Soc Choice Welf (2017) 48:717–745
DOI 10.1007/s00355-017-1033-0

ORIGINAL PAPER

On the likelihood of single-peaked preferences

Marie-Louise Lackner[1] · **Martin Lackner**[2]

Received: 20 May 2015 / Accepted: 27 January 2017 / Published online: 7 March 2017

Abstract This paper contains an extensive combinatorial analysis of the single-peaked domain restriction and investigates the likelihood that an election is single-peaked. We provide a very general upper bound result for domain restrictions that can be defined by certain forbidden configurations. This upper bound implies that many domain restrictions (including the single-peaked restriction) are very unlikely to appear in a random election chosen according to the Impartial Culture assumption. For single-peaked elections, this upper bound can be refined and complemented by a lower bound that is asymptotically tight. In addition, we provide exact results for elections with few voters or candidates. Moreover, we consider the Pólya urn model and the Mallows model and obtain lower bounds showing that single-peakedness is considerably more likely to appear for certain parameterizations.

1 Introduction

The single-peaked restriction (Black 1948) is an extensively studied domain restriction in social choice theory. An election, i.e., a collection of preferences represented as total orders on a set of candidates, is single-peaked if the candidates can be ordered linearly—on a so-called axis—so that each preference is either strictly increasing along this ordering, or strictly decreasing, or first increasing and then decreasing. See Fig. 1 for examples and Sect. 2 for formal definitions. Intuitively, the axis reflects the

✉ Marie-Louise Lackner
marie-louise.lackner@tuwien.ac.at

Martin Lackner
martin.lackner@cs.ox.ac.uk

[1] Institute of Discrete Mathematics and Geometry, TU Wien, Vienna, Austria

[2] Department of Computer Science, University of Oxford, Oxford, UK

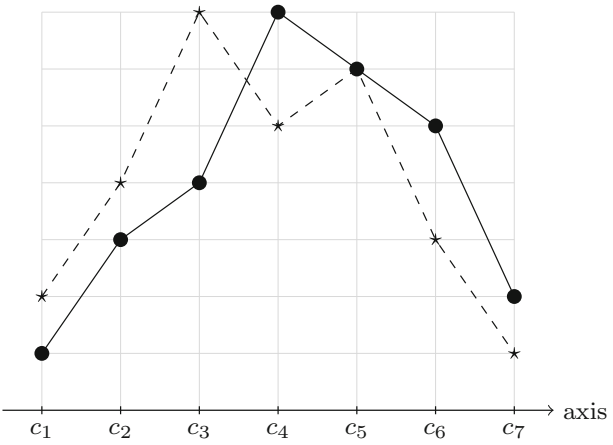

Fig. 1 The vote $V_1 : c_4 > c_5 > c_6 > c_3 > c_2 > c_7 > c_1$, shown as a *solid line*, is single-peaked with respect to the axis $c_1 < c_2 < c_3 < c_4 < c_5 < c_6 < c_7$. The vote $V_2 : c_3 > c_5 > c_4 > c_2 > c_6 > c_1 > c_7$, depicted as a *dashed line*, is not single-peaked with respect to this axis since both c_3 and c_5 form a peak. However, note that both votes are single-peaked with respect to the axis $c_1 < c_2 < c_3 < c_5 < c_4 < c_6 < c_7$

society's ordering of the candidates and voters always prefer candidates that are closer to their ideal candidate over those farther away. In political elections, for example, this axis could reflect the left-right spectrum of the candidates or a natural ordering of political issues such as the maximum income tax.

Single-peaked preferences have several nice properties. First, they guarantee that a Condorcet winner exists and further that the pairwise majority relation is transitive (Inada 1969). Thus single-peaked preferences are a way to escape Arrow's paradox (Arrow 1950). Second, non-manipulable voting rules exist for single-peaked preferences (Moulin 1980) and hence the single-peaked restriction also offers a way to circumvent the Gibbard-Satterthwaite paradox (Gibbard 1973; Satterthwaite 1975). By adopting an algorithmic viewpoint, a third advantage becomes apparent. Restricting the input to single-peaked preferences often allows for faster algorithms for computationally hard voting problems (Brandt et al. 2015; Betzler et al. 2013; Walsh 2007; Faliszewski et al. 2011b).

In this paper we perform an extensive combinatorial analysis of the single-peaked domain. Our aim is to establish results on the likelihood that an election is single-peaked for some axis. To be more precise, we allow the axis to be chosen depending on the preferences and do not assume that it is given together with the election. We consider three probability distributions for elections: the Impartial Culture (IC) assumption in which all total orders are equally likely and are chosen independently, the Pólya urn model which assumes a certain homogeneity among voters and Mallows model in which the probability of a vote depends on the closeness to a given reference vote (with respect to the Kendall-tau distance).

Our main results are the following:

– *Configuration definable restrictions* Many domain restrictions can be characterized by forbidden configurations, in particular the single-peaked domain. We

prove a close connection between configurations and permutations patterns. This novel connection allows us to obtain a very general result (Theorem 8), showing that many domain restrictions characterized by forbidden configurations are very unlikely to appear in a random election chosen according to the Impartial Culture assumption. More precisely, while the total number of elections with n votes and m candidates is equal to $(m!)^n$, the number of elections belonging to such a domain restriction can be bounded by $m! \cdot c^{nm}$ for some constant c.

- *Counting single-peaked elections* We perform a detailed combinatorial analysis of the single-peaked domain by counting the number of single-peaked elections. The number of single-peaked elections immediately yields the corresponding probability with respect to the Impartial Culture (IC) assumption, which is the number of single-peaked elections divided by the total number of elections. We establish an upper bound for the number of single-peaked elections which asymptotically matches our lower bound result (Theorem 11). In addition, we show exact enumeration results for elections with two voters or up to four candidates (Theorem 12). Our results rigorously show that the single-peaked restriction is highly unlikely to appear in random elections chosen according to the IC assumption. This holds even for elections with few votes and candidates (cf. Sect. 8). Most of our results can easily be translated to the Impartial Anonymous Culture (IAC) assumption (Proposition 13).
- *Pólya urn model* In contrast to the IC assumption, single-peaked elections are considerably more likely if elections are chosen according to the Pólya urn model. We provide a lower bound on the corresponding likelihood (Theorem 14) and show that, if a sufficiently strong homogeneity is assumed, the probability of an election with n votes being single-peaked is larger than $1/n$ (Corollary 15).
- *Mallows model* We encounter the most likely occurrence of single-peaked elections under Mallows model. As for the Pólya urn model we establish a lower bound result on the likelihood (Theorem 16). If the dispersion parameter ϕ is sufficiently small , we are able to show that single-peaked elections are likely to appear (Corollary 17 and Table 4).

Related work Computing the likelihood of properties related to voting has been the focus of a large body of research. The most fundamental question in this line of research is the choice of appropriate probability distributions, see the survey of Critchlow et al. (1991). We would like to mention two particular properties of elections that have been studied from a probability theoretic point of view: the likelihood of manipulability and the likelihood of having a Condorcet winner.

An election is manipulable if a voter or a coalition of voters is better off by not voting sincerely but by misrepresenting their true preferences. The Gibbard-Satterthwaite paradox (Gibbard 1973; Satterthwaite 1975) states that every reasonable voting rule for more than two candidates is susceptible to manipulation. However, the Gibbard-Satterthwaite paradox does not offer insight into how likely it is that manipulation is possible. Determining this likelihood both for single manipulators and coalitions of manipulators has been the focus of intensive research. Results have been obtained under a variety of probability distributions: for example under the Impartial Culture assumption (Slinko 2002a, b; Friedgut et al. 2008; Isaksson et al. 2012), the Pólya urn

model (Lepelley and Valognes 2003), the Impartial Anonymous Culture (Favardin et al. 2002; Slinko 2005).

The likelihood that an election has a Condorcet winner or, its converse, the likelihood of the Condorcet paradox has been the focus of many publications (see the survey of Gehrlein (2006) as well as more recent work of Gehrlein et al. (2013, 2015) for more recent research). In particular, we would like to mention that the likelihood of an election with three candidates having a Condorcet winner under the Impartial Anonymous Culture assumption is $\frac{15(n+3)^2}{16(n+2)(n+4)}$ for odd n and $\frac{15(n+2)(n^2+8n+8)}{16(n+1)(n+3)(n+5)}$ for even n (Gehrlein 2002). We will comment on the relation between these result and our results in Sect. 9.

Organization Preliminaries are established in Sect. 2. The results on configuration definable restrictions can be found in Sect. 3, results on counting single-peaked elections in Sect. 4, results on the IAC assumption in Sect. 5, results on the Pólya urn model in Sect. 6 and results on the Mallows model in Sect. 7. In Sect. 8 we provide numerical evaluations of our results and discuss their implications. We conclude the paper in Sect. 9 with directions for future research.

2 Preliminaries

Sets and orders Let S be a finite set. A relation on S is total if for every $a, b \in S$, either the pair (a, b) or (b, a) is contained in the relation. A *total order on* S is a reflexive, antisymmetric, transitive and total relation. Let T be a total order of S. Instead of writing $(a, b) \in T$, we write $a \leq_T b$ or $b \geq_T a$. We write $a <_T b$ or $b >_T a$ to state that $a \leq_T b$ and $a \neq b$. As a short form, we write $T : s_1 s_2 s_3 \dots s_i$ instead of $s_1 >_T s_2 >_T s_3 >_T \dots >_T s_i$ for $s_1, s_2, \dots, s_i$ in S. We write $T(i)$ to denote the i-th largest element with respect to T.

Permutations A permutation π of a finite set S is a bijective function from S to S. We write π^{-1} for the inverse function of π. A permutation of the set $\{1, \dots, m\}$ is called an m-permutation. We shall write an m-permutation π as the sequence of values $\pi(1)\pi(2)\dots\pi(m)$. For example $\pi = 321$ is the permutation with $\pi(1) = 3$, $\pi(2) = 2$ and $\pi(3) = 1$. Every pair (T_1, T_2) of total orders on a set with m elements yields an m-permutation $p(T_1, T_2)$, which is defined as follows: i maps to j if the i-th largest element in T_1 equals the j-th largest element in T_2. For $T_1 : bac$ and $T_2 : cab$ we have $p(T_1, T_2) = 321$. Note that $p(T_1, T_2) = p(T_2, T_1)^{-1}$.

Elections An (n, m)-election $(C, \mathcal{P})$ consists of a size-m set C and an n-tuple $\mathcal{P} = (V_1, \dots, V_n)$ of total orders on C. The total orders $V_1, \dots, V_n$ represent votes or preferences. We write $V \in \mathcal{P}$ to denote that there exists an index $i \in [n]$ such that $V = V_i$. Given a vote $V_i \in \mathcal{P}$, the intuitive meaning of $V_i : c_j c_k$ is that the i-th voter prefers candidate c_j to candidate c_k.

We assume that candidate sets are chosen from a fixed, infinite set. When counting elections we do not care about the specific names these candidates have. That means when we count elections we fix the candidate set to $\{c_1, c_2, \dots, c_m\}$. Note that the

number of (n, m)-elections is $(m!)^n$. Throughout the paper we only consider (n, m)-elections with $n \geq 2$ and $m \geq 2$.

Probability distributions over elections We consider four probability distributions in this paper. The first and simplest is the Impartial Culture (IC) which assumes that in an election all votes, i.e., total orders of candidates, are equally likely and are chosen independently. Thus, the IC assumption can be seen as the uniform distribution over total orders on candidates. The results in our paper concerning IC do not state probabilities but rather count the number of elections. If, e.g., the number of single-peaked (n, m)-elections is $a(n, m, SP)$, then the probability under the IC assumption that an (n, m)-election is single-peaked is $\frac{a(n,m,SP)}{(m!)^n}$. It is important to note that in our paper elections contain an ordered list of votes. Thus, we distinguish elections that consist of the same votes but these votes appear in a different order. This is in contrast to the Impartial Anonymous Culture (IAC) assumption, in which elections contain a multiset of votes and thus elections are not ordered. The IAC assumption is briefly considered in Sect. 5.

In addition to the IC assumption, we consider the Pólya urn model and the Mallows model. Both distributions are generalizations of the IC assumption and generate more structured elections. We are going to define the Pólya urn and the Mallows model in Sects. 6 and 7, respectively.

Single-peaked preferences The single-peaked restriction assumes that the candidates can be ordered linearly on a so-called axis and voters prefer candidates close to their ideal point to candidates that are further away.

Definition 1 Let $(C, \mathcal{P})$ be an election and A a total order of C. A vote V on C *contains a valley with respect to* A *on the candidates* $c_1, c_2, c_3 \in C$ if $A : c_1\, c_2\, c_3$ and V ranks c_2 below c_1 and c_3. The election $(C, \mathcal{P})$ is *single-peaked with respect to* A if for every $V \in \mathcal{P}$ and for all candidates $c_1, c_2, c_3 \in C$, V does not contain a valley with respect to A on c_1, c_2, c_3. We then call the total order A the axis. The election $(C, \mathcal{P})$ is *single-peaked* if there exists a total order A of C such that $(C, \mathcal{P})$ is single-peaked with respect to A.

Remark 1 Given an axis on m candidates, there are 2^{m-1} votes that are single-peaked with respect to this axis (Escoffier et al. 2008). This can be seen as follows: The last ranked candidate has to be one of the two outermost candidates on the axis and hence there are two possibilities. Once we have picked this last candidate, we can iterate the argument for the next lowest ranked candidate, where we again have two possibilities. Thus, for all positions in the total order (except for the top ranked candidate), there are two candidates to choose from—which yields 2^{m-1} possibilities in total.

3 A general result based on permutation patterns

Before we study the single-peaked domain in detail, we prove a general result that is applicable to a large class of domain restrictions including the single-peaked domain.

To precisely define this class of domain restrictions, we require the notion of configuration definability.

3.1 Configuration definable domain restrictions

Single-peaked elections may also be defined in the following way:

Theorem 2 (Ballester and Haeringer 2011) *An (n, m)-election $(C, \mathcal{P})$ is single-peaked if and only if there do not exist candidates $a, b, c, d \in C$ and indices $i, j \in [n]$ such that*

- $V_i : abc$, $V_i : db$,
- $V_j : cba$ *and* $V_j : db$ *holds*

and there do not exist candidates $a, b, c \in C$ and indices $i, j, k \in [n]$ such that

- $V_i : ba, ca$ *(i.e., a is ranked below b and c),*
- $V_j : ab, cb$ *and*
- $V_k : ac, bc$ *holds.*

Note that this theorem defines single-peakedness without referring to an axis. Indeed, single-peakedness is now defined as a local property in the sense that certain *configurations* must not be contained in the election. Similar definitions have also been found for the single-crossing (Bredereck et al. 2013b) and group-separable (Ballester and Haeringer 2011) domain. For other domains such as value-, worst-, medium and best-restricted preferences, a characterization via configurations follows immediately from the original definitions (Sen 1966; Sen and Pattanaik 1970). Let us now exactly define what it means for a domain restriction to be *configuration definable*.

Definition 2 An (l, k)-*configuration* $(S, \mathcal{T})$ consists of a finite set S of cardinality k and a tuple $\mathcal{T} = (T_1, \ldots, T_l)$, where $T_1, \ldots, T_l$ are total orders on S. An election $(C, \mathcal{P})$ *contains* configuration $\mathcal{C}$ if there exist an injective function f from $[l]$ into $[n]$ and an injective function g from S into C such that, for any $x, y \in S$ and $i \in [l]$, it holds that $T_i : xy$ implies $V_{f(i)} : g(x)g(y)$.

We use $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$ as a shorthand notation to denote that the election $(C, \mathcal{P})$ contains the configuration $(S, \mathcal{T})$. An election $(C, \mathcal{P})$ *avoids* a configuration $(S, \mathcal{T})$ if $(C, \mathcal{P})$ does not contain $(S, \mathcal{T})$. In such a case we say that $(C, \mathcal{P})$ is $(S, \mathcal{T})$-*restricted*. If the set S is clear from the context, we omit it and just use $\mathcal{T}$ to describe a configuration.

Example 3 Let us consider an election $(C, \mathcal{P})$ with $C = \{u, v, w, x, y\}$ and $\mathcal{P} = (uvwxy, wyvux, yuxwv)$ and a configuration $(S, \mathcal{T})$ with $S = \{a, b, c, d\}$ and $\mathcal{T} = (dabc, cdba)$. Election $(C, \mathcal{P})$ contains the configuration $(S, \mathcal{T})$ as witnessed by the functions $f : \{1 \mapsto 1, 2 \mapsto 3\}$ and $g : \{a \mapsto v, b \mapsto x, c \mapsto y, d \mapsto u\}$. In Fig. 2, the functions f and g are depicted graphically.

By considering all linearizations of the partial orders appearing in Theorem 2 we can now restate it as follows.

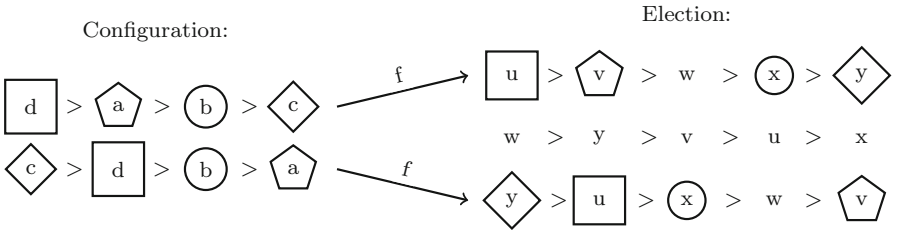

Fig. 2 The configuration on the *left-hand* side is contained in the election on the *right-hand* side as witnessed by $f : \{1 \mapsto 1, 2 \mapsto 3\}$ and $g : \{a \mapsto v, b \mapsto x, c \mapsto y, d \mapsto u\}$

Theorem 4 (Ballester and Haeringer 2011) *An election is single-peaked if and only if it avoids*

- *the following* $(2, 4)$*-configurations:* $(dabc, dcba)$, $(adbc, dcba)$, $(dabc, cdba)$ *and* $(adbc, cdba)$
- *as well as the following* $(3, 3)$*-configurations:* (bca, acb, abc), (cba, acb, abc), (bca, cab, abc), (cba, cab, abc), (bca, acb, bac), (cba, acb, bac), (bca, cab, bac), (cba, cab, bac).

The first four configurations correspond to the first condition in Theorem 2, the remaining eight correspond to the second condition.

Definition 3 Let Γ be a set of configurations. A set of elections Π is *defined by* Γ if Π consists exactly of those elections that avoid all configurations in Γ. We call Π *configuration definable* if there exists a set of configurations Γ which defines Π. If Π is definable by a finite set of configurations, it is called *finitely configuration definable*.

By Theorem 4 we know that the set of all single-peaked elections is finitely configuration definable. This is also true for the set of group-separable elections (Ballester and Haeringer 2011) and for the set of single-crossing elections (Bredereck et al. 2013b).

We are now going to characterize which sets of elections are configuration definable. In the following definition, for two elections $(C, \mathcal{P})$ and $(C', \mathcal{P}')$, we write $(C', \mathcal{P}') \sqsubseteq (C, \mathcal{P})$ if $(C', \mathcal{P}')$, considered as a configuration, is contained in $(C, \mathcal{P})$. Since every election can be seen as a configuration, the configuration containment relation immediately translates to election containment.

Definition 4 A set of elections Π is *hereditary* if for every election $(C', \mathcal{P}')$ it holds that if there exists an election $(C, \mathcal{P}) \in \Pi$ with $(C', \mathcal{P}') \sqsubseteq (C, \mathcal{P})$, then $(C', \mathcal{P}') \in \Pi$.

Proposition 5 *A set of elections is configuration definable if and only if it is hereditary.*

Proof Let a set of elections Π be defined by a set of configurations Γ and $(C, \mathcal{P}) \in \Pi$. Let $(C', \mathcal{P}') \sqsubseteq (C, \mathcal{P})$. Since $(C, \mathcal{P}) \in \Pi$, $(C, \mathcal{P})$ avoids all configurations in Γ. Due to $(C', \mathcal{P}') \sqsubseteq (C, \mathcal{P})$, also $(C', \mathcal{P}')$ avoids all configurations in Γ and is therefore contained in Π.

For the other direction, let Π be a hereditary set of elections. We define Π^c to be the *complement of* Π, i.e., an election is contained in Π^c if it is not contained in Π.

We claim that Π^c, considered as a set of configurations, defines Π. If $(C, \mathcal{P}) \in \Pi$, it avoids all $(C', \mathcal{P}') \in \Pi^c$ since if there existed a $(C', \mathcal{P}') \in \Pi^c$ with $(C', \mathcal{P}') \sqsubseteq (C, \mathcal{P})$, this would imply that $(C', \mathcal{P}') \in \Pi$. It remains to show that if an election $(C'', \mathcal{P}'')$ avoids all $(C', \mathcal{P}') \in \Pi^c$, then $(C'', \mathcal{P}'') \in \Pi$. This follows from noting that $(C'', \mathcal{P}'')$ avoiding all $(C', \mathcal{P}') \in \Pi^c$ implies $(C'', \mathcal{P}'') \notin \Pi^c$, which in turn implies $(C'', \mathcal{P}'') \in \Pi$. □

As a consequence of Proposition 5, we know that 2D single-peaked (Barberà et al. 1993) and 1D Euclidean elections (Coombs 1964; Knoblauch 2010) are configuration definable. However, Proposition 5 does not help to answer whether these restrictions are *finitely* configuration definable. For the 1D Euclidean domain it is even known that it is not finitely configuration definable (Chen et al. 2015). Finite configuration definability has been crucial for establishing algorithmic results (Bredereck et al. 2013a; Elkind and Lackner 2014).

A natural example of a meaningful restriction that is not configuration definable is the set of all elections that have a Condorcet winner. The property of having a Condorcet winner is not hereditary and thus cannot be defined by configurations. Another example is the "single-peaked on a tree" restriction (Demange 1982).

3.2 The connection to permutation patterns

In this section, we establish a strong link between the concept of configuration containment and the concept of pattern containment in permutations. Pattern containment in permutations is defined as follows.

Definition 5 A k-permutation π is *contained as a pattern* in an m-permutation τ if there is a subsequence of τ that is order-isomorphic to π. In other words, π is contained in τ, if there is a strictly increasing map $\mu : \{1, \dots, k\} \to \{1, \dots, m\}$ so that the sequence $\mu(\pi) = \big(\mu(\pi(1)), \mu(\pi(2)), \dots, \mu(\pi(k))\big)$ is a subsequence of τ. This map μ is called a *matching* of π into τ. If there is no such matching, τ *avoids the pattern* π.

For example, the pattern $\pi = 132$ is contained in $\tau = 32514$ since the subsequence 254 of τ is order-isomorphic to π. However, the pattern 123 is avoided by τ. Note that τ contains π if and only if τ^{-1} contains π^{-1}.

We are going to prove two lemmas. The first lemma (Lemma 6) states that every permutation pattern matching query can naturally be translated into a configuration containment query. The second lemma (Lemma 7) states that for $(2, k)$-configurations, a configuration containment query can naturally be translated in a permutation pattern query.

Lemma 6 *Let π be a k-permutation and τ an m-permutation. We define the corresponding configuration and election as follows: Let $(C, \mathcal{P})$ be a $(3, m)$-election with $C = \{c_1, \dots, c_m\}$, $\mathcal{P} = (V_1, V_2, V_3)$ and*

$$V_1 : c_1 c_2 \cdots c_m \qquad V_2 : c_1 c_2 \cdots c_m \qquad V_3 : c_{\tau(1)} c_{\tau(2)} \cdots c_{\tau(m)}.$$

Furthermore, let $(S, \mathcal{T})$ *be a* $(3, k)$*-configuration with* $S = \{x_1, \ldots, x_k\}$, $\mathcal{T} = (T_1, T_2, T_3)$ *and*

$$T_1 : x_1 x_2 \cdots x_k \qquad T_2 : x_1 x_2 \cdots x_k \qquad T_3 : x_{\pi(1)} x_{\pi(2)} \cdots x_{\pi(k)}.$$

Then $(C, \mathcal{P})$ *contains* $(S, \mathcal{T})$ *if and only if* τ *contains* π.

Proof Assume that we have a matching μ from π into τ. We have to find an injective function f from $\{1, 2, 3\}$ into $\{1, 2, 3\}$ and an injective function g from S into C such that, for any $x, y \in S$ and $i \in \{1, 2, 3\}$, it holds that $T_i : xy$ implies $V_{f(i)} : g(x)g(y)$. Let f be the function $\{1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 3\}$ and $g = \mu$. It holds for $x_i, x_j \in S$ that $T_1 : x_i\, x_j$ if and only if $V_1 : c_{\mu(i)}\, c_{\mu(j)}$ since μ is monotone. The same holds for T_2 and V_2. For T_3 and V_3 observe that $T_3 : x_i\, x_j$ implies $V_3 : c_{\mu(i)}\, c_{\mu(j)}$ since μ is a matching. Thus, the election fulfils $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$.

For the other direction, assume that $(C, \mathcal{P})$ contains $(S, \mathcal{T})$. Consequently, there exists an injective function f from $\{1, 2, 3\}$ into $\{1, 2, 3\}$ and an injective function g from S into C such that, for any $x, y \in S$ and $i \in \{1, 2, 3\}$, it holds that $T_i : xy$ implies $V_{f(i)} : g(x)g(y)$. First, we claim that $f(3) = 3$. Observe that f has to map T_1 and T_2 to identical total orders. Thus, unless $V_1 = V_2 = V_3$, $f(3) = 3$. In the case that $V_1 = V_2 = V_3$, we can assume without loss of generality that $f(3) = 3$. We will construct a function μ and show that μ is a matching from π into τ. Let us define $\mu(i) = j$ if $g(x_i) = c_j$. Observe that μ is strictly increasing since for $i < j$, $V_1 : c_{g(i)}\, c_{g(j)}$ and $V_1 : c_1 c_2 \cdots c_m$. In addition, $\mu(\pi) = \big(\mu(\pi(1)), \mu(\pi(2)), \ldots, \mu(\pi(k))\big)$ is a subsequence of τ since, by definition of T_3 and V_3 and the fact that $f(3) = 3$, $\big(g(x_{\pi(1)}), g(x_{\pi(2)}), \ldots, g(x_{\pi(k)})\big)$ is a subsequence of $\big(c_{\tau(1)}, c_{\tau(2)}, \ldots, c_{\tau(m)}\big)$. □

Next, we will prove the second lemma, which is essential for the main theorem of this section (Theorem 8). As of now, we shall denote by $S_m(\pi_1, \ldots, \pi_l)$ the cardinality of the set of m-permutations that avoid the patterns $\pi_1, \ldots, \pi_l$.

Lemma 7 *Let* $(S, \mathcal{T})$ *be a* $(2, k)$*-configuration with* $\mathcal{T} = (T_1, T_2)$. *Furthermore, let* V_1 *be a total order on the candidate set* $C = \{c_1, \ldots, c_m\}$. *Then the number of total orders* V_2 *such that the election* $(C, \mathcal{P})$ *with* $\mathcal{P} = (V_1, V_2)$ *avoids* $(S, \mathcal{T})$ *is equal to* $S_m(\pi, \pi^{-1})$, *where* $\pi = p(T_1, T_2)$.

Proof Let us start by proving the following statement: The configuration $(S, \mathcal{T})$ is contained in an election $(C, \mathcal{P})$ with $\mathcal{P} = (V_1, V_2)$ if and only if the permutation π or the permutation π^{-1} is contained in $p(V_1, V_2)$. In order to alleviate notation, we will assume in the following that $C = \{1, 2, \ldots m\}$ and $S = \{1, 2, \ldots k\}$.

Let μ be a matching witnessing that π is contained in $p(V_1, V_2)$. We can assume without loss of generality that $T_1 : 12 \ldots k$ and $V_1 : 12 \ldots m$. Then the functions $f = \{1 \mapsto 1, 2 \mapsto 2\}$ and $g = \mu$ show that $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$ (cf. Definition 2). If π^{-1} is contained in $p(V_1, V_2)$ as witnessed by a matching μ, then the functions $f = \{1 \mapsto 2, 2 \mapsto 1\}$ and $g = \mu$ show that $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$.

For the other direction, let $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$. Without loss of generality we assume that $T_1 : 12 \ldots k$. Note that renaming C does not change whether $(S, \mathcal{T}) \sqsubseteq (C, \mathcal{P})$. Thus, it is safe to rename the candidates according to the f function: If $f = \{1 \mapsto$

$1, 2 \mapsto 2\}$, let $V_1 : 12 \dots n$. Since $f(1) = 1$, g is monotonic. It is easy to verify that g is a matching from π into $p(V_1, V_2)$. If $f = \{1 \mapsto 2, 2 \mapsto 1\}$, let $V_2 : 12 \dots n$. Now, g is a matching from π into $p(V_2, V_1) = (p(V_1, V_2))^{-1}$. This is equivalent to g being a matching from π^{-1} into $p(V_1, V_2)$.

It follows that $(C, \mathcal{P})$ avoids the configuration $(S, \mathcal{T})$ if and only if the permutation $p(V_1, V_2)$ avoids both the patterns π and π^{-1}. Moreover, for the fixed total order V_1 and a fixed m-permutation τ, there is a single total order V_2 such that $p(V_1, V_2) = \tau$. Thus the number of votes V_2 such that $p(V_1, V_2)$ avoids π and π^{-1} (and equivalently the number of votes V_2 such that $(C, \mathcal{P})$ avoids $(S, \mathcal{T})$) is equal to $S_m(\pi, \pi^{-1})$, the number of m-permutations avoiding π and π^{-1}. □

From this lemma follows the main theorem of this section that is applicable to any set of configurations that contains at least one configuration of cardinality two.

3.3 Elections that avoid a $(2, k)$-configuration

With the help of Lemma 7, we are able to establish the following result.

Theorem 8 *Let $a(n, m, \Gamma)$ be the number of (n, m)-elections avoiding a set of configurations Γ. Let $k \geq 2$. If a set of configurations Γ contains a $(2, k)$-configuration, then it holds for all $n, m \in \mathbb{N}$ that*

$$a(n, m, \Gamma) \leq m! \cdot c_k^{(n-1)m},$$

where c_k is a constant depending only on k.

This result shows that forbidding any $(2, k)$-configuration is a very strong restriction. Indeed, $m! \cdot c_k^{(n-1)m}$ is very small compared to the total number of (n, m)-elections which is $(m!)^n$. This result allows us to bound the number of single-peaked and group-separable elections. However, let us prove this result first before we explore its consequences.

In order to prove this result we make use of the link between configuration avoiding elections and pattern avoiding permutations established in Lemma 7 and profit from a very strong result within the theory of pattern avoidance in permutations, the Marcus-Tardos theorem (former Stanley-Wilf conjecture).

Proof We are going to provide an upper bound on the number of (n, m)-elections avoiding a $(2, k)$-configuration $(S, \mathcal{T})$ with $\mathcal{T} = (T_1, T_2)$. Let us start by choosing the first vote V_1 of the election at random. For this there are $m!$ possibilities. When choosing the remaining $(n - 1)$ votes $V_2, \dots, V_n$, we have to make sure that no selection of two votes contains the forbidden configuration $(S, \mathcal{T})$. We relax this condition and only demand that for none of the pairs (V_1, V_i) with $i \neq 1$, the election $(C, (V_1, V_i))$ contains the forbidden configuration. Hereby we obtain an upper bound for $a(n, m, \{(S, \mathcal{T})\})$. Now Lemma 7 tells us that there are—under this relaxed condition—$S_m(\pi, \pi^{-1})$ choices for every V_i where $\pi = p(T_1, T_2)$. Thus we have the following upper bound:

$$a(n, m, \{(S, \mathcal{T})\}) \leq m! S_m(\pi, \pi^{-1})^{n-1} \leq m! S_m(\pi)^{n-1}, \tag{1}$$

where the second inequality follows since all permutations avoiding both π and π^{-1} clearly avoid π.

Now we apply the famous Marcus-Tardos theorem (Marcus and Tardos 2004): For every permutation π of length k there exists a constant c_k such that for all positive integers m we have $S_m(\pi) \leq {c_k}^m$. Putting this together with Equation (1) and noting that $a(n, m, \{(S, \mathcal{T})\})$ is an upper bound for $a(n, m, \Gamma)$, we obtain the desired upper bound. □

The proof of the Marcus-Tardos theorem provides an explicit exponential formula for the constants c_k. Indeed, it holds that

$$S_m(\pi) \leq \left(15^{2k^4\binom{k^2}{k}}\right)^m.$$

These constants are however far from being optimal and there is an ongoing effort to find minimal values for c_k with fixed k. In particular it has been shown that $c_2 = 1$, $c_3 = 4$ (Simion and Schmidt 1985) and $c_4 \leq 13.738$ (Bóna 2014).

Let us discuss the implications of this theorem. It is applicable to all (not necessarily finite) configuration definable domain restrictions that contain a configuration of cardinality two. In particular, we obtain the following upper bounds for single-peaked and group-separable elections.

Corollary 9 *Let $a(n, m, \Gamma_{sp})$ denote the number of single-peaked (n, m)-elections. For $n, m \geq 2$ it holds that $a(n, m, \Gamma_{sp}) \leq m! \cdot 4^{(m-1)(n-1)}$.*

Proof We know from Theorem 2 that the single-peaked domain avoids the $(2, 4)$-configurations $(dabc, dcba)$, $(adbc, dcba)$, $(dabc, cdba)$ and $(adbc, cdba)$. We can use Eq. (1) in the proof of Theorem 8 to bound $a(n, m, \Gamma_{sp})$. For this, we have to compute the permutations and their inverses corresponding to the four configurations. We obtain the permutations $\pi_1 = p(dabc, dcba) = 1432$, $\pi_2 = p(adbc, dcba) = 4132$, $\pi_3 = p(dabc, cdba) = 2431$ and $\pi_4 = p(adbc, cdba) = 4231$. Their inverses are $\pi_1^{-1} = \pi_1$, $\pi_2^{-1} = \pi_3$, $\pi_3^{-1} = \pi_2$ and $\pi_4^{-1} = \pi_4$. Hence it holds that the number of (n, m)-elections that avoid these four configurations is bounded by $m! \cdot S_m(\pi_1, \pi_2, \pi_3, \pi_4)^{n-1}$. The enumeration problem for this permutation class has been solved by Guibert (1995) in his PhD thesis with the help of the method of generating trees. A more direct and combinatorial approach to this permutation class can be found in the first author's PhD thesis (Bruner 2015). It holds that $S_m(\pi_1, \pi_2, \pi_3, \pi_4) = \binom{2m-2}{m-1}$, which, in turn, is bounded by 4^{m-1}. □

This upper bound also holds for the 1D Euclidean domain (Coombs 1964; Knoblauch 2010), since this domain is a subset of the single-peaked domain. In the next section, we will see that the growth rate of $a(n, m, \Gamma_{sp})$ is indeed of the form $m! \cdot c^{(m-1)(n-1)}$ for some constant c. However, the constant found in Corollary 9 is not optimal as we will see by providing a better bound for the single-peaked restriction that is even asymptotically optimal.

As another corollary of Theorem 8, we prove a bound on the number of group-separable elections. An election is group separable if for every subset of candidates

C' there exists a partition C_1, C_2 of C' such that in every vote either all candidates in C_1 are preferred to all candidates in C_2 or vice versa. Ballester and Haeringer (2011) showed that the group-separable domain is finitely configuration definable. In particular, this domain avoids the configuration $(abcd, bdac)$. Therefore, Theorem 8 is applicable.

Corollary 10 *Let $a(n, m, \Gamma_{gs})$ denote the number of group-separable (n, m)-elections. For $n, m \geq 2$ it holds that $a(n, m, \Gamma_{gs}) \leq m! \cdot (3 + 2\sqrt{2})^{m(n-1)}$.*

Proof The proof is similar to the one of Corollary 9. We use Equation (1) in the proof of Theorem 8 to bound $a(n, m, \Gamma_{gs})$, i.e., $a(n, m, \Gamma_{gs}) \leq m! \cdot S_m(\pi, \pi^{-1})^{n-1}$, where $\pi = p(abcd, bdac) = 3142$ and $\pi^{-1} = 2413$. Permutations avoiding these two patterns are known under the name of *separable permutations*. It is known that separable permutations are counted by the large Schröder numbers (OEIS A006318) and that $S_m(\pi, \pi^{-1}) \leq (3 + 2\sqrt{2})^m$ (West 1995). □

4 Counting results and the Impartial Culture assumption

As in the previous section, let $a(n, m, \Gamma_{\text{sp}})$ denote the number of single-peaked elections. In this section, we prove a lower and upper bound on $a(n, m, \Gamma_{\text{sp}})$. These two bounds are asymptotically optimal, i.e., the lower bound converges to the upper bound for every fixed m and $n \to \infty$. In addition, we prove exact enumeration results for $a(2, m, \Gamma_{\text{sp}})$, $a(n, 3, \Gamma_{\text{sp}})$ and $a(n, 4, \Gamma_{\text{sp}})$.

Our results immediately imply bounds on the probability that an (n, m)-election is single-peaked assuming that elections are drawn uniformly at random, i.e., according to the Impartial Culture assumption. The probability is simply $a(n, m, \Gamma_{\text{sp}})/(m!)^n$.

Theorem 11 *It holds that*

$$\frac{m!}{2} \cdot 2^{(m-1)\cdot n} \cdot (1 - \epsilon(n, m)) \leq a(n, m, \Gamma_{sp}) \leq \frac{m!}{2} \cdot 2^{(m-1)\cdot n},$$

where $\epsilon(n, m) \to 0$ for every fixed m and $n \to \infty$.

Proof First observe that an election is single-peaked with respect to an axis if and only if it is single-peaked with respect to its reverse, i.e., the axis read from right to left. Thus the total number of axes on m candidates that need to be considered is $m!/2$. Second, recall that the number of votes that are single-peaked with respect to a given axis is 2^{m-1} (cf. Remark 1).

Now we have gathered all facts necessary for the upper bound. For every one of the $m!/2$ axes considered, select an ordered set of votes from the 2^{m-1} votes that are single-peaked with respect to this axis. There are exactly $2^{(m-1)\cdot n}$ such possibilities, which yields the upper bound. Since an election may be single-peaked with respect to more than two axes, this number is only an upper bound for $a(n, m, \Gamma_{\text{sp}})$.

Let us turn to the lower bound. Given a vote V, there are only two axes with respect to which both V and its reverse $\bar{V}$ are single-peaked, namely the total orders V and $\bar{V}$ themselves. Thus the presence of the votes V and $\bar{V}$ in an election forces the axis to

be equal to either V or $\bar{V}$. If we fix a vote V, the number of single-peaked elections containing both V and $\bar{V}$ can be determined easily using the inclusion-exclusion principle. Indeed, since $2^{(m-1)\cdot n}$ is the number of single-peaked elections for the axis V, $\left(2^{m-1}-1\right)^n$ is the number of single-peaked elections for the axis V that do not contain the vote V (analogous for single-peaked elections that do not contain $\bar{V}$) and $\left(2^{m-1}-2\right)^n$ is the number of single-peaked elections for the axis V that neither contain V nor $\bar{V}$. Thus, the number of single-peaked elections containing both V and $\bar{V}$ for some fixed vote V is equal to:

$$2^{(m-1)\cdot n} - 2\cdot\left(2^{m-1}-1\right)^n + \left(2^{m-1}-2\right)^n.$$

Multiplying this by the number of possibilities for the vote V leads to the lower bound for $a(n, m, \Gamma_{sp})$:

$$\begin{aligned}&\frac{m!}{2}\cdot\left(2^{(m-1)\cdot n} + \left(2^{m-1}-2\right)^n - 2\cdot\left(2^{m-1}-1\right)^n\right)\\ &= \frac{m!}{2}\cdot 2^{(m-1)\cdot n}\cdot\left(1-\epsilon(n,m)\right),\\ &\text{where } \epsilon(n,m) = \frac{2\cdot(2^{m-1}-1)^n - \left(2^{m-1}-2\right)^n}{2^{(m-1)n}}.\end{aligned}$$

Since $\epsilon(n,m) \le 2\cdot\left(\frac{2^{m-1}-1}{2^{m-1}}\right)^n$, $\epsilon(n,m)$ tends to 0 for every fixed m and $n\to\infty$. Clearly, not all single-peaked elections contain a pair of votes where one is the reverse of the other. Thus this number is indeed only a lower bound. □

In the next theorem we prove exact enumeration formulæ for $a(n, m, \Gamma_{\mathrm{sp}})$ for $n = 2$, $m = 3$ and $m = 4$. Note that for $m \le 2$ and for $n = 1$ all (n, m)-elections are single-peaked. For $n > 2$ and for $m > 4$ we have not been able to find exact enumeration formulas.

Theorem 12 *It holds that*

(i.) $a(2, m, \Gamma_{sp}) = m!\cdot\binom{2m-2}{m-1}$ *for* $m \ge 1$,
(ii.) $a(n, 3, \Gamma_{sp}) = 6\cdot 2^{n-1}\left(2^n-1\right)$ *and*
(iii.) $a(n, 4, \Gamma_{sp}) = 24\cdot 4^{n-1}\cdot\left(2^{n+1}-3\right)$.

Proof (i.) $a(2, m, \Gamma_{\mathrm{sp}}) = m!\cdot\binom{2m-2}{m-1}$: This follows from Lemma 7. We choose the first vote arbitrarily ($m!$ possibilities). The second vote has to be chosen in such a way that all configurations that characterize single-peakedness are avoided. Since we consider only elections with two votes, the relevant configurations are $(dabc, dcba)$, $(adbc, dcba)$, $(dabc, cdba)$ and $(adbc, cdba)$ (Theorem 4). We obtain the permutations $\pi_1 = p(dabc, dcba) = 1432$, $\pi_2 = p(adbc, dcba) = 4132$, $\pi_3 = p(dabc, cdba) = 2431$ and $\pi_4 = p(adbc, cdba) = 4231$. Their inverses are $\pi_1^{-1} = \pi_1$, $\pi_2^{-1} = \pi_3$, $\pi_3^{-1} = \pi_2$ and $\pi_4^{-1} = \pi_4$. Thus, the number of $a(2, m, \Gamma_{\mathrm{sp}}) = S_m(\pi_1, \pi_2, \pi_3, \pi_4)$, which is equal to $\binom{2m-2}{m-1}$, as shown by Guibert (1995) and Bruner (2015).

(ii.) $a(n, 3, \Gamma_{\text{sp}}) = m! \cdot 2^{n-1}(2^n - 1)$: We consider all elections with three candidates. There are $m!$ many possibilities for the first vote V_1. Without loss of generality, let us consider only the vote $V_1 : abc$. Since we have only three candidates, single-peakedness boils down to having at most two last ranked candidates (cf. Theorem 2). Due to our assumption that $V_1 : abc$, we distinguish three cases: elections in which the votes rank either a or c last, elections in which the votes rank either b or c last and elections in which all votes rank c last. The number of elections in which the votes rank either a or c last can be determined as follows: every vote can either be abc, bac, cba or bca. Hence, there are 4^{n-1} possibilities for elections in which the votes rank either a or c last and where $V_1 : abc$ holds. By the same argument, the number of elections in which the votes rank either b or c last is 4^{n-1} as well. The number of elections where c is always ranked last is 2^{n-1}. We obtain a total number of single-peaked elections with a fixed first vote of $4^{n-1} + 4^{n-1} - 2^{n-1} = 2^{n-1} \cdot (2 \cdot 2^{n-1} - 1)$. Given that 6 options for the first vote exist, we obtain the stated enumeration result.

(iii.) $a(n, 4, \Gamma_{\text{sp}}) = m! \cdot 4^{n-1} \cdot \left(2^{n+1} - 3\right)$: As in the previous proof, we fix $V_1 : abcd$. This vote V_1 already rules out some possible axes. Indeed, only eight axes are single-peaked axes for V_1, namely $A_1 : abcd$, $A_2 : bacd$, $A_3 : cabd$, $A_4 : cbad$, and their reverses. Since the reverse of an axis permits the same single-peaked votes, we have to consider only A_1, A_2, A_3, A_4. For $1 \leq i \leq 4$, let W_i denote the set of four-candidate votes that are single-peaked with respect to axis A_i. We count the number of single-peaked elections with four candidates by using the inclusion-exclusion principle, i.e.,

$$\begin{aligned} a(n, 4, \Gamma_{\text{sp}}) = m! \cdot (&|W_1| + |W_2| + |W_3| + |W_4| \\ &- |W_1 \cap W_2| - |W_1 \cap W_3| - |W_1 \cap W_4| \\ &- |W_2 \cap W_3| - |W_2 \cap W_4| - |W_3 \cap W_4| \\ &+ |W_1 \cap W_2 \cap W_3| + |W_1 \cap W_2 \cap W_4| \\ &+ |W_1 \cap W_3 \cap W_4| + |W_2 \cap W_3 \cap W_4| \\ &- |W_1 \cap W_2 \cap W_3 \cap W_4|) \end{aligned}$$

It is easy to verify that $W_1 \cap W_2 = W_1 \cap W_4 = W_2 \cap W_3 = \{abcd, bacd\}$. Consequently, all intersections of three or four sets consist also of these two votes. The remaining intersections look as follows:

$$\begin{aligned} W_1 \cap W_3 &= \{abcd, bacd, cbad, bcad\}, \\ W_2 \cap W_4 &= \{abcd, bacd, cabd, acbd\}, \\ W_3 \cap W_4 &= \{abcd, bacd, badc, abdc\}. \end{aligned}$$

The number of votes single-peaked with respect to one axis is 2^{m-1} (see Remark 1), i.e., in our case 8. We obtain

$$\begin{aligned} a(n, 4, \Gamma_{\text{sp}}) &= 4! \cdot \left(4 \cdot 8^{n-1} - 3 \cdot 2^{n-1} - 3 \cdot 4^{n-1} + 4 \cdot 2^{n-1} - 2^{n-1}\right) \\ &= 24 \cdot \left(4 \cdot 8^{n-1} - 3 \cdot 4^{n-1}\right). \end{aligned}$$

□

5 The Impartial Anonymous Culture assumption

The counting results from the previous section on the Impartial Culture (IC) assumption can easily be adapted to the Impartial Anonymous Culture (IAC) assumption as we will see in the following. For the proofs of these results, it is important to keep in mind that an election sampled according to the IAC model is a multiset of votes, i.e., the order of the votes is of no relevance. Thus, the total number of (n, m)-elections is equal to $\left(\!\!\binom{m!}{n}\!\!\right) = \binom{m!+n-1}{n}$.

In the following, let $p_A(n, m)$ denote the probability that an (n, m)-election created according to the IAC assumption is single-peaked.

Proposition 13 *It holds that*

(i.)

$$\frac{m!}{2}\frac{\left(\!\!\binom{2^{m-1}}{n}\!\!\right)}{\left(\!\!\binom{m!}{n}\!\!\right)} \cdot (1+\epsilon(n,m)) \leq p_A(n,m) \leq \frac{m!}{2}\frac{\left(\!\!\binom{2^{m-1}}{n}\!\!\right)}{\left(\!\!\binom{m!}{n}\!\!\right)},$$

where $\epsilon(n, m) \to 0$ *as* $n \to \infty$ *for* $n, m \geq 2$,

(ii.) $p_A(2, m) = \frac{1}{m!+1}\left(\binom{2m-2}{m-1} + 1\right)$ *for* $m \geq 1$ *and*

(iii.) $p_A(n, 3) = \frac{60n}{(n+2)(n+3)(n+4)}$ *for* $n \geq 1$.

Proof We follow the proofs of Theorems 11 and 12.

(i.) Given a fixed axis, there are $\left(\!\!\binom{2^{m-1}}{n}\!\!\right)$ (n, m)-elections that are single-peaked with respect to this axis. Multiplying with the number of axes that need to be considered and dividing by the total number of elections leads to the upper bound on the probability.
For the lower bound, we fix a vote V and determine the number of elections that are single-peaked and contain both V and $\bar{V}$:

$$\begin{aligned}
&\left(\!\!\binom{2^{m-1}}{n}\!\!\right) - 2 \cdot \left(\!\!\binom{2^{m-1}-1}{n}\!\!\right) + \left(\!\!\binom{2^{m-1}-2}{n}\!\!\right) \\
&= \left(\!\!\binom{2^{m-1}}{n}\!\!\right) \cdot \left(1 - 2 \cdot \frac{2^{m-1}-1}{2^{m-1}+n-1} + \frac{(2^{m-1}-1)(2^{m-1}-2)}{(2^{m-1}+n-1)(2^{m-1}+n-2)}\right) \\
&= \left(\!\!\binom{2^{m-1}}{n}\!\!\right) \cdot (1+\epsilon(n,m)),
\end{aligned}$$

where it can be checked easily that $\epsilon(n, m) \to 0$ as n tends to infinity. This gives the lower bound.

(ii.) We pick one vote at random, the second vote can be one of $\binom{2m-2}{m-1}$ possibilities (cf. Theorem 12). However, the order of votes does not matter and we are thus double-counting profiles that consist of two distinct votes–there are $m!(\binom{2m-2}{m-1}-$

$1)/2$ such profiles. Adding the $m!$ profiles in which the same vote occurs twice, we obtain the following number of possibilities:

$$\frac{m!}{2}\left(\binom{2m-2}{m-1}+1\right)$$

Dividing by the total number of $(2, m)$-elections leads to the probability.

(iii.) An election with three candidates is single-peaked if and only if it has at most two last-ranked candidates. Using inclusion-exclusion one obtains that the total number of possible elections is:

$$3\cdot\left(\left(\!\binom{4}{n}\!\right)-\left(\!\binom{2}{n}\!\right)\right)=\frac{n+1}{2}\left((n+2)(n+3)-6\right).$$

Again, dividing by the total number of elections gives the probability. □

The case with 4 candidates that corresponds to case *(iv)* in Theorem 12 can not be directly derived from the IC case and would need a far more involved inclusion-exclusion argument. It is thus omitted here.

6 The Pólya urn model

The Pólya urn model (also refereed to as the Pólya-Eggenberger urn model) (Johnson and Kotz 1977; Berg 1985; Mahmoud 2008) is an approach to sample elections with a variable degree of *social homogeneity*, i.e., where preferences are not independent but voters tend to have the same preferences as other voters. In the following the parameter a, a non-negative integer, describes the degree of social homogeneity. As we will see in a moment, the case $a = 0$ corresponds to the Impartial Culture assumption, i.e., a population with no homogeneity.

The setting of the Pólya urn model for an election with n votes and m candidates can be described as follows. Consider a large urn containing $m!$ balls. Every ball represents one of the $m!$ possible votes on the candidate set and has a different color. An election is then created by subsequently pulling n balls out of the urn according to the following rule. The first ball is pulled at random and constitutes the first vote of the election. Then the pulled ball is returned to the urn and a other balls of the same color are added to the urn. This procedure is repeated n times until an election consisting of n votes is created.

At a first glance, it might seem that the probability assigned to a certain election within the Pólya urn model depends on the order of the votes. However this is not the case: Any election that can be obtained by rearranging a given election $(C, \mathcal{P})$, i.e, by changing the order of the votes, has exactly the same probability of occurring as the election $(C, \mathcal{P})$ itself. First, when the i-th ball is drawn from the urn, i.e., when the i-th vote is chosen, there are always $m! + (i-1)\cdot a$ balls present in the urn. Second, for any vote V the number of balls corresponding to V, i.e., the number of favourable cases, only depends on how often the vote V has already been pulled out of the urn and is equal to $(1 + k \cdot a)$ where k is the number of times V has already been pulled.

It is now easy to give a concise characterization of this discrete distribution. In order to alleviate notation, let us use the so called *Pochhammer k-symbol* as introduced by Díaz and Pariguan (2007).

Definition 6 The *Pochhammer k-symbol* is defined as $(x)_{n,k} = \prod_{i=1}^{n}(x+(i-1)\cdot k)$ where in our context $x \in \mathbb{R}$ and n, k are non-negative integers. Note that $(x)_{n,1} = x(x+1)(x+2)\ldots(x+n-1)$ is the ordinary Pochhammer symbol (also known as rising factorial) and $(1)_{n,1} = n!$.

We can now define the probability of a given (n, m)-election with ℓ distinct votes. Let $n_i, i \in [m!]$ be non-negative integers with $\sum_{i=1}^{m!} n_i = n$ such that, for all $i \in [\ell]$ vote V_i appears n_i times. The probability of such an election is given by:

$$\binom{n}{n_1,\ldots,n_\ell} \cdot \frac{\prod_{i=1}^{\ell}(1)_{n_i,a}}{(m!)_{n,a}} = \frac{n!}{\prod_{i=1}^{\ell} n_i!} \cdot \frac{\prod_{i=1}^{\ell}(1)_{n_i,a}}{(m!)_{n,a}}. \tag{2}$$

Note that setting $a = 0$ corresponds to the case where every one of the n votes is drawn from exactly the same urn, namely the urn containing every one of the $m!$ balls exactly once. Thus, the votes are chosen independently and every vote has the same probability of occurring; this corresponds to the Impartial Culture assumption.

Setting the homogeneity factor to $a = 1$ leads to the Impartial Anonymous Culture (IAC) assumption in which elections are considered as multisets and not as lists of votes and every election has the same probability of occurring. This can be seen as follows: Under IAC an election is fully characterized by the numbers $n_i, i \in [m!]$, as defined above. Setting $a = 1$ in equation (2) in order to determine the probability of an election in which vote V_i appears n_i times, we obtain:

$$\frac{n!}{\prod_{i=1}^{\ell} n_i!} \cdot \frac{\prod_{i=1}^{\ell}(1)_{n_i,1}}{(m!)_{n,1}} = \frac{n!}{\prod_{i=1}^{\ell} n_i!} \cdot \frac{\prod_{i=1}^{\ell} n_i!}{\prod_{j=1}^{n}(m!+(j-1))} = \frac{1}{\binom{m!+n-1}{n}}.$$

Clearly, this probability does not depend on the choice of the numbers $n_i, i \in [m!]$, and thus every multiset of votes has the same probability of being sampled.

Since we have consider IC and IAC already in Sects. 4 and 5 and have obtained asymptotically optimal results, the following lower bound theorem is interesting only for $a > 1$.

Theorem 14 *Let $p_P(n, m, a)$ denote the probability that an (n, m)-election created according to the Pólya urn model with homogeneity $a > 0$ is single-peaked. It holds that:*

$$p_P(n,m,a) \geq \frac{m!(n-1)!}{a\left(\frac{m!}{a}\right)_{n,1}} \cdot \left[1 + \frac{2}{a}\binom{2m-2}{m-1}H_{n-1} + \frac{n}{a}\sum_{l=2}^{n-1}\frac{(2^{m-1}-2)_{n-l,a}}{a^{n-l}} \cdot \frac{H_{l-1}}{l(n-l)!}\right],$$

where H_k denotes the k-th harmonic number $\sum_{i=1}^{k}\frac{1}{i}$.

Proof Before we start with the actual proof, let us collect a few useful observations. In the following we will use that

$$(m!)_{n,a} = a^n \cdot \prod_{i=1}^{n} \left(\frac{m!}{a} + (i-1) \right) = a^n \left(\frac{m!}{a} \right)_{n,1}. \tag{3}$$

Moreover, we will use the following bound:

$$(1)_{k,a} = \prod_{i=1}^{k} (1 + (i-1) \cdot a) \geq \prod_{i=2}^{k} (i-1) \cdot a = (k-1)! \cdot a^{k-1}. \tag{4}$$

Since it holds that

$$\frac{1}{i \cdot (l-i)} = \frac{1}{l} \cdot \left(\frac{l-i+i}{i \cdot (l-i)} \right) = \frac{1}{l} \cdot \left(\frac{1}{i} + \frac{1}{l-i} \right)$$

the following sum can be expressed with the help of the harmonic numbers:

$$\sum_{i=1}^{l-1} \frac{1}{i \cdot (l-i)} = \frac{1}{l} \cdot \left(\sum_{i=1}^{l-1} \frac{1}{i} + \sum_{i=1}^{l-1} \frac{1}{l-i} \right) = \frac{2}{l} \cdot \sum_{i=1}^{l-1} \frac{1}{i} = \frac{2}{l} H_{l-1}. \tag{5}$$

The proof of the theorem is now split in three parts: First, we consider elections with only one distinct vote. Then, we determine a lower bound on the probability of single-peaked elections that consist of exactly two distinct votes. Third, we give a lower bound on the number of single-peaked elections that contain at least three distinct votes.

Let us now start with the first part of the lower bound. Clearly, an election in which all votes are identical is single-peaked. Let us denote the probability of this event by p_1; in the following we fix m, n, a an omit them in the notation. According to the discrete probability distribution of the Pólya urn model (2), the probability of this event is:

$$p_1 = m! \cdot \frac{(1)_{n,a}}{(m!)_{n,a}} \geq \frac{m!(n-1)!a^{n-1}}{a^n \left(\frac{m!}{a} \right)_{n,1}} = \frac{m!(n-1)!}{a \left(\frac{m!}{a} \right)_{n,1}},$$

where we used Eqs. (3) and (4).

Next, we want to determine the probability that an election is sampled that is single-peaked and consists of exactly two distinct votes. Let us denote the probability of this event by p_2. From the first statement of Theorem 12 we know that there are exactly $m! \cdot \binom{2m-2}{m-1}$ elections with two voters and m candidates that are single-peaked. That is, if we pick a first vote V_1 at random, there are $\binom{2m-2}{m-1}$ votes V_2 that form a single-peaked election together with V_1. We thus have the following (again according to (2)):

$$p_2 = m!\binom{2m-2}{m-1} \cdot \sum_{i=1}^{n-1} \underbrace{\mathbb{P}(i \text{ votes equal to } V_1 \text{ and } n-i \text{ votes equal to } V_2)}_{=p'(i)}$$

According to (2) the probability $p'(i)$ is equal to

$$p'(i) = \binom{n}{i} \frac{(1)_{i,a} \cdot (1)_{n-i,a}}{(m!)_{n,a}}.$$

Using the bound in Eq. (4) and the equality in Eq. (3), $p'(i)$ can be bounded from below as follows:

$$p'(i) \geq \binom{n}{i} \frac{(i-1)! \cdot a^{i-1} \cdot (n-i-1)! \cdot a^{n-i-1}}{(m!)_{n,a}} = \frac{n! \cdot a^{n-2}}{a^n \left(\frac{m!}{a}\right)_{n,1} \cdot i(n-i)}.$$

For p_2 we thus obtain

$$\begin{aligned} p_2 &\geq m! \cdot \binom{2m-2}{m-1} \frac{n!}{a^2 \left(\frac{m!}{a}\right)_{n,1}} \sum_{i=1}^{n-1} \frac{1}{i(n-i)} \\ &= \frac{m!(n-1)!}{\left(\frac{m!}{a}\right)_{n,1}} \cdot \binom{2m-2}{m-1} \frac{2}{a^2} H_{n-1}, \end{aligned}$$

where the transformation from the first to the second line is done with the identity in Eq. (5).

Finally, for single-peaked elections that have more than two distinct votes, we only consider elections that contain a vote V and also its reverse $\bar{V}$. Let us denote the probability of this event by p_3. As in the proof of the bounds under the IC assumption this idea is based on the following fact about single-peakedness: If a vote V and its reverse vote $\bar{V}$ are both present within an election, then there are at most two axes with respect to which this election can be single-peaked, namely the axes V and $\bar{V}$. Thus, if a single-peaked election contains both the vote V and $\bar{V}$, all the other votes must be among the $2^{m-1} - 2$ votes that are also single-peaked with respect to the axis V (respectively $\bar{V}$). Let us denote this set of votes that are not equal to V or $\bar{V}$ and that are single-peaked with respect to the axis V by S_V.

The probability p_3 is then given as follows:

$$\begin{aligned} p_3 =& \frac{m!}{2} \sum_{l=2}^{n-1} \sum_{i=1}^{l-1} \mathbb{P}(i \text{ votes equal to } V, l-i \text{ votes equal to } \bar{V}, n-l \text{ votes in } S_V) \\ =& \frac{m!}{2} \sum_{l=2}^{n-1} \sum_{i=1}^{l-1} \binom{n}{i} \binom{n-i}{l-i} \frac{(1)_{i,a} \cdot (1)_{l-i,a} \cdot (2^{m-1}-2)_{n-l,a}}{(m!)_{n,a}}, \end{aligned}$$

where $m!/2$ stands for the number of possible choices for the vote V, i is the number of times the vote V appears and $l - i$ is the number of times the vote $\bar{V}$ appears.

By using the bound in Eq. (4) as well as the identity in Eq. (5), we obtain the following:

$$\begin{aligned} p_3 &\geq \frac{m!}{2(m!)_{n,a}} \sum_{l=2}^{n-1} (2^{m-1}-2)_{n-l,a} \sum_{i=1}^{l-1} \binom{n}{i}\binom{n-i}{l-i} (i-1)!(l-i-1)!a^{l-2} \\ &= \frac{m!}{2(m!)_{n,a}} \sum_{l=2}^{n-1} (2^{m-1}-2)_{n-l,a} \cdot a^{l-2} \sum_{i=1}^{l-1} \frac{n!(n-i)!(i-1)!(l-i-1)!}{i!(n-i)!(l-i)!(n-l)!} \\ &= \frac{m!n!}{2(m!)_{n,a}} \sum_{l=2}^{n-1} (2^{m-1}-2)_{n-l,a} \cdot \frac{a^{l-2}}{(n-l)!} \sum_{i=1}^{l-1} \frac{1}{i(l-i)} \\ &= \frac{m!n!}{\left(\frac{m!}{a}\right)_{n,1}} \sum_{l=2}^{n-1} \frac{(2^{m-1}-2)_{n-l,a}}{a^{n+2-l}} \cdot \frac{1}{l(n-l)!} H_{l-1}. \end{aligned}$$

Since $p_1 + p_2 + p_3 \leq p_P(n, m, a)$, we obtain the desired lower bound. □

To illustrate the rather involved lower bound of Theorem 14, we consider the special case of $a = m!$. This special case corresponds to highly homogeneous elections; the probability that the first and the second vote are identical is roughly 50%. It is a typical assumption that a is a multiple of $m!$ (McCabe-Dansted and Slinko 2006; Walsh 2010, 2011) since otherwise, i.e., for a fixed a, the actual homogeneity of elections drawn according to the Pólya urn model would depend on the number of candidates m.

Corollary 15 *Let $p_P(n, m, m!)$ denote the probability that an (n, m)-election created according to the Pólya urn model with homogeneity $m!$ is single-peaked. It holds that:*

$$p_P(n, m, m!) \geq \frac{1}{n} \cdot \left(1 + 2\frac{\ln(n-1)}{m!} \cdot \frac{(2m-2)!}{((m-1)!)^2}\right).$$

We see that for $a = m!$ and small n, there is a significant probability that the Pólya urn model produces single-peaked elections.

7 Mallows model

The Mallows model (Mallows 1957) assumes that there is a *reference vote* and votes are more likely to appear in an election if they are close to this reference vote. Closeness is measured by the Kendall tau rank distance, defined as follows.

Definition 7 Given two votes V and W contained in an election $(C, \mathcal{P})$, the *Kendall tau rank distance* $\kappa(V, W)$ is a metric that counts the number of pairwise disagreements between V and W. To be more precise:

$$\kappa(V, W) = |\left\{\{c, c'\} \subseteq C : (V : cc' \wedge W : c'c) \vee (V : c'c \wedge W : cc')\right\}|.$$

Note that $\kappa(V, W)$ is also the minimum number of transpositions, i.e., swaps, of adjacent elements, needed to transform V into W or vice versa. We can now define the Mallows model.

Definition 8 Let C be a set of candidates with $|C| = m$ and let $T(C)$ be the set of all total orders on C. Given a *reference vote* V and a real number $\phi \in (0, 1]$, the so-called *dispersion parameter*, the *Mallows model* is defined as follows. Every vote W of an (n, m)-election is determined independently from the others according to the following probability distribution:

$$\mathbb{P}_{V,\phi}(W) = \frac{1}{Z} \cdot \phi^{\kappa(V,W)}, \tag{6}$$

where the normalization constant $Z = \sum_{W \in T(C)} \phi^{\kappa(V,W)}$ fulfils $Z = 1 \cdot (1 + \phi) \cdot (1 + \phi + \phi^2) \cdots (1 + \cdots + \phi^{m-1})$.

Note that choosing $\phi = 1$ corresponds the Impartial Culture assumption and as $\phi \to 0$ one obtains a distribution that concentrates all mass on V.

Theorem 16 *Let $p_M(n, m, \phi)$ denote the probability of an (n, m)-election being single-peaked if it is created according to the Mallows model with dispersion parameter ϕ. Then the following lower bound holds:*

$$p_M(n, m, \phi) \geq \left(\frac{1 + \phi \cdot (m - 1) + \phi^2 \cdot (m - 2)(m - 3)/2}{Z}\right)^n.$$

Proof Without loss of generality, we can assume that the reference vote is $V : c_1c_2 \ldots c_m$. We define an axis A as follows: $A : \ldots c_6c_4c_2c_1c_3c_5 \ldots$. Clearly, V is single-peaked with respect to A and it will turn out that "many" other votes that are close to V with respect to the Kendall tau distance are also single-peaked with respect to this axis. See Fig. 3 for a representation of the axis A and the reference vote V for the case of seven candidates. In the following, we will write "W is SP" as a short form of "the total order W on C is single-peaked with respect to axis A".

The idea of this proof is to bound the probability $p_M(n, m, \phi)$ from below as follows:

$$p_M(n, m, \phi) \geq \left(\mathbb{P}_{V,\phi}(W : W \text{ is SP})\right)^n.$$

Moreover, we use the following bound:

$$\mathbb{P}_{V,\phi}(W : W \text{ is SP}) \geq \mathbb{P}_{V,\phi}(V) + \sum_{\substack{W \text{ is SP and} \\ \kappa(V,W)=1}} \mathbb{P}_{V,\phi}(W) + \sum_{\substack{W \text{ is SP and} \\ \kappa(V,W)=2}} \mathbb{P}_{V,\phi}(W)$$

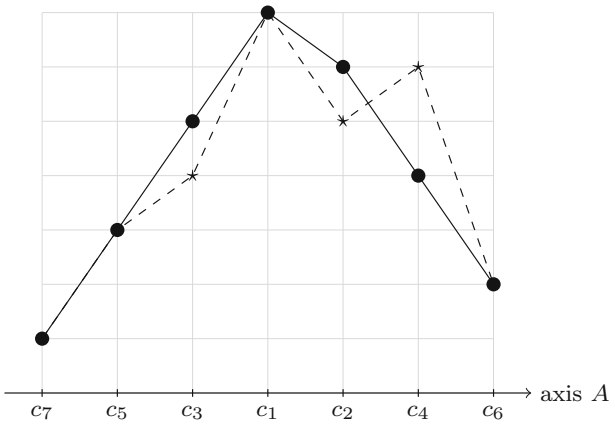

Fig. 3 The axis A and the reference vote $V : c_1c_2c_3c_4c_5c_6c_7$, shown as a *solid line*. The *dashed line* represents the vote $W : c_1c_4c_2c_3c_5c_6c_7$ and is not single-peaked with respect to A. The Kendall tau distance of V and W is 2. Note that all votes with a Kendall tau distance to V of 1 are single-peaked with respect to A

First, it is clear that $\mathbb{P}(V) = 1/Z$.

Second, we need to compute the number of votes W that are single-peaked with respect to A and that fulfil $\kappa(V, W) = 1$. Votes W with $\kappa(V, W) = 1$ are votes in which the order of exactly one pair of candidates (c_i, c_{i+1}) has been changed in V. Since there are $(m - 1)$ pairs of adjacent candidates in V, there are exactly $(m - 1)$ votes W with $\kappa(V, W) = 1$.

All these votes are single-peaked with respect to the axis A since:

- If c_1 and c_2 are interchanged, the position of the peak on the axis A is changed, but clearly no new peaks arise.
- If two other candidates c_i and c_{i+1} are interchanged, one of these two candidates lies to the left of the peak on A and the other one to the right of the peak. Thus, interchanging only these two candidates does not create a new peak either.

Therefore we have the following:

$$\sum_{\substack{W \text{ is SP and} \\ \kappa(V,W)=1}} \mathbb{P}_{V,\phi}(W) = (m-1) \cdot \frac{\phi}{Z}.$$

Third, we need to compute the number of votes W that are single-peaked with respect to A and that fulfil $\kappa(V, W) = 2$. Here we have a different situation: Not all votes that can be obtained by exactly two swaps of adjacent candidates in V are single-peaked with respect to A. For instance, first swapping the candidates (c_3, c_4) and then swapping (c_2, c_4) in V, does not lead to a vote that is single-peaked with respect to A. For this example, see the vote shown as a dashed line in Fig. 3. The problem here is that the swapping of these two pairs changes the order of c_2 and c_4, two elements that both lie on the same side of the peak on A, and thus a valley is created by the elements c_1, c_2 and c_4. In general, a pair of swaps (c_i, c_{i+1}) and (c_j, c_{j+1}) is always allowed if

the two pairs of candidates do not have any elements in common. Note that the order of the two (disjoint) swaps is of no importance and without loss of generality we can assume that $i + 1 < j$.

Knowing this, we can bound the number of votes W with $\kappa(V, W) = 2$ that are single-peaked with respect to A as follows: If the first swap is (c_i, c_{i+1}) for some $i \in [1, m-3]$ and the second swap (c_j, c_{j+1}) is disjoint from the first one, j has to fulfil $j \in [i+2, m-1]$ and thus there are $(m - i - 2)$ possibilities for (c_j, c_{j+1}). Summing over all possible i we obtain that there are at least

$$\sum_{i=1}^{m-3} m - i - 2 = \frac{(m-2)(m-3)}{2}$$

many votes W with $\kappa(V, W) = 2$ that are single-peaked with respect to A. Thus we have

$$\sum_{\substack{W \text{ is SP and} \\ \kappa(V,W)=2}} \mathbb{P}_{V,\phi}(W) \geq \frac{\phi^2}{Z} \cdot \frac{(m-2)(m-3)}{2}.$$

Putting the results for Kendall tau distance equal to 0, 1 and 2 together we obtain the desired lower bound. □

The lower bound result of Theorem 16 does not give an immediate intuition for the likelihood of single-peakedness under Mallows model. Hence we consider the special case $\phi = \frac{1}{m}$. This substitution yields a simpler lower bound, which is considerably larger than, e.g., the lower bound of roughly $(2^m/m!)^n$ obtained for the Impartial Culture Assumption (Theorem 11). For a short discussion on "realistic" parameter values ϕ we refer to Sect. 8.

Corollary 17 *Assuming $\phi = \frac{1}{m}$, it holds that*

$$p_M(n, m, \phi) \geq \left(1.5 \left(\frac{1 - \frac{1}{m}}{1 - \left(\frac{1}{m}\right)^m}\right)^{m-1}\right)^n > \left(1 - \frac{1}{m}\right)^{(m-1)n}.$$

Proof Inserting $\phi = \frac{1}{m}$ in the lower bound of Theorem 16 yields

$$p_M(n, m, \phi) \geq \left(\frac{1 + \frac{m-1}{m} + \frac{(m-2)(m-3)}{2m^2}}{\left(1 + \frac{1}{m}\right) \cdot \left(1 + \frac{1}{m} + \frac{1}{m^2}\right) \cdots \left(1 + \frac{1}{m} + \cdots + \frac{1}{m^{m-1}}\right)}\right)^n. \quad (7)$$

As can be checked easily, the numerator is larger than 1.5. Moreover the finite geometric sums in the denominator are all bounded from above by

$$\frac{1 - \left(\frac{1}{m}\right)^m}{1 - \frac{1}{m}}.$$

Table 1 The likelihood that an (n, m)-election is single-peaked when drawn according to the Impartial Culture assumption

(n, m)	Exact probability	(n, m)	Lower bound	Upper bound
(2, 3)	1	(2, 5)	0.58	0.58
(5, 3)	0.38	(5, 5)	1.6×10^{-4}	2.6×10^{-3}
(10, 3)	0.05	(10, 5)	2.2×10^{-8}	1.1×10^{-7}
(25, 3)	1.19×10^{-4}	(25, 5)	5.0×10^{-21}	8.0×10^{-21}
(50, 3)	4.70×10^{-9}	(50, 5)	9.7×10^{-43}	1.1×10^{-42}
(2, 4)	0.83	(2, 10)	1.3×10^{-2}	1.3×10^{-2}
(5, 4)	0.05	(5, 10)	7.6×10^{-18}	1.1×10^{-13}
(10, 4)	2.03×10^{-4}	(10, 10)	1.9×10^{-36}	5.7×10^{-33}
(25, 4)	1.42×10^{-11}	(25, 10)	2.3×10^{-93}	1.0×10^{-90}
(50, 4)	1.67×10^{-23}	(50, 10)	4.6×10^{-189}	5.5×10^{-187}

The probabilities in the left part of the table are obtained via Theorem 12; those in the right part are obtained via Theorem 11 (except for $n = 2$, which are obtained via Theorem 12 and are exact)

We thus obtain

$$p_M(n, m, \phi) \geq \left(1.5 \left(\frac{1 - \frac{1}{m}}{1 - \left(\frac{1}{m}\right)^m}\right)^{m-1}\right)^n.$$

Replacing the numerator in (7) by 1 and every one of the finite geometric sums in the denominator by an infinite geometric row leads to the second, much rougher lower bound. Note that this is simply a bound on the probability of sampling an (n, m)-election in which all votes are equal to the reference vote. □

8 Numerical evaluations

In this section we provide numerical evaluations of our probability results from the previous sections and make some observations based on these evaluations. In Table 1, we list exact probabilities that an (n, m)-election is single-peaked assuming the Impartial Culture assumptions for small values of m and bounds for these probabilities for a larger number of candidates. Table 2 shows probabilities for elections of the same size assuming the Impartial Anonymous Culture.

Finally, Table 3 shows lower bounds for the Pólya urn model and Table 4 shows lower bounds for the Mallows model.

We conclude this section with the following observations:

– The probabilities shown in Table 1 illustrate how unlikely it is that an election drawn according to IC is single-peaked. Single-peakedness is a strong combinatorial property, so it is not surprising that is is not satisfied by elections sampled uniformly at random. However, it is noteworthy that even for very small n and m

Table 2 The likelihood that an (n, m)-election is single-peaked when sampled according to the Impartial Anonymous Culture assumption

(n, m)	Lower bound	Upper bound	(n, m)	Lower bound	Upper bound
(2, 3)	1	1	(2, 5)	0.58	0.58
(5, 3)	0.59	0.59	(5, 5)	2.17×10^{-4}	4.13×10^{-3}
(10, 3)	0.27	0.27	(10, 5)	1.19×10^{-7}	7.98×10^{-7}
(25, 3)	0.068	0.068	(25, 5)	1.44×10^{-16}	3.76×10^{-16}
(50, 3)	0.02	0.02	(50, 5)	2.91×10^{-28}	4.94×10^{-28}
(2, 4)	0.84	0.84	(2, 10)	1.3×10^{-2}	1.3×10^{-2}
(5, 4)	1.46×10^{-2}	9.67×10^{-2}	(5, 10)	7.78×10^{-18}	1.03×10^{-13}
(10, 4)	8.34×10^{-4}	2.52×10^{-3}	(10, 10)	2.06×10^{-36}	6.19×10^{-33}
(25, 4)	7.89×10^{-7}	1.30×10^{-6}	(25, 10)	3.69×10^{-93}	1.76×10^{-90}
(50, 4)	4.28×10^{-10}	5.58×10^{-10}	(50, 10)	4.29×10^{-188}	5.05×10^{-186}

The probabilities in both the left and right part of the table are obtained via Proposition 13

the probability is small, e.g., for $m = n = 5$ it is less than 0.0026. Conversely, our results indicate that even for very small real-world single-peaked data sets it is highly unlikely that their single-peakedness is the product of mere chance.

– As can be seen in Table 2, the probability that an election is single-peaked is slightly higher when it is sampled according to the IAC model than when it is sampled according to the IC model. This can be explained heuristically as follows: Under the IAC model, elections with many coinciding votes have the same likelihood of appearing as elections consisting of many different votes. However, in the IC model, elections consisting of many different votes have a higher chanced of being sampled because of the many ways in which the votes can be rearranged. It is clear that an election where most votes are the same has a higher chance of being single-peaked than an election in which very many different votes appear. Thus, it is not surprising if the likelihood of single-peakedness is higher under the IAC than under the IC model.
– For the Pólya urn model (Table 3) we observe significantly higher probabilities. This is of course due to our chosen parameter values a—recall that $a = 0$ implies IC and $a = 1$ implies IAC. In particular for $a = m!$ we see that single-peaked profiles arise with considerable likelihood. The assumption of $a = m!$ is common in the literature (McCabe-Dansted and Slinko 2006; Walsh 2010, 2011) and even values of up to $a = 3m!$ have been considered (Dominique Lepelley 2003). As a consequence, we learn from these probabilities that setting $a = m!$ generates extremely homogeneous profiles, even to the extent that they become single-peaked.
– Even higher probabilities are shown in Table 3. We see that for $\phi = 0.05$ single-peakedness is likely to be observed, e.g., with a probability > 0.49 for $n = 50$ and $m = 5$. Clearly, $\phi = 0.05$ is a strong assumption and profiles obtained in this way are highly homogeneous; in fact, $\phi = 0.05$ implies that all voters share the same preferences except for minor deviations, which, in turn, enables the single-peakedness property to hold. For $\phi = 0.1$ we still see a significant chance of single-peakedness, for larger values of ϕ the likelihood deteriorates quickly.

Table 3 Lower bounds obtained from Theorem 14 on the likelihood that an (n, m)-election is single-peaked when sampled according to the Pólya urn model with homogeneity a

(n, m)	$a = 10$	$a = m!/2$	$a = m!$
(10, 5)	1.6×10^{-4}	0.13	0.43
(25, 5)	8.4×10^{-8}	3.0×10^{-2}	0.21
(50, 5)	1.5×10^{-10}	9.1×10^{-3}	0.12
(10, 10)	3.6×10^{-36}	2.0×10^{-2}	0.10
(25, 10)	2.3×10^{-91}	3.6×10^{-3}	4.4×10^{-2}
(50, 10)	2.6×10^{-181}	9.7×10^{-4}	2.2×10^{-2}

The question arises: what are typical values for ϕ? Betzler et al. (2014) compute maximum-likelihood estimates of ϕ for different real-world data sets. They find values[1] ranging from 0.7 to almost 1. They also generate elections using values for ϕ ranging from 0, 37 to almost 1. Other publications generate election with ϕ in the interval [0.3, 1] (Boutilier et al. 2014) and [0.6, 0.9] (Oren et al. 2013). We see that all these parameter values are too large to imply single-peakedness with non-negligible probability. On the one hand, this implies that values for ϕ small enough to generate single-peakedness profiles are generally too restrictive to be found in (published) experiments. On the other hand, our results allow to argue that the parameter values in the aforementioned papers have been chosen sensibly since the accordingly generated elections contain (at least) enough disagreement as to prevent single-peakedness to arise with significant likelihood.

9 Conclusions and directions for future research

We have seen that the likelihood of single-peaked preferences varies significantly for the Impartial (Anonymous) Culture assumption, the Pólya urn and the Mallows model. For elections chosen according to the IC or the IAC assumption, it is extremely unlikely that single-peakedness arises (cf. Table 1). With Theorem 8, we have shown that unlikeliness also holds for arbitrary domain restrictions that avoid a $(2, k)$-configuration. In contrast, for the Pólya urn and the Mallows model with parameter a (ϕ) chosen sufficiently large (small) it is rather likely that elections are single-peaked. Numerical probabilities in Tables 3 and 4 affirm this claim.

Let us conclude with directions for future research. Theorem 8 requires that the domain restriction avoids a $(2, k)$-configuration and thus is not applicable to domain restrictions such as the single-crossing restriction (Roberts 1977; Bredereck et al. 2013b) or 2D single-peaked restriction (Barberà et al. 1993). It remains open whether this result can be extended to such domain restrictions as well and how the corresponding bound would look like. It would also be interesting to complement Theorem 8 with a corresponding lower bound result. In general, the likelihood of other domain restrictions such as the single-crossing (Roberts 1977) or the 2D single-peaked restriction (Barberà et al. 1993) has yet to be studied.

[1] Their parameter θ is related to ϕ via the equation $\phi = e^{-\theta}$.

Table 4 Lower bounds via Theorem 16 on the likelihood that an (n, m)-election is single-peaked when sampled according to the Mallows model with dispersion parameter ϕ

(n, m)	$\phi = 0.3$	$\phi = 0.2$	$\phi = 0.1$	$\phi = 0.05$	$\phi = 0.01$
(10, 5)	0.02	0.15	0.59	0.86	0.99
(25, 5)	5.7×10^{-5}	8.7×10^{-3}	0.26	0.70	0.98
(50, 5)	3.3×10^{-9}	7.6×10^{-5}	7.2×10^{-2}	0.49	0.97
(10, 10)	3.7×10^{-6}	2.7×10^{-3}	0.20	0.66	0.98
(25, 10)	2.7×10^{-14}	3.7×10^{-7}	1.9×10^{-2}	0.36	0.96
(50, 10)	7.5×10^{-28}	1.4×10^{-13}	3.7×10^{-4}	0.13	0.92

In Sect. 5 we studied the likelihood of single-peakedness under the IAC assumption. In particular, it follows from Theorem 13 that under IAC the likelihood that an $(n, 3)$-election is single-peaked is $\frac{60n}{(n+2)(n+3)(n+4)}$. The likelihood that an $(n, 3)$-election has a Condorcet winner is $\frac{15(n+3)^2}{16(n+2)(n+4)}$ for odd n and $\frac{15(n+2)(n^2+8n+8)}{16(n+1)(n+3)(n+5)}$ for even n (Gehrlein 2002). Note that the probability for single-peakedness is significantly smaller than the latter two and, in particular, the former converges to 0 whereas the latter converge to $15/16$ for $n \to \infty$. Recently, the top monotonicity restriction has been proposed (Barberà and Moreno 2011) which is a generalization of the single-peaked and single-crossing domain but still guarantees a Condorcet winner. It would be highly interesting to know the likelihood of top monotonicity restricted preferences and whether this probability is non-zero for $n \to \infty$.

Another direction is to consider other probability distributions such as the Plackett-Luce model (Plackett 1975; Luce 1959) or Mallows mixture models where more than one reference vote is considered (Murphy and Martin 2003). One could also analyze the probability distribution that arises when assuming that all elections are single-peaked and that all elections of the same size are equally likely. This would allow allow to ask questions such as "How likely is it that a single-peaked election is also single-crossing?". Finally, a recent research direction is to consider elections that are nearly single-peaked, i.e., elections that have a small distance to being single-peaked according to some notion of distance (Faliszewski et al. 2011a; Elkind et al. 2012; Cornaz et al. 2012, 2013; Erdélyi et al. 2013; Bredereck et al. 2013a). The likelihood that elections are nearly single-peaked remains a worthwhile direction for future research.

Acknowledgements The first author was supported by the Austrian Science Foundation FWF, Grant P25337-N23, the second author by the FWF, Grant P25518-N23 and Y698 and by the European Research Council (ERC) under Grant Number 639945 (ACCORD).

References

Arrow KJ (1950) A difficulty in the concept of social welfare. J Polit Econ 58(4):328–346

Ballester MA, Haeringer G (2011) A characterization of the single-peaked domain. Social Choice Welfare 36(2):305–322

Barberà S, Moreno B (2011) Top monotonicity: A common root for single peakedness, single crossing and the median voter result. Games Econ Behav 73(2):345–359

Barberà S, Gul F, Stacchetti E (1993) Generalized median voter schemes and committees. J Econ Theory 61(2):262–289

Berg S (1985) Paradox of voting under an urn model: The effect of homogeneity. Publ Choice 47(2):377–387

Betzler N, Slinko A, Uhlmann J (2013) On the computation of fully proportional representation. J Art Intell Res 47:475–519

Betzler N, Bredereck R, Niedermeier R (2014) Theoretical and empirical evaluation of data reduction for exact Kemeny rank aggregation. Auton Agents Multi-Agent Syst 28(5):721–748

Black D (1948) On the rationale of group decision making. J Polit Econ 56(1):23–34

Bóna M (2014) A new record for 1324-avoiding permutations. Eur J Math 1(1):198–206

Boutilier C, Lang J, Oren J, Palacios H (2014) Robust winners and winner determination policies under candidate uncertainty. In: Proc. of AAAI-14, AAAI Press, pp 1391–1397

Brandt F, Brill M, Hemaspaandra E, Hemaspaandra LA (2015) Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. J Art Intell Res 53:439–496

Bredereck R, Chen J, Woeginger GJ (2013) Are there any nicely structured preference profiles nearby? In: Proc. of IJCAI-13, pp 62–68

Bredereck R, Chen J, Woeginger GJ (2013b) A characterization of the single-crossing domain. Social Choice Welfare 41(4):989–998

Bruner ML (2015) Patterns in labelled combinatorial objects. PhD thesis, TU Wien

Chen J, Pruhs K, Woeginger GJ (2015) The one-dimensional Euclidean domain: finitely many obstructions are not enough. Soc Choice Welf 48(2):409–432. doi:10.1007/s00355-016-1011-y

Coombs CH (1964) A theory of data. Wiley, Nee York

Cornaz D, Galand L, Spanjaard O (2012) Bounded single-peaked width and proportional representation. In: Proc. of ECAI-12, IOS Press, pp 270–275

Cornaz D, Galand L, Spanjaard O (2013) Kemeny elections with bounded single-peaked or single-crossing width. In: Proc. of IJCAI-13, IJCAI/AAAI

Critchlow DE, Fligner MA, Verducci JS (1991) Probability models on rankings. J Math Psychol 35(3):294–318

Demange G (1982) Single-peaked orders on a tree. Math Social Sci 3(4):389–396

Díaz R, Pariguan E (2007) On hypergeometric functions and Pochhammer k-symbol. Divulg Matemáticas 15(2):179–192

Dominique Lepelley FV (2003) Voting rules, manipulability and social homogeneity. Public Choice 116(1/2):165–184

Elkind E, Lackner M (2014) On detecting nearly structured preference profiles. In: Proc. of AAAI-14, AAAI Press, USA pp 661–667

Elkind E, Faliszewski P, Slinko AM (2012) Clone structures in voters' preferences. In: Proc. of EC-12, ACM, New York pp 496–513

Erdélyi G, Lackner M, Pfandler A (2013) Computational aspects of nearly single-peaked electorates. In: Proc. of AAAI-13, AAAI Press, USA

Escoffier B, Lang J, Öztürk M (2008) Single-peaked consistency and its complexity. In: Proc. of ECAI-08, IOS Press, FAIA, vol 178, pp 366–370

Faliszewski P, Hemaspaandra E, Hemaspaandra LA (2011a) The complexity of manipulative attacks in nearly single-peaked electorates. In: Proc. of TARK-11, pp 228–237

Faliszewski P, Hemaspaandra E, Hemaspaandra LA, Rothe J (2011b) The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. Inform Comput 209(2):89–107

Favardin P, Lepelley D, Serais J (2002) Borda rule, Copeland method and strategic manipulation. Rev Econ Design 7(2):213–228

Friedgut E, Kalai G, Nisan N (2008) Elections can be manipulated often. In: Proc. of the 49th IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, pp 243–249

Gehrlein WV (2002) Condorcet's paradox and the likelihood of its occurrence: different perspectives on balanced preferences. Theory Decis 52(2):171–199
Gehrlein WV (2006) Condorcet's paradox. Springer, Berlin Heidelberg
Gehrlein WV, Moyouwou I, Lepelley D (2013) The impact of voters preference diversity on the probability of some electoral outcomes. Math Soc Sci 66(3):352–365
Gehrlein WV, Lepelley D, Moyouwou I (2015) Voters' preference diver-sity, concepts of agreement and Condorcet's paradox. Quality and Quantity 49(6):2345–2368
Gibbard A (1973) Manipulation of voting schemes. Econometrica 41(4):587–601
Guibert O (1995) Combinatoire des permutations à motifs exclus en liason avec mots, cartes planaires et tableaux de Young. PhD thesis, Université de Bordeaux I
Ki Inada (1969) The simple majority decision rule. Econometrica 37(3):490–506
Isaksson M, Kindler G, Mossel E (2012) The geometry of manipulation–a quantitative proof of the Gibbard-Satterthwaite theorem. Combinatorica 32(2):221–250
Johnson NL, Kotz S (1977) Urn models and their application. Wiley, New York
Knoblauch V (2010) Recognizing one-dimensional Euclidean preference profiles. J Math Econ 46(1):1–5
Lepelley D, Valognes F (2003) Voting rules, manipulability and social homogeneity. Public Choice 116(1–2):165–184
Luce RD (1959) Individual choice behavior. Wiley, UK
Mahmoud H (2008) Pólya urn models. Texts in Statistical Science, Chapman and Hall/CRC, London
Mallows CL (1957) Non-null ranking models. i. Biometrika 44(1/2):114–130
Marcus A, Tardos G (2004) Excluded permutation matrices and the Stanley-Wilf conjecture. J Comb Theory A 107(1):153–160
McCabe-Dansted JC, Slinko A (2006) Exploratory analysis of similarities between social choice rules. Group Decis Negot 15(1):77–107
Murphy TB, Martin D (2003) Mixtures of distance-based models for ranking data. Comput Stat Data Anal 41(3–4):645–655
Oren J, Filmus Y, Boutilier C (2013) Efficient vote elicitation under candidate uncertainty. In: Proc. of IJCAI-13, IJCAI/AAAI, pp 309–316
Plackett RL (1975) The analysis of permutations. J Royal Stat Soc Series C (Appl Stat) 24(2):193–202
Roberts KW (1977) Voting over income tax schedules. J Publ Econ 8(3):329–340
Satterthwaite M (1975) Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. J Econ Theory 10(2):187–217
Sen AK (1966) A possibility theorem on majority decisions. Econometrica 34:491–499
Sen AK, Pattanaik PK (1970) Necessary and sufficient conditions for rational choice under majority decision. J Econ Theory 1:178–202
Simion R, Schmidt FW (1985) Restricted permutations. Eur J Comb 6:383–406
Slinko A (2002a) On asymptotic strategy-proofness of classical social choice rules. Theory Decis 52(4):389–398
Slinko A (2002b) On asymptotic strategy-proofness of the plurality and the run-off rules. Social Choice Welfare 19(2):313–324
Slinko A (2005) How the size of a coalition affects its chances to influence an election. Social Choice Welfare 26(1):143–153
Walsh T (2007) Uncertainty in preference elicitation and aggregation. In: Proc. of AAAI-07, AAAI Press, pp 3–8
Walsh T (2010) An empirical study of the manipulability of single transferable voting. In: Proc. of ECAI-10, IOS Press, pp 257–262
Walsh T (2011) Where are the hard manipulation problems? J Art Intell Res 42:1–29
West J (1995) Generating trees and the Catalan and Schröder numbers. Discret Math 146(13):247–262