

The Likelihood of Structure in Preference Profiles

Marie-Louise Bruner

marie-louise.bruner@tuwien.ac.at
Vienna University of Technology
Vienna, Austria

Martin Lackner

lackner@dbai.tuwien.ac.at
Vienna University of Technology
Vienna, Austria

Abstract

In the field of computational social choice, structure in preferences is often described by so-called domain restrictions. Domain restrictions are of major importance since they allow for the circumvention of Arrow’s paradox and for faster algorithms. On the other hand, such structure might be disadvantageous if one seeks to protect voting mechanisms against manipulation and control with the help of computational complexity. So far, it is unclear how likely it is that domain restrictions arise. In this paper, we answer this question from a combinatorial point of view. Our results show how unlikely it is that a preference profile belongs to a restricted domain if it is chosen at random under the Impartial Culture assumption.

Introduction

Detecting and exploiting the structure of data is a major topic in algorithmics and computer science in general. In computational social choice, the most prevalent form of data consists of preferences. Structure in preference data has been studied as *domain restrictions*, such as the single-peaked, single-crossing or 2D single-peaked restriction. There are three main reasons for studying domain restrictions:

Historically, the main motivation for studying domain restrictions was to find a way to escape Arrow’s paradox (Arrow 1950). For example, every single-peaked profile has a Condorcet winner and thus allows for a voting system that is non-dictatorial, Pareto efficient, and independent of irrelevant alternatives. More broadly speaking, voting systems restricted to certain domains may have desirable properties that do not hold in general.

By adopting the algorithmic viewpoint of computational social choice, a second reason for studying domain restrictions becomes apparent: Restricting the domain of preference data often allows for faster algorithms for computationally hard voting problems (Brandt et al. 2010; Betzler, Slinko, and Uhlmann 2013; Walsh 2007). Computational advantages can prevail even if the preference profiles are only close to a certain domain restriction (Cornaz, Galand, and Spanjaard 2013; Skowron et al. 2013). To be able to speak about closeness, several notions of distance have been proposed and studied in the literature (Faliszewski, Hemaspaandra, and Hemaspaandra 2011; Elkind, Faliszewski, and

Slinko 2012; Cornaz, Galand, and Spanjaard 2012; Erdélyi, Lackner, and Pfandler 2013; Bredereck, Chen, and Woeginger 2013a).

A third reason for studying domain restrictions is also related to their influence on computational complexity. A major research topic in computational social choice is the use of complexity to protect elections from manipulation, control and other forms of dishonest behavior. For an overview of this research area we refer to the surveys by Faliszewski, Hemaspaandra, and Hemaspaandra (2010) and Rothe and Schend (2013). As domain restrictions tend to decrease the complexity of voting problems, they have the undesirable effect that, for example, manipulation and control become computationally easier on restricted domains (Brandt et al. 2010; Faliszewski et al. 2011). To some degree, this problem arises even if domain restrictions are relaxed by aforementioned notions of distance (Faliszewski, Hemaspaandra, and Hemaspaandra 2011).

Despite the vast literature on domain restrictions, a fundamental question has not received much attention so far: How likely is it that preference profiles lie in a restricted domain?

There are two experimental studies on that topic: Mattei, Forshee, and Goldsmith (2012) report that in their data sets almost no evidence for the single-peaked restriction was found. Similarly, Sui, Francois-Nienaber, and Boutilier (2013) report also no occurrences of the single-peaked restriction in their data sets. However, they found that these preferences are close to being 2D single-peaked.

Our work, in contrast, is of theoretical nature. We employ combinatorial methods to study the likelihood of structure in preference profiles. In this paper, likelihood is considered with respect to the Impartial Culture assumption, where each vote is equally likely to appear. While this is not a realistic assumption for real-world preference data (cf. Popova, Regenwetter, and Mattei (2013)), the Impartial Culture is the most basic distribution and thus it is generally used to obtain baseline results. Our paper is the first extensive combinatorial analysis of domain restrictions. Its main contributions are listed in the following.

- Many domain restrictions can be characterized by *forbidden configurations*: for example, the single-peaked domain (Ballester and Haeringer 2011) and the single-crossing domain (Bredereck, Chen, and Woeginger 2013b). We prove a close connection between configurations and permutations

(n, m)	SP (lb)	SP (ub)	SC (ub)	2D (ub)
(10, 5)	$8 \cdot 10^{-8}$	$2 \cdot 10^{-7}$	$1 \cdot 10^{-4}$	0.54
(25, 5)	$2 \cdot 10^{-21}$	$8 \cdot 10^{-21}$	$3 \cdot 10^{-14}$	$2 \cdot 10^{-2}$
(10, 10)	$5 \cdot 10^{-33}$	$6 \cdot 10^{-33}$	$3 \cdot 10^{-16}$	$2 \cdot 10^{-5}$
(25, 10)	$9 \cdot 10^{-91}$	$1 \cdot 10^{-90}$	$2 \cdot 10^{-87}$	$9 \cdot 10^{-19}$
(50, 10)	$5 \cdot 10^{-187}$	$6 \cdot 10^{-187}$	$9 \cdot 10^{-204}$	$2 \cdot 10^{-40}$

Figure 1: The likelihood that a random profile (assuming Impartial Culture) with n votes and m candidates is single-peaked (SP), single-crossing (SC) and 2D single peaked (2D). The table contains lower and upper bounds (lb, ub).

(n, m)	k	SP	SC	2D-SP
(25, 10)	5	$1 \cdot 10^{-66}$	$3 \cdot 10^{-59}$	$2 \cdot 10^{-9}$
	10	$2 \cdot 10^{-45}$	$1 \cdot 10^{-33}$	$2 \cdot 10^{-3}$
	15	$2 \cdot 10^{-26}$	$9 \cdot 10^{-10}$	–
(50, 10)	10	$2 \cdot 10^{-138}$	$1 \cdot 10^{-147}$	$7 \cdot 10^{-22}$
	25	$2 \cdot 10^{-76}$	$2 \cdot 10^{-73}$	$2 \cdot 10^{-4}$
	40	$6 \cdot 10^{-23}$	$3 \cdot 10^{-6}$	–

Figure 2: Upper bounds on the likelihood that a random profile (assuming Impartial Culture) with n votes and m candidates is single-peaked (SP), single-crossing (SC) and 2D single peaked (2D) if k votes may be deleted.

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

- We perform a detailed combinatorial analysis of the most commonly used domain restrictions: We study single-peaked, single-crossing and 2D single-peaked profiles. For all these restrictions, we prove upper bounds on the number of domain restricted profiles. In particular, the upper bound for single-peaked profiles is asymptotically tight. Our results indicate that these domain restrictions are highly unlikely to appear in random profiles chosen according to the Impartial Culture assumption. This holds even for profiles with few votes and candidates (cf. Table 1).

- In addition, we study the distance notions of voter deletion and candidate deletion. Our results show that finding a subset of votes belonging to a domain restriction is unlikely as well: When optimally deleting k votes, the remaining profile is still unlikely to be domain restricted (cf. Table 2), even for rather large k . Similar results are obtained for candidate deletion.

- For local candidate deletion we show a completely different type of result: We determine how many candidates have to be deleted at most per vote so that the remaining profile is guaranteed to be single-peaked or 2D single-peaked.

Due to the space limitations most of the proofs had to be omitted.

Preliminaries

Sets and orders. In our paper, two kinds of orders appear: partial and total orders. Let S be a finite set. A partial order of S is a binary relation that is reflexive, antisymmetric and transitive. A total order of S is a partial order that is total, i.e., for every $a, b \in S$, either the pair (a, b) or (b, a) is contained in the relation. Let P be a partial order of S . Instead of writing $(a, b) \in P$, we write $a \leq_P b$ or $b \geq_P a$. We write $a <_P b$ or $b >_P a$ to state that $a \leq_P b$ and $a \neq b$. Given two subsets A and B of S , we write $A >_P B$ to denote that every element in A is larger than every element in B with respect to P . We write $\text{dom}(P)$ to denote the domain of P , i.e., $\text{dom}(P) = S$. Given a set or tuple of partial orders \mathcal{S} , we use $\text{dom}(\mathcal{S})$ to denote $\bigcup_{P \in \mathcal{S}} \text{dom}(P)$. Let T be a total order on S . We write $T(i)$ to denote the i -th largest element with respect to T . We say that $a \in S$ has rank i in T if $T(i) = a$. A total order T is a linearization of a partial order P if $\text{dom}(T) = \text{dom}(P)$ and for all $a, b \in \text{dom}(P)$, $a <_P b$ implies $a <_T b$.

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 on 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 can be identified with the m -permutation $p(T_1, T_2) := \{i \mapsto j : T_1(i) \text{ has rank } j \text{ in } T_2\}$. For $T_1 = b < a < c$ and $T_2 = c < a < b$ we have $p(T_1, T_2) = 321$. Note that $p(T_1, T_2) = p(T_2, T_1)^{-1}$.

Profiles. An (n, m) -profile \mathcal{P} is an n -tuple (V_1, \dots, V_n) of total orders of the candidate set $\{c_1, \dots, c_m\}$. The total orders in a profile represent votes (or preferences). To easier distinguish between votes and other orders, we use \prec_V and \succ_V to compare candidates with respect to a vote V . If $c_i \prec_V c_j$ holds, this means that candidate c_j is preferred to candidate c_i in vote V . We write $\mathcal{P}[S]$ to denote \mathcal{P} restricted to $S \subseteq \text{dom}(\mathcal{P})$. In the following, \mathcal{P} always denotes a profile.

Randomness. The two main probability distributions in social choice theory are the Impartial Culture (IC) and the Impartial Anonymous Culture (IAC) assumption. IC assumes that votes are chosen uniformly at random from the set of all possible votes. In contrast, IAC does not differentiate between profiles that can be obtained from one another by rearranging the list of votes. By this one obtains equivalence classes of profiles. IAC assumes that each equivalence class is equally likely. Thus, the number of distinct profiles under IC is $m!^n$ whereas the number of distinct profiles under IAC is $\binom{m+n-1}{n}$, the number of multisets with n elements chosen from a base set of cardinality $m!$. In our paper, when we speak of a random profile, we always mean a profile randomly chosen under the IC assumption.

Domain restrictions

The single-peaked restriction (Black 1948) is the most widely used restriction. It assumes that the candidates can be ordered linearly and voters prefer candidates close to their

ideal point to candidates that are further away.

Definition 1. Let A be a total order of $\text{dom}(\mathcal{P})$, the so-called axis. A vote $V \in \mathcal{P}$ contains a valley with respect to A on the candidates $c_1, c_2, c_3 \in \text{dom}(\mathcal{P})$ if $c_1 <_A c_2 <_A c_3$, $c_2 \prec_V c_1$ and $c_2 \prec_V c_3$ holds. The profile \mathcal{P} is single-peaked with respect to A if for every $V \in \mathcal{P}$ and for all candidates $c_1, c_2, c_3 \in \text{dom}(\mathcal{P})$, V does not contain a valley with respect to A on c_1, c_2, c_3 . The profile \mathcal{P} is single-peaked if there exists a total order A of $\text{dom}(\mathcal{P})$ such that \mathcal{P} is single-peaked with respect to A .

The single-peaked restriction can be relaxed to a two-dimensional setting (Barberà, Gul, and Stacchetti 1993), in which valleys are less likely to arise.

Definition 2. Let A and B be total orders of $\text{dom}(\mathcal{P})$, the so-called axes. A vote $V \in \mathcal{P}$ contains a 2D-valley with respect to (A, B) on the candidates $c_1, c_2, c_3 \in \text{dom}(\mathcal{P})$ if V contains a (1D) valley with respect to A on c_1, c_2, c_3 as well as a valley with respect to B on c_1, c_2, c_3 . The profile \mathcal{P} is 2D single-peaked with respect to (A, B) if for every vote $V \in \mathcal{P}$ and for all candidates $c_1, c_2, c_3 \in \text{dom}(\mathcal{P})$, V does not contain a 2D-valley with respect to (A, B) on c_1, c_2, c_3 . The profile \mathcal{P} is 2D single-peaked if there exist two total orders A, B of $\text{dom}(\mathcal{P})$ such that \mathcal{P} is 2D single-peaked with respect to (A, B) .

We continue with the single-crossing restriction (Roberts 1977), where the votes and not the candidates are ordered along a linear axis.

Definition 3. Let A be a total order of $\{1, \dots, |\mathcal{P}|\}$. The profile \mathcal{P} is single-crossing with respect to A if the set $\{V \in \mathcal{P} \mid c_1 \prec_V c_2\}$ is an interval with respect to A for every pair of candidates (c_1, c_2) . The profile \mathcal{P} is single-crossing if there exists a total order A of $\{1, \dots, |\mathcal{P}|\}$ such that \mathcal{P} is single-crossing with respect to A .

We will now see that all these configurations share a property: they are definable by a (possibly infinite) set of so-called forbidden configurations. This unified view of domain restrictions will allow us to prove a very general result about domain restrictions in the next section.

Definition 4. (Configurations and containment) An (l, k) -configuration $\mathcal{C} = (C_1, \dots, C_l)$ is an l -tuple of partial orders over $\{x_1, \dots, x_k\}$. A profile \mathcal{P} contains configuration \mathcal{C} if there exist an injective function f from \mathcal{C} into \mathcal{P} and an injective function g from $\text{dom}(\mathcal{C})$ into $\text{dom}(\mathcal{P})$ such that, for any $x, y \in \text{dom}(\mathcal{C})$ and $C \in \mathcal{C}$, it holds that $x <_C y$ implies $g(x) \prec_{f(C)} g(y)$. We use $\mathcal{C} \sqsubseteq \mathcal{P}$ as a shorthand notation to denote that \mathcal{P} contains \mathcal{C} . A profile avoids a configuration \mathcal{C} if it does not contain \mathcal{C} . In such a case we say that \mathcal{P} is \mathcal{C} -restricted.

In Figure 3, we can see a $(2, 4)$ -configuration that is contained in a $(3, 5)$ -profile.

Definition 5. (Configuration definable) Let Γ be a set of configurations. A set of profiles Π is defined by Γ if Π consists exactly of those profiles 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.

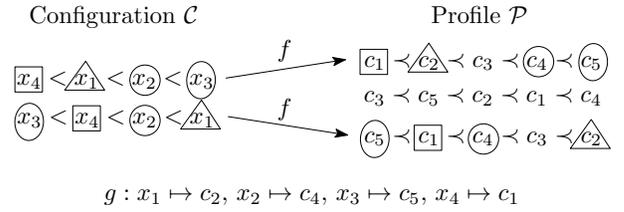


Figure 3: The configuration on the left-hand side is contained in the profile on the right-hand side.

The sets of single-peaked (Ballester and Haeringer 2011) and single-crossing (Bredereck, Chen, and Woeginger 2013b) profiles are known to be finitely configuration definable. The following definition and proposition will allow us to prove that many other domain restrictions, especially 2D single-peakedness, are configuration definable as well.

Definition 6. A set of profiles Π is hereditary if for every profile \mathcal{P}' it holds that $\mathcal{P} \in \Pi$ and $\mathcal{P}' \sqsubseteq \mathcal{P}$ implies $\mathcal{P}' \in \Pi$.

Proposition 1. A set of profiles is configuration definable if and only if it is hereditary.

Proof. Let Π be defined by Γ and $\mathcal{P} \in \Pi$. Assume towards a contradiction that there exists a $\mathcal{P}' \sqsubseteq \mathcal{P}$ with $\mathcal{P}' \notin \Pi$. Since $\mathcal{P} \in \Pi$, \mathcal{P} avoids all configurations in Γ . Since $\mathcal{P}' \sqsubseteq \mathcal{P}$, \mathcal{P}' also avoids all configurations in Γ and thus $\mathcal{P}' \in \Pi$ — a contradiction.

For the other direction we assume that for every profile $\mathcal{P} \in \Pi$ it holds that $\mathcal{P}' \sqsubseteq \mathcal{P}$ implies $\mathcal{P}' \in \Pi$. Let Π^c denote the set of all profiles that are not contained in Π . It is easy to observe that Π^c is a (possibly infinite) set of configurations that defines Π . \square

Corollary 2. The 2D single-peaked restriction is hereditary and hence configuration definable.

It remains open whether this restriction is finitely configuration definable. Finite configuration definability has been used for obtaining algorithms (Bredereck, Chen, and Woeginger 2013a; Elkind and Lackner 2014). A natural example of a meaningful restriction that is not configuration definable is the set of all preference profiles that have a Condorcet winner. The property of having a Condorcet winner is not hereditary and thus cannot be defined by configurations. We also note that there exist sets of profiles that are configuration definable but not finitely configuration definable. The proof of this statement builds upon the relation of permutation patterns and configurations but had to be omitted.

The connection to permutation patterns

In this section, we establish a strong link between the concept of configuration containment in profiles and the concept of pattern containment in permutations. We refer the interested reader to *Combinatorics of permutations* (Bóna 2004) which gives a very good overview of the field of pattern avoidance in permutations. The central definition within this field is the following:

Definition 7. A k -permutation π is contained as a pattern in an n -permutation σ , if there is a subsequence of σ that is order-isomorphic to π . In other words, π is contained in σ , if there is a strictly increasing map $\mu : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ so that the sequence $\mu(\pi) = (\mu(\pi(1)), \mu(\pi(2)), \dots, \mu(\pi(k)))$ is a subsequence of σ . This map μ is called a matching of π into σ . If there is no such matching, σ avoids the pattern π .

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

First we will need the following lemma that establishes a link between configuration containment in profiles and pattern containment in permutations. As of now, we shall denote by $S_m(\pi_1, \dots, \pi_l)$ the cardinality of the set of m -permutations that avoid the permutations π_1, \dots, π_l .

Lemma 3. Let $\mathcal{C} = (C_1, C_2)$ be a configuration containing two total orders and m a positive integer. Furthermore, let V_1 be a vote on m candidates. Then the number of votes V_2 such that $\mathcal{P} = (V_1, V_2)$ avoids \mathcal{C} is equal to $S_m(\pi, \pi^{-1})$, where $\pi = p(C_1, C_2)$.

From this lemma follows a very general result that is applicable to any set of configurations that contains at least one configuration of cardinality two.

Theorem 4. Let $a(n, m, \Gamma)$ be the number of (n, m) -profiles avoiding a set of configurations Γ . Let $k \geq 2$. If a set of configurations Γ contains a $(2, k)$ -configuration $\mathcal{C} = (C_1, C_2)$, 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 on preference profiles. Indeed, $m! \cdot c_k^{(n-1)m}$ is very small compared to the total number of (n, m) -profiles which is $(m!)^n$.

Proof. Without loss of generality we can assume that \mathcal{C} consists of two total orders. Indeed, if \mathcal{C} consists of partial orders, we can simply choose any linearization \tilde{C}_1 of C_1 and \tilde{C}_2 of C_2 and take $\tilde{\mathcal{C}} = (\tilde{C}_1, \tilde{C}_2)$ instead of \mathcal{C} . Then it clearly holds that $a(n, m, \{\mathcal{C}\}) \leq a(n, m, \{\tilde{\mathcal{C}}\})$.

Let us start by choosing the first vote V_1 of the profile 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 \mathcal{C} . If we relax this condition and only demand that none of the pairs (V_1, V_i) for $i \neq 1$ contain the forbidden configuration, we clearly obtain an upper bound for $a(n, m, \{\mathcal{C}\})$. Now Lemma 3 tells us that there are — under this relaxed condition — $S_m(\pi, \pi^{-1})$ choices for every V_i where $\pi := p(C_1, C_2)$. Thus we have the following upper bound:

$$a(n, m, \{\mathcal{C}\}) \leq m! S_m(\pi, \pi^{-1})^{n-1} \leq m! S_m(\pi)^{n-1}, \quad (1)$$

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 $a(n, m, \{\mathcal{C}\})$ is an upper bound for $a(n, m, \Gamma)$ we obtain the desired upper bound. \square

The proof of the Marcus-Tardos theorem provides an explicit exponential formula for the constants c_k , but these constants are far from being optimal. There is an ongoing effort to find exact formulas for $S_m(\pi)$ with fixed π .

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. This includes the single-peaked restriction as well as the 1D Euclidean (Coombs 1964; Knoblauch 2010) and group separable (Ballester and Haeringer 2011) restrictions. In the next section, we prove a better bound for the single-peaked restriction that is even asymptotically optimal.

Combinatorial Results for Domain Restrictions

In this section, we present our combinatorial results on the number of profiles avoiding a set of configurations. We shall always denote by $a(n, m, D)$ the number of (n, m) -profiles belonging to the domain restriction D . In the following we derive upper bounds for $a(n, m, D)$ where D is one of the following domain restrictions: single-peaked (SP), single-crossing (SC) or 2D single-peaked ($2D$). From our results it is easy to derive bounds on the probability that a random (n, m) -profile is within one of the mentioned domain restrictions. This is simply $a(n, m, \Gamma)/m!^n$, where $m!^n$ is the total number of (n, m) -profiles.

Theorem 5. For $n, m \geq 2$ it holds that

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

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

Proof. First observe that a profile 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, as shown by Escoffier, Lang, and Öztürk (2008), the number of votes that are single-peaked with respect to a given axis is 2^{m-1} .

For every one of the $m!/2$ axes considered, select an n -tuple 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.

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 V and \bar{V} themselves. Thus the presence of the votes V and \bar{V} in a profile forces the axis to be equal to either V or \bar{V} . If we fix a vote V , the number of single-peaked profiles containing both V and \bar{V} can thus be determined exactly. Multiplying this by the number of possible choices for V leads to:

$$\frac{m!}{2} \cdot \sum_{\substack{1 \leq i, j \\ i+j \leq n}} \binom{n}{i} \cdot \binom{n-i}{j} (2^{m-1} - 2)^{n-i-j}, \quad (2)$$

where i is the number of times the vote V appears in the profile and j the number of times the vote \bar{V} appears. The other votes may be any of the $2^{m-1} - 2$ remaining votes that are single-peaked with respect to the axis V and $n-i-j$ of them must be chosen. By simple manipulations of equation (2) we obtain the lower bound $2^{(m-1) \cdot n} \cdot (1 - \epsilon(n, m))$ where $\epsilon(n, m) := (2 \cdot (2^{m-1} - 1)^n - (2^{m-1} - 2)^n) / 2^{(m-1)n}$. As can easily be seen, $\epsilon(n, m)$ tends to 0 for every fixed m and $n \rightarrow \infty$. \square

The proof for the upper bound on the number of single-crossing and 2D single-peaked profiles had to be omitted.

Theorem 6. *If $n, m \geq 2$ it holds that $a(n, m, SC) \leq$*

$$\leq \min \left(n!m!n^{m \cdot (m-1)/2}, \frac{(n + \binom{m}{2})! \cdot m!}{\prod_{i=0}^{m-2} (2i+1)^{m-1-i}} \right).$$

Theorem 7. *For $m > 4$ and $m \geq 2$ it holds that*

$$a(n, m, 2D) \leq m! \cdot \left(\frac{4^n}{24} \right)^{\lfloor \frac{m}{4} \rfloor} \cdot \left(\frac{(m-1)!}{\prod_{i=1}^{\lceil \frac{m}{4} \rceil - 1} (m-4i)} \right)^n.$$

Distances to Domain Restrictions

As mentioned in the introduction, domain restrictions are often too restrictive to describe real-world preference data. Notions of distance make domain restrictions more flexible. Here, we study three distances. The first is *voter deletion* (or *Maverick*) (Faliszewski, Hemaspaandra, and Hemaspaandra 2011) which is the number of voters that have to be removed from a profile for it to belong to a restricted domain. The second one is *candidate deletion* (Escoffier, Lang, and Öztürk 2008), where candidates instead of voters are removed from the domain of the profile. The third is *local candidate deletion* (Erdélyi, Lackner, and Pfandler 2013). Here we ask for the number of candidates that have to be removed per vote such that the corresponding (partial) profile belongs to a certain domain restriction.

We start with two upper bounds that are applicable to arbitrary domain restrictions.

Theorem 8. *Let $v(n, m, k, \Gamma)$ denote the number of profiles that have a voter deletion distance of at most k to the set of (n, m) -profiles avoiding Γ . It holds that $v(n, m, k, \Gamma) \leq \binom{n}{k} \cdot (m!)^k \cdot a(n-k, m, \Gamma)$.*

Theorem 9. *Let $c(n, m, k, \Gamma)$ denote the number of profiles that have a candidate deletion distance of at most k to the set of (n, m) -profiles avoiding Γ . It holds that*

$$c(n, m, k, \Gamma) \leq \left(\frac{m!}{(m-k)!} \right)^n \cdot a(n, m-k, \Gamma).$$

The next theorem tells us how many local candidate deletions are needed at most to make a profile single-peaked.

Theorem 10. *Let u be a positive integer such that $m \geq u \cdot (u-1)/2 + 1$. For every axis A and every vote V on m candidates there is a subset $S \subseteq \text{dom}(V)$ of size at least u such that $V[S]$ is single-peaked with respect to A .*

m	3	4	5	6	7	8	9	10	11	12
SP	1	1	2	3	3	4	5	6	6	7
$2D$	0	0	1	1	2	2	3	4	4	5

Table 1: Maximal number of candidates that have to be deleted per vote in order to make an arbitrary profile with m candidates (2D) single-peaked.

This theorem is a refinement of the famous Erdős-Szekeres theorem which states that every sequence of length at least $(r-1)(s-1)+1$ contains a subsequence of length r that is monotonically increasing or a subsequence of length s that is monotonically decreasing. Our result implies that at most $\lceil m - \frac{1}{2}(\sqrt{8m-7}+1) \rceil$ candidates have to be locally deleted in an (n, m) -profile to make it single-peaked.

This theorem can be extended to 2D single-peaked with the help of the following result:

Proposition 11. *Let \mathcal{P} be a profile. If there exist disjoint sets C_1, C_2 with $C_1 \cup C_2 = \text{dom}(\mathcal{P})$ such that $\mathcal{P}[C_1]$ as well as $\mathcal{P}[C_2]$ are single-peaked then \mathcal{P} is 2D single-peaked.*

We can use Theorem 10 and Proposition 11 to compute the maximum local candidate deletion distance of any profile with m candidates to the (2D) single-peaked restriction. The results for profiles with few candidates are exemplarily shown in Table 1. In the experimental study of Sui, Francois-Nienaber, and Boutilier (2013) they find that a (3800, 9)-profile is 3-Local Candidate Deletion 2D Single-Peaked. Our results show that this is necessarily the case for every profile with 9 candidates (cf. Table 1).

Conclusions

At a first glance, our results seem to have a negative flavor since they show that random profiles are unlikely to belong to a restricted domain. However, this can also be interpreted in a positive way: If structure is found in a profile, this is almost certainly not the mere product of chance. For example, Sui, Francois-Nienaber, and Boutilier (2013) studied a (3800, 9)-profile that contained a subset of 2498 votes which were 2D single-peaked. Our results show that the probability of such an event is less than $7.4 \cdot 10^{-61}$. Thus, it can be concluded that this structure is very unlikely to occur randomly.

Another positive aspect is in relation to manipulation and control. Although domain restrictions break the complexity barrier against undesired attacks, our results lead to the conclusion that even a small amount of randomness in a profile can yield protection.

We would like to mention one specific application of our paper. We provide means to compute the maximal amount of noise in a preference profile (e.g. uninformed or disinterested voters) such that there is still a reasonable chance for structure. While IC is not realistic for real-world preference profiles, we believe it is appropriate to model this kind of noise. Our results rigorously show how fragile notions of domain restrictions really are. This is relevant information since many algorithmic results in computational social choice assume restricted domains.

Let us conclude with directions for future research. First and foremost, our result have to be extended to other, more realistic probability distributions. We hope that our results may serve as a starting point for such investigations. It would also be interesting to complement our upper bound results with corresponding lower bounds, both for domain restrictions as well as for notions of distance. This would allow us to compute the guaranteed maximal distance to each domain restriction, similar to our results for the local candidate deletion distance.

Finally, the connection between configurations in profiles and patterns in permutations established in this paper cannot only be used to derive combinatorial results but also for algorithmic advances. Indeed, as preliminary studies of the authors show, this connection can be used to show that the CONFIGURATION CONTAINMENT problem, asking whether a configuration \mathcal{C} is contained in a profile \mathcal{P} , is NP-complete even if $|\mathcal{P}| = 2$ and $|\mathcal{C}| = 2$. It seems promising to adapt algorithms from permutation pattern matching to the CONFIGURATION CONTAINMENT problem, possibly yielding algorithms for structure detection in preferences that are not tailored to a single domain restriction.

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 by the Vienna Science and Technology Fund (WWTF) project ICT12-15.

References

- Arrow, K. J. 1950. A difficulty in the concept of social welfare. *The Journal of Political Economy* 58(4):328–346.
- Ballester, M. A., and Haeringer, G. 2011. A characterization of the single-peaked domain. *Social Choice and Welfare* 36(2):305–322.
- Barberà, S.; Gul, F.; and Stacchetti, E. 1993. Generalized median voter schemes and committees. *Journal of Economic Theory* 61(2):262–289.
- Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research (JAIR)* 47:475–519.
- Black, D. 1948. On the rationale of group decision making. *Journal of Political Economy* 56(1):23–34.
- Bóna, M. 2004. *Combinatorics of permutations*. Discrete Mathematics and Its Applications. Chapman & Hall/CRC.
- Brandt, F.; Brill, M.; Hemaspaandra, E.; and Hemaspaandra, L. A. 2010. Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. In *Proc. of AAAI-10*, 715–722.
- Bredereck, R.; Chen, J.; and Woeginger, G. J. 2013a. Are there any nicely structured preference profiles nearby? In *Proc. of IJCAI-13*, 62–68.
- Bredereck, R.; Chen, J.; and Woeginger, G. J. 2013b. A characterization of the single-crossing domain. *Social Choice and Welfare* 41(4):989–998.
- Coombs, C. H. 1964. *A Theory of Data*.
- Cornaz, D.; Galand, L.; and Spanjaard, O. 2012. Bounded single-peaked width and proportional representation. In *Proc. of ECAI-12*, volume 242 of *FAIA*, 270–275. IOS Press.
- Cornaz, D.; Galand, L.; and Spanjaard, O. 2013. Kemeny elections with bounded single-peaked or single-crossing width. In *Proc. of IJCAI-13*. IJCAI/AAAI.
- Elkind, E., and Lackner, M. 2014. On detecting nearly structured preference profiles. In *Proc. of AAAI-2014*. To appear.
- Elkind, E.; Faliszewski, P.; and Slinko, A. M. 2012. Clone structures in voters’ preferences. In *Proc. of EC-12*, 496–513. ACM.
- Erdélyi, G.; Lackner, M.; and Pfandler, A. 2013. Computational complexity of nearly single-peaked electorates. In *Proc. of AAAI-13*. AAAI Press.
- Escoffier, B.; Lang, J.; and Öztürk, M. 2008. Single-peaked consistency and its complexity. In *Proc. of ECAI-08*, volume 178 of *FAIA*, 366–370. IOS Press.
- Faliszewski, P.; Hemaspaandra, E.; Hemaspaandra, L. A.; and Rothe, J. 2011. The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. *Information and Computation* 209(2):89–107.
- Faliszewski, P.; Hemaspaandra, E.; and Hemaspaandra, L. A. 2010. Using complexity to protect elections. *Commun. ACM* 53(11):74–82.
- Faliszewski, P.; Hemaspaandra, E.; and Hemaspaandra, L. A. 2011. The complexity of manipulative attacks in nearly single-peaked electorates. In *Proc. of TARK-11*, 228–237.
- Knoblauch, V. 2010. Recognizing one-dimensional Euclidean preference profiles. *Journal of Mathematical Economics* 46(1):1–5.
- Marcus, A., and Tardos, G. 2004. Excluded permutation matrices and the stanley–wilf conjecture. *Journal of Combinatorial Theory, Series A* 107(1):153–160.
- Mattei, N.; Forshee, J.; and Goldsmith, J. 2012. An empirical study of voting rules and manipulation with large datasets. In *Proc. of COMSOC-12*.
- Popova, A.; Regenwetter, M.; and Mattei, N. 2013. A behavioral perspective on social choice. *Annals of Mathematics and Artificial Intelligence* 68(1-3):5–30.
- Roberts, K. W. 1977. Voting over income tax schedules. *Journal of Public Economics* 8(3):329–340.
- Rothe, J., and Schend, L. 2013. Challenges to complexity shields that are supposed to protect elections against manipulation and control: a survey. *Annals of Mathematics and Artificial Intelligence* 68(1-3):161–193.
- Skowron, P.; Yu, L.; Faliszewski, P.; and Elkind, E. 2013. The complexity of fully proportional representation for single-crossing electorates. 1–12.
- Sui, X.; Francois-Nienaber, A.; and Boutilier, C. 2013. Multi-dimensional single-peaked consistency and its approximations. In *Proc. of IJCAI-13*. IJCAI/AAAI.
- Walsh, T. 2007. Uncertainty in preference elicitation and aggregation. In *Proc. of AAAI-07*, 3–8. AAAI Press.