2}$, and the anti-plurality rule (Bossert and Suzumura, 2016) to $\lambda = 1$.

3. Plurality, Condorcet, Borda

At its core, we examine how the polarizing candidate — or one of the polarizing candidates — performs under different voting systems. In the related literature, the three most well-known voting systems are frequently compared: the plurality rule, the Condorcet method, and the Borda rule. At first, we also focus on these three voting methods, but later generalize our results.

Our first result is that the Borda rule shows the strongest resilience to the success of polarizing candidates.

Theorem 1 (*Robustness of the Borda Rule*)**.**

1. *If a polarizing candidate is elected under the Borda rule for any profile R in the domain of polarized profiles $\mathscr{R}$, then a polarizing candidate is also elected under plurality rule, but not vice versa.*
2. *If a polarizing candidate is elected under the Borda rule for any profile $R \in \mathscr{R}$, then a polarizing candidate is also elected under the Condorcet method (if a Condorcet winner exists), but not vice versa.*

A detailed proof is provided in Appendix A, while Table 3 presents a concise summary. The columns PC and CC indicate the sets of polarizing and consensus candidates in each profile $R \in \mathscr{R}$. The columns Condorcet, Borda, and Plurality specify the candidate who wins under each respective voting method. If the winner is not a polarizing candidate, this is highlighted in bold.

In some cases, the Borda or plurality winner (A or B) depends on the specific configuration of the γ-values. We denote these cases by $A|B$ in Table 3. The Borda winner in profile ② is B if the condition $\gamma_1 > 2-4\gamma_2$ is satisfied. In profile ⑤, B emerges as the Borda winner when γ_2 is greater than one-third. In profiles ⑤ to ⑧, A emerges as the plurality winner if $\gamma_1 > \gamma_2$.

In profiles ②, ③, ④, ⑦, and ⑧, the polarizing candidate — or one of the polarizing candidates — prevails under the plurality rule.

[3] Quasi-Agreement holds if all voters agree on the relative ranking of one candidate, i.e., all voters regard that candidate as either best, or medium, or worst. If Quasi-Agreement holds, the Borda winner coincides with the Condorcet winner (Dasgupta and Maskin, 2008, p. 961). As an example, consider a three-candidate setup with $\{x, y, z\}$ as candidates in which x is the most-preferred candidate of voters in Group 1, and y is the most-preferred candidate of voters in Group 2 and 3. Then, y emerges as the Condorcet winner. If both x and y are consensus candidates, then all voters must rank z last, so Quasi-Agreement holds, and x cannot defeat y under the Borda rule.

[4] We are aware that with the term 'regarded' we deliberately use a vague term. Candidate B, according to Definition 1, is not classified as a polarizing candidate. However, the definition may be considered overly restrictive, particularly when $\gamma_3 \to 1/3$. Ultimately, Definition 1 establishes a cutoff level, and here we examine whether this threshold determines the key results.

[5] In the literature, two principal approaches have been proposed to address the indeterminacy arising from cyclic majorities. The first approach, which we adopt in this study, assumes an "empty choice set" (e.g. Sen, 2014; Dasgupta and Maskin, 2008, p. 956), meaning that no winner is declared in the occurrence of a Condorcet paradox. The second approach, by contrast, assigns all candidates to the set of winners in such cases (e.g., Holliday and Pacuit, 2021, Def. 6). Applying the latter approach to our analysis, any candidate could emerge as the winner in profile ⑤, including non-polarizing candidates, even in scenarios where the plurality rule or the Borda rule selects a polarizing candidate. Such an outcome would ultimately be random and, within the context of this study, less compelling.

Similarly, in profiles ⑤ and ⑥, the polarizing candidate is the plurality winner if $\gamma_1 > \gamma_2$. In other words, only in profile ① does the plurality rule reliably prevent the election of a polarizing candidate. In this profile, the non-polarizing candidate B prevails under each voting method.

When the winner is determined by the Condorcet method, polarizing candidates do not prevail in profiles ⑥ through ⑧. This is because, in these profiles, the polarizing candidate(s) benefit from vote splitting, a phenomenon to which the Condorcet method is *immune* (Maskin, 2025).

Whenever a polarizing candidate fails to win under the Condorcet method, the same candidate also fails to win under the Borda rule. However, the converse does not hold: in profile ②, there exists a domain where candidate B defeats candidate A under the Borda rule.

In Profile ⑤, the non-polarizing candidate B emerges as the Borda winner as soon as the share of despisers (γ_2)) exceeds one-third. The condition for B to emerge as plurality winner is $\gamma_2 > \gamma_1$, which implies $\gamma_2 > \frac{1}{3}$ (see Appendix A).

We already noted at the end of Section 2.1 that the results lose persuasiveness as $\gamma_3 \to 1/3$.

Regarding profile ②, the first point to note is that candidate C cannot emerge as the Borda winner (as proven in Appendix A). When $\gamma_3 \to 1/3$, candidate B loses their status as consensus/unifying candidate and γ_1 and γ_2 must be close to each other, rendering the condition for B to emerge as Borda winner ($\gamma_1 > 2 - 4\gamma_2$) unsatisfied. Consequently, for high values of γ_3, candidate A wins for all three voting methods.

Regarding ⑤, we will postpone the discussion to the next section, where we will explore the electoral results by applying Condorcet-consistent rules.

4. Generalizations

4.1. Condorcet-consistent rules

A Condorcet-consistent voting rule (aka Condorcet extension) is defined by its ability to elect a Condorcet winner whenever such a candidate exists and use some other procedure to select a candidate in the case of cyclical majorities (Moulin, 1991, Def. 9.2). Most of the widely-known Condorcet-consistent voting rules (for example, maximin, Young's rule, Kemeny's rule, Dodgson's rule, Split Cycle) follow a procedure established by Condorcet himself, breaking cyclical majorities at the point with the smallest majority (Brandt et al., 2024).

Remark 1 (*Condorcet-Consistent Rules are More Polarizing Than the Borda Rule*)**.** When no Condorcet winner exists, Condorcet-consistent rules that break cyclical majorities at the point with the smallest majority select the plurality winner instead of the Borda winner for any profile R in the domain of polarized profiles $\mathscr{R}$. Consequently, in profiles with cyclical majorities, Condorcet-consistent rules tend to favor polarizing candidates more than the Borda rule does.

Remark 1 complements Theorem 1 (Statement 2) for the case in which no Condorcet winner exists and the winner is determined according to the smallest majority principle within Condorcet-consistent rules.[6]

The proof of Remark 1, which we present in Appendix B, proceeds in two steps. First, candidate C cannot win under Condorcet-consistent rules. Second, the winner (either A or B) always coincides with the plurality winner.

[6] The voting rules mentioned above are designed to follow the smallest-margin principle. Using computational methods, we verify that all Condorcet-consistent rules implemented in standard Python and R packages produce the same outcomes as described in Remark 1. The corresponding scripts are available from the author upon request.

Table 4
Non-polarizing candidates and scores.

$\mathscr{R}$	NPC	Score$_A$	Score$_B$	Score$_C$
①	B, C	$\gamma_1 + \lambda\gamma_3$	$1 + \gamma_1(\lambda - 1)$	$\lambda\gamma_2$
②	B, C	$\gamma_1 + \gamma_3$	$\lambda\gamma_1 + \gamma_2$	$\lambda(\gamma_2 + \gamma_3)$
③	C	$\gamma_1 + \lambda\gamma_3$	$\gamma_2 + \gamma_3$	$\lambda(\gamma_1 + \gamma_2)$
④	C	$\gamma_1 + \gamma_3$	$\gamma_2 + \lambda\gamma_3$	$\lambda(\gamma_1 + \gamma_2)$
⑤	B, C	$\gamma_1 + \lambda\gamma_3$	$\gamma_2 + \lambda\gamma_1$	$\lambda\gamma_2 + \gamma_3$
⑥	B, C	γ_1	$\gamma_2 + \lambda(\gamma_1 + \gamma_3)$	$\lambda\gamma_2 + \gamma_3$
⑦	C	$\gamma_1 + \lambda\gamma_3$	γ_2	$\lambda + \gamma_3(1 - \lambda)$
⑧	C	γ_1	$\gamma_2 + \lambda\gamma_3$	$\lambda + \gamma_3(1 - \lambda)$

4.2. Scoring rules

Under the plurality rule ($\lambda = 0$), a candidate's electoral outcome depends solely on whether they are ranked first by the voters. The picture is different for scoring rules with $\lambda \in (0, 1]$. Second-ranked candidates are considered in the tally, which implies a relative disadvantage for the least-preferred candidates. This mechanism is inherently relative because a candidate consistently ranked either first or last receives the same score regardless of the value of λ. However, their prospects for victory diminish as second-ranked candidates accrue higher scores, an effect that intensifies with increasing λ.

For the sake of clarity, Table 4 displays the scores of the respective candidates. The second column, headed NPC, denotes the set of candidates that are not polarizing.

Theorem 2 (*Increasing λ Hinders the Success of Polarizing Candidates*)**.** *If a scoring rule with $\lambda \in [0, 1]$ applied on any profile R in the domain of polarized profiles $\mathscr{R}$ elects a non-polarizing candidate, then a scoring rule with $\dot{\lambda} > \lambda$ elects the non-polarizing candidate on the same profile, too.*

We split the proof into three parts. In the sequel, we denote by $x\, P^{\lambda}\, y$ that, for a given λ, candidate x defeats y (i.e., Score$_x$ > Score$_y$).

Proof. The first and most straightforward part considers ①. Here, candidate B defeats the polarizing candidate A regardless of λ. To demonstrate this, observe that $A P^{\lambda} B \iff \lambda < \frac{2\gamma_1 - 1}{\gamma_1 - \gamma_3} < 0$. Thus, for the polarizing candidate A to defeat B, λ must be negative, ensuring that B prevails for all $\lambda \in [0, 1]$.

The second part concerns profiles ②, ⑤, and ⑥. First, in these profiles, candidate B defeats candidate C for all $\lambda \in [0, 1]$. To see this, consider that

1. $C P^{\lambda}_{②} B \iff \lambda(\gamma_2 + \gamma_3) > \lambda\gamma_1 + \gamma_2 \iff \lambda > \frac{\gamma_2}{1 - 2\gamma_1} > 1$. The λ-value which would be required for C to beat B lies beyond its maximum value of 1.
2. For $\gamma_1 \neq \gamma_2$, the relation $C P^{\lambda}_{⑤} B$ holds if and only if
$$\lambda \cdot \operatorname{sgn}(\gamma_2 - \gamma_1) > \frac{\gamma_1 + 2\gamma_2 - 1}{|\gamma_2 - \gamma_1|}.$$
If $\gamma_2 > \gamma_1$, then $C P^{\lambda}_{⑤} B$ requires $\lambda > 1$; if $\gamma_1 > \gamma_2$, a higher score for C than for B necessitates a negative λ-value. In either case, the condition implies that $C P^{\lambda}_{⑤} B$ holds only if $\lambda \notin [0, 1]$. In the special case where $\gamma \equiv \gamma_1 = \gamma_2$, $C P^{\lambda}_{⑤} B \iff 1 - \gamma(2 - \lambda) > \gamma(1 + \lambda) \iff \gamma < 1/3$. This condition would require $\gamma_3 > \gamma$, which is excluded by assumption.
3. $C P^{\lambda}_{⑥} B \iff \lambda(\gamma_2 - \gamma_1 - \gamma_3) > \gamma_2 - \gamma_3$. Since the left-hand side expression in brackets is negative, for C to beat B, we must have $\lambda < \frac{\gamma_2 - \gamma_3}{\gamma_2 - \gamma_3 - \gamma_1} < 0$, hence $\lambda \notin [0, 1]$.

Secondly, we show that the derivative of the scores with respect to λ is higher for the non-polarizing candidate B than for the polarizing candidate A. For better clarity, let $\sigma_x^{ⓟ}$ denote the derivative of the score for candidate $x \in \{A, B, C\}$ on profile ⓟ $\in \mathscr{R}$. We have:

$\sigma_B^{②} = \gamma_1 > \sigma_A^{②} = 0,$

Table 5
Eliminated and winning candidates in multi-stage scoring rules.

$\mathscr{R}$	PW	Plur. Elim.		BW	Borda Elim.		APW	Anti-Plur. Elim.	
		Elim.	Win		Elim.	Win		Elim.	Win
①	**B**	*C*	**B**	**B**	*C*	**B**	**B**	*C*	**B**
②	*A*	*C*	*A*	*A*|**B**	*C*	*A*	**B**	*C*|*A*	**B**
③	*B*	*C*	*B*	*A*|*B*	*C*	*B*	**C**	*A*|*B*	*A*|*B*
④	*A*	*C*	*A*	*A*|*B*	*C*	*A*	**C**	*A*|*B*	*A*|*B*
⑤	*A*|**B**	*C*	*A*	*A*|**B**	*C*	*A*	**B**	*A*|*C*	*A*|**B**
⑥	*A*|**B**	*C*	**B**	**B**	*A*|*C*	**B**	**B**	*A*	**B**
⑦	*A*|*B*	*C*	*A*	**C**	*A*|*B*	**C**	**C**	*B*	**C**
⑧	*A*|*B*	*C*	*B*	**C**	*A*|*B*	**C**	**C**	*A*	**C**

$$\sigma_B^{⑤} = \gamma_1 > \sigma_A^{⑤} = \gamma_3,$$
$$\sigma_B^{⑥} = \gamma_1 + \gamma_3 > \sigma_A^{⑥} = 0.$$

Hence, if the non-polarizing candidate B is elected by a given λ-scoring rule, then they must continue to be elected as λ increases.

The third part treats the remaining profiles ③, ④, ⑦, and ⑧. In each of these profiles, candidate C is the only non-polarizing candidate. We have to show that if C is elected for a given λ in a profile, then they are also elected with $\dot{\lambda} > \lambda$:

$$\sigma_C^{③} = \gamma_1 + \gamma_3 > \sigma_A^{③} = \gamma_3 > \sigma_B^{③} = 0,$$
$$\sigma_C^{④} = \gamma_1 + \gamma_2 > \sigma_B^{④} = \gamma_3 > \sigma_A^{④} = 0,$$
$$\sigma_C^{⑦} = \gamma_1 + \gamma_2 > \sigma_A^{⑦} = \gamma_3 > \sigma_B^{⑦} = 0,$$
$$\sigma_C^{⑧} = \gamma_1 + \gamma_2 > \sigma_B^{⑧} = \gamma_3 > \sigma_A^{⑧} = 0. \quad \square$$

4.3. Scoring elimination rules

The experimental and empirical studies mentioned in the Introduction primarily focus on sequential voting systems and Approval Voting. Against this backdrop, and given that sequential voting systems (particularly runoff and Hare systems) are frequently applied in practice, they will be included in our analysis as well.

Lepelley et al. (2018) distinguished three types of scoring elimination rules: plurality elimination (eliminating candidates with the fewest votes), Borda elimination rules (removing candidates with the lowest Borda scores[7]), and anti-plurality-based elimination.

Table 5 summarizes the resilience of each rule to the success of polarizing candidates. The columns `Elim.` indicate the candidate eliminated by the respective family of rules in the first round, while the columns `Win` show the corresponding winners. As in Table 3, non-polarizing winners are typeset in bold. For clarity and ease of comparison, the columns `PW`, `BW`, and `APW` refer to the winners under the plurality rule, Borda rule, and anti-plurality rule, respectively.

Remark 2 (*Resilience of Elimination Rules to Polarization*)**.** For profiles R in the domain $\mathscr{R}$ of polarized profiles:

1. The plurality-elimination rule is not, in general, more resilient to the election of a polarizing candidate than the standard plurality rule. Only in Profile ⑥ does the elimination variant clearly prevent the victory of a polarizing candidate, in contrast to the plurality rule, under which either a polarizing or a non-polarizing candidate may win. Conversely, in Profile ⑤, a non-polarizing candidate can win under the plurality rule but not under its elimination variant.
2. The Borda-elimination rule is more susceptible to the election of a polarizing candidate than the Borda rule. In Profiles ② and ⑤, a non-polarizing candidate can emerge as the Borda winner but cannot win under its elimination variant.
3. The anti-plurality elimination rule is markedly less resilient to polarization than the anti-plurality rule. Under the elimination variant, a polarizing candidate wins in Profiles ③ and ④, and possibly also in Profile ⑤. In contrast, no polarizing candidate ever wins under the anti-plurality rule in these cases.
4. Overall, Borda and anti-plurality elimination rules exhibit greater resilience to polarization than the plurality rule.

The fourth statement in Remark 2 aligns with empirical findings (Bordignon et al., 2016). However, the effect is modest and appears to result less from the high resilience of plurality-elimination rules than from the pronounced susceptibility of the plurality rule to elect polarizing candidates. Overall, we find no positive contribution of elimination-based rules compared to their respective single-stage counterparts regarding the resilience to polarization.

4.4. Approval voting

Most experimental studies focus on Approval Voting (AV), but incorporating AV into our framework hinges on voters' cut-off levels. If all voters approve the same number of candidates, AV can either favor or hinder polarization. This occurs because AV mirrors plurality voting when voters set their cut-off level between the first and second candidates. Conversely, when voters approve two candidates, the results are the same as anti-plurality voting.

AV's effects become more varied when voters have different cut-off levels, as discussed by Gehrlein et al. (2016, Sec. 2.2). In some cases, AV can bolster polarizing candidates even more than does the plurality rule, particularly when major groups approve of only one candidate while moderates approve of two.[8]

AV does not have an unambiguously moderating effect. While AV may perform well with multiple candidates, as demonstrated in field experiments, voter cut-off levels remain an empirical question. It must be acknowledged, however, that the effect of AV cannot be adequately represented within a three-candidate model.

5. Conclusion

This paper has studied the impact of voting rules on the success of polarizing candidates. We have found that while the frequently used plurality rule incentivize candidates to deepen existing division, the Condorcet method and, even more so, the Borda rule create stronger incentives for moderating policy platforms. Generally, the greater the weight a scoring rule assigns to second-ranked candidates, the stronger the disincentive for polarization. Under the anti-plurality rule, incentives for societal division are absent.

These findings have implications for the ongoing debates about electoral reform. Having said this, the analysis in the current paper is not a plea for the anti-plurality rule but rather an effort to highlight the intricate relationship between polarization and election methods. While the anti-plurality rule is unlikely to gain political traction, we emphasize the importance of considering additional desirable criteria for a voting system beyond its effect on polarization. The Borda rule is well known for its axiomatic advantages (Young, 1975; Moulin, 1991,

[7] This method is known as the Baldwin method. A related version is the Nanson rule, which eliminates based on the Borda scores but uses the average Borda score as threshold. We neglect the differences for the sake of simplicity.

[8] For example, in Profile ①, if $\gamma_1 > \gamma_2$, candidate B emerges as plurality winner, but the polarizing candidate A wins under Approval Voting. This occurs because moderates inadvertently enable A's victory by approving two candidates despite preferring B. Whether such voting behavior can be rationalized cannot be adequately addressed within the framework of this paper, which assumes honest voting throughout.

Ch. 9.3); our contribution adds to these by highlighting the significant amount of discouragement it gives to political polarization.

One limitation of this study lies in its exclusion of voters' strategic considerations. Future research is encouraged to incorporate strategic behavior into the framework. Another limitation is that our model focuses on elections with three candidates. However, we can demonstrate that the main results remain robust when more candidates are included.[9] In fact, the three-candidate scenario represents the most favorable conditions for a polarizing candidate, whereas a larger number of candidates may dilute their influence.

The core message persists: the plurality rule, which dominates real-world elections, creates conditions that encourage polarization—a phenomenon widely recognized as a threat to democratic norms. Addressing this challenge in discussions of electoral reform could be pivotal in combating what has been described as the global "pandemic of authoritarianism" (Sen, 2020).

Funding

The author declares: No funding received for conducting this work.

Appendix A. Proof of Theorem 1

Proof. Recall that $xP^b y$ indicates that candidate x receives a higher Borda score than candidate y. Furthermore, the γ_i-values are subject to the restriction $\gamma_1, \gamma_2 \in \left(\gamma_3, \frac{1}{2}\right)$, as assumed above.

① The non-polarizing candidate B is both the plurality winner and the Condorcet winner. Candidate B also defeats the polarizing candidate A under the Borda rule: $B\,P^b A \iff \frac{1}{2} > \gamma_1 - \frac{1}{2}\gamma_2$. Since $\gamma_1 < \frac{1}{2}$ and $\frac{\gamma_2}{2} > 0$, candidate B is victorious. Additionally, $C\,P^b B$ is not possible because C is Pareto-dominated by B.

② The unique polarizing candidate, A, is both the plurality winner and the Condorcet winner. To show that there exists a domain in profile ② where the PC is defeated under the Borda rule, observe that $B\,P^b A \iff \frac{1}{2}\gamma_1 + \gamma_2 > \gamma_1 + \gamma_3 \iff \gamma_1 > 2 - 4\gamma_2$. For completeness, it is straightforward to verify that candidate C cannot be the Borda winner, since $B\,P^b C \iff \frac{1}{2}\gamma_1 + \gamma_2 > \frac{1}{2}\gamma_2 + \frac{1}{2}\gamma_3 \iff \gamma_1 + \gamma_2 > \gamma_3$, which holds throughout the admissible range of γ_i-values.

③ The unique non-polarizing candidate, C, cannot emerge as winner. Proof: $C\,P^b B \iff 3\gamma_1 + \gamma_2 > 2$. Since $\gamma_1, \gamma_2 < \frac{1}{2}$, the left-hand side of the inequality cannot exceed the right-hand side.

④ We observe a pattern similar to that of the previous profile: $C\,P^b A \iff \gamma_1 + 3\gamma_2 > 2$.

⑤ First, we demonstrate that C cannot become the Borda winner: $B\,P^b C \iff \frac{\gamma_1}{2} + \gamma_2 > \frac{\gamma_2}{2} + \gamma_3 \iff \gamma_3 < \frac{1}{3}$. Since $\gamma_1, \gamma_2 \in \left(\gamma_3, \frac{1}{2}\right)$, this condition is satisfied.
Next, we show that if the non-polarizing candidate B defeats A under the plurality rule ($B\,P^p A$), then B also emerges as Borda winner, but not vice versa ($B\,P^p A \implies B\,P^b A$). To prove this, consider that $B\,P^p A \iff \gamma_2 > \gamma_1$. This condition implies $\gamma_2 > \frac{1}{3}$. Otherwise, $\gamma_3 > \gamma_2 > \gamma_1$, which is outside the admissible range of γ_i-values. For B to defeat A under the Borda rule, we have $B\,P^b A \iff \frac{\gamma_1}{2} + \gamma_2 > \gamma_1 + \frac{\gamma_3}{2} \iff \gamma_2 > \frac{1}{3}$. On the other hand, B can emerge as Borda winner but A wins under plurality rule (for example, $\gamma_1 = 0.48, \gamma_2 = 0.40$).

⑥ By Lemma 1, the unique polarizing candidate, A, cannot win under Borda's rule; instead, the Condorcet winner B prevails. However, B does not emerge as the plurality winner for all cases where $\gamma_1 > \gamma_2$.

⑦ A and B are polarizing candidates. Due to the numerical inferiority of the moderates, the plurality rule selects one of the polarizing candidates. However, under the Borda rule, the Condorcet winner C is elected (cf. Lemma 1).

⑧ Similar to ⑦, C, the only non-polarizing candidate, is the Condorcet winner and (by Lemma 1) emerges victorious under the Borda rule. □

[9] The related material is available from the author on request.

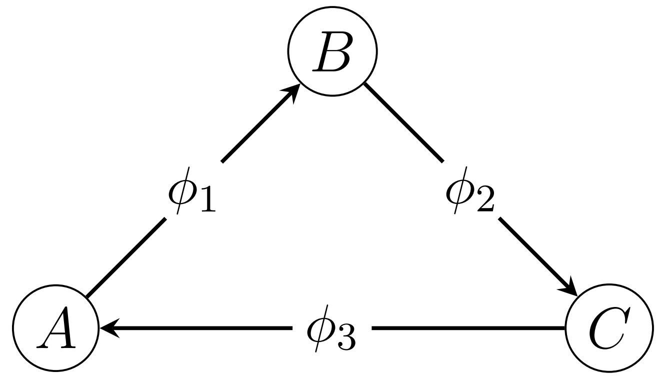

Fig. B.1. Margin graph for profile ⑤.

Appendix B. Proof of Remark 1

Let the margin of candidate A over candidate B in profile ⑤ be defined as the number of voters who rank A above B, minus the number of voters who rank B above A. We denote by $\phi_1 \equiv \gamma_1 + \gamma_3 - \gamma_2 = 1 - 2\cdot\gamma_2$ the margin for the pair (A, B), by $\phi_2 \equiv \gamma_1 + \gamma_2 - \gamma_3 = 2(\gamma_1 + \gamma_2) - 1$ the margin for (B, C), and by $\phi_3 \equiv \gamma_2 + \gamma_3 - \gamma_1 = 1 - 2\cdot\gamma_1$ the margin for (C, A).

The smallest majority principle identifies the weakest majority relation, i.e., the pairwise margin with the smallest absolute value—and removes the corresponding edge in the majority graph (Holliday and Pacuit, 2021). The winner is the candidate to whom no arrow points, but from whom at least one arrow originates.

Fig. B.1 illustrates the margin graph to profile ⑤.

We prove Remark 1 in two steps. First, we show that C never prevails as the winner. Secondly, we show that the plurality winner (A or B) emerges as winner in Condorcet-consistent rules.

Lemma 2 (*Candidate C Cannot Emerge as the Winner*). *The margin graph connecting B and C will never be withdrawn.*

Proof of Lemma (*Lemma 2*).

1. $\phi_2 > \phi_1 \iff 2(\gamma_1 + \gamma_2) - 1 > 1 - 2\gamma_2 \iff 2\gamma_2 > 1 - \gamma_1$. Since $\sum_i \gamma_i = 1$, we have $1 - \gamma_1 = \gamma_2 + \gamma_3$ and hence $\gamma_2 > \gamma_3$. □
2. $\phi_2 > \phi_3 \iff 2(\gamma_1 + \gamma_2) - 1 > 1 - 2\gamma_1 \iff 2\gamma_1 + \gamma_2 > 1$. The last inequality simplifies to $\gamma_1 > \gamma_3$. □

Because $\phi_1 = 1 - 2\gamma_2$ and $\phi_3 = 1 - 2\gamma_1$, it is obvious that $\gamma_1 > \gamma_2 \implies \phi_1 > \phi_3$. In this case, A is the plurality winner and also selected by a Condorcet-consistent voting rule. $\gamma_1 < \gamma_2$ in turn implies $\phi_1 < \phi_3$, such that the plurality winner in this case, B, also emerges as winner by applying a Condorcet-consistent voting rule. □

Data availability

Replication materials for the computational methods described in Section 4.1 are available from the author upon request.

References

Alós-Ferrer, C., Granić, Đ.G., 2012. Two field experiments on approval voting in Germany. Soc. Choice Welf. 39, 171–205. http://dx.doi.org/10.1007/s00355-011-0550-5.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A., 2018. Exposure to opposing views on social media can increase political polarization. Proc. Natl. Acad. Sci. 115, 9216–9221. http://dx.doi.org/10.1073/pnas.1804840115.

Baujard, A., Igersheim, H., Lebon, I., Gavrel, F., Laslier, J.F., 2014. Who's favored by evaluative voting? An experiment conducted during the 2012 french presidential election. Elect. Stud. 34, 131–145. http://dx.doi.org/10.1016/j.electstud.2013.11.003.

Bordignon, M., Nannicini, T., Tabellini, G., 2016. Moderating political extremism: Single round versus runoff elections under plurality rule. Am. Econ. Rev. 106, 2349–2370. http://dx.doi.org/10.1257/aer.20131024.

Bossert, W., Suzumura, K., 2016. The greatest unhappiness of the least number. Soc. Choice Welf. 47, 187–205. http://dx.doi.org/10.1007/s00355-016-0951-6.

Brandt, F., Dong, C., Peters, D., 2024. Condorcet-consistent choice among three candidates. URL: arXiv:2411.19857.

Campante, F.R., Hojman, D.A., 2013. Media and polarization: Evidence from the introduction of broadcast TV in the United States. J. Public Econ. 100, 79–92. http://dx.doi.org/10.1016/j.jpubeco.2013.02.006.

Dasgupta, P., Maskin, E., 2008. On the robustness of majority rule. J. Eur. Econ. Assoc. 6, 949–973. http://dx.doi.org/10.1162/jeea.2008.6.5.949.

Dasgupta, P., Maskin, E., 2020. Strategy-proofness, independence of irrelevant alternatives, and majority rule. Am. Econ. Rev.: Insights 2, 459–474. http://dx.doi.org/10.1257/aeri.20200178.

Gehrlein, W.V., 2005. Probabilities of election outcomes with two parameters: The relative impact of unifying and polarizing candidates. Rev. Econ. Des. 9, 317–336. http://dx.doi.org/10.1007/s10058-005-0132-z.

Gehrlein, W.V., Lepelley, D., Moyouwou, I., 2016. A note on approval voting and electing the Condorcet loser. Math. Soc. Sci. 80, 115–122. http://dx.doi.org/10.1016/j.mathsocsci.2016.02.009.

Grofman, B., Feld, S.L., Fraenkel, J., 2017. Finding the threshold of exclusion for all single seat and multi-seat scoring rules: Illustrated by results for the Borda and Dowdall rules. Math. Soc. Sci. 85, 52–56. http://dx.doi.org/10.1016/j.mathsocsci.2016.11.004.

Holliday, W.H., Pacuit, E., 2021. Axioms for defeat in democratic elections. J. Theor. Politics 33, 475–524. http://dx.doi.org/10.1177/09516298211043236.

Laslier, J.F., Van der Straeten, K., 2008. A live experiment on approval voting. Exp. Econ. 11, 97–105. http://dx.doi.org/10.1007/s10683-006-9149-6.

Lepelley, D., Moyouwou, I., Smaoui, H., 2018. Monotonicity paradoxes in three-candidate elections using scoring elimination rules. Soc. Choice Welf. 50, 1–33. http://dx.doi.org/10.1007/s00355-017-1069-1.

Levy, R., 2021. Social media, news consumption, and polarization: Evidence from a field experiment. Am. Econ. Rev. 111, 831–870. http://dx.doi.org/10.1257/aer.20191777.

Maskin, E., 2025. Borda's rule and Arrow's independence condition. J. Political Econ. 133, 385–420. http://dx.doi.org/10.1086/732892.

Miller, N.R., 2013. The alternative vote and Coombs rule versus first-past-the-post: a social choice analysis of simulated data based on English elections, 1992–2010. Public Choice 158, 399–425. http://dx.doi.org/10.1007/s11127-013-0067-9.

Moulin, H., 1991. Axioms of cooperative decision making. In: Econometric Society Monographs, vol. 15, Cambridge Univ. Press, Cambridge.

Sen, A., 2014. Arrow and the impossibility theorem. In: Maskin, E., Sen, A. (Eds.), The Arrow Impossibility Theorem. Columbia University Press, pp. 29–42.

Sen, A., 2020. Books and freedom. In: MVB (Ed.), Peace Prize of the German Book Trade. MVB, Frankfurt / Main, pp. 65–75.

Young, H.P., 1975. Social choice scoring functions. SIAM J. Appl. Math. 28, 824–838. http://dx.doi.org/10.1137/0128067.