

# How Hard Is the Manipulative Design of Scoring Systems?\*

Dorothea Baumeister and Tobias Högbe

Institut für Informatik, Heinrich-Heine-Universität Düsseldorf

{d.baumeister, tobias.hogbe}@uni-duesseldorf.de

## Abstract

In an election, votes are often given as ordered lists over candidates. A common way of determining the winner is then to apply some scoring system, where each position is associated with a specific score. This setting is also transferable to other situations, such as sports tournaments. The design of such systems, i.e., the choice of the score values, may have a crucial influence on the outcome. We study the computational complexity of two related decision problems. In addition, we provide a case study of data from Formula 1 using ILP formulations. Our results show that under some mild conditions there are cases where the actual scoring system has no influence, whereas in other cases very small changes may lead to a different winner. This may be seen as a measure of robustness of the winning candidate.

## 1 Introduction

A common task in diverse situations is to choose a winner among a group of alternatives. If the underlying information consists of linear orders over the alternatives a usual approach is to use scoring systems. Here, each position is associated with a respective score value and the candidates with the highest scores are the winners. This approach is not only useful for elections in the classical sense, but also for other situations such as sports competitions. A Formula 1 season can be seen as an election where the candidates correspond to the drivers and the voters represent the placements for each race. Another example of a competition using scoring systems is the Eurovision Song Contest.

When designing such a scoring system some requirements obviously have to be respected. For example the points should be non-increasing from the first to the last place, and when considering sports tournaments it is often reasonable to have a significant distance between the front positions. Taking the scoring systems for the Formula 1 as an example, it can be observed that there were several changes over the years. When a lot of money and/or prestige is involved, it is

\*A preliminary version of this paper appeared as an extended abstract in AAMAS2019.

particularly important to examine the manipulative design of scoring systems, since an external agent may therefore have an influence on manipulating the outcome.

We investigate this problem from two different perspectives. On the one hand we provide a theoretical background for the manipulative design of scoring systems and on the other hand we give an experimental evaluation of data from the Formula 1. We introduce the decision problem SCORING SYSTEM EXISTENCE which captures the simple setting where a manipulator seeks for a scoring system that makes some distinguished candidate win, regardless of any external requirements on the scores. As argued above, the existence of only weird scoring systems may not help, as there is no chance to implement them. So we additionally consider restrictions on the system. A slightly different approach is captured by the problem CLOSE SCORING SYSTEMS, where the choice of a new system is bounded by some given distance. This is especially important when an existing system will be changed since in most cases small changes will be accepted more easily. We study the computational complexity of both problems in several variants. Our results range from hardness results for NP and W[2] to polynomial-time algorithms that are obtained through linear programming approaches. In addition to the theoretical results, we also provide a case study on Formula 1 data. For all seasons from 1961 to 2008, we checked whether there exist scoring systems that result in a different winner. We show that there are years with perfect winners, i.e., the current winner is the only possible unique winner under every other scoring system that fulfills some very mild criteria, whereas in other years very small changes may already lead to a different winner. Hence the obtained results are also a measure of robustness of the actual winner.

## 2 Preliminaries

Formally an election is a pair  $(C, V)$ , with  $C$  being the set of candidates and  $V = (v_1, \dots, v_n)$  a profile of voters consisting of  $n$  linear orders over  $C$ . A scoring system used to define the winner of an election with  $m$  candidates is defined through the vector  $\vec{\alpha} = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}_{\geq 0}^m$  with  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_m$  and  $\alpha_1 > \alpha_m$ . By  $\text{score}_{(C,V)}^{\vec{\alpha}}(c) = \sum_{i=1}^n \alpha_{\text{pos}(v_i, c)}$  we denote the number of points for candidate  $c$  in election  $(C, V)$ , where  $\text{pos}(v, c)$  is the position of candidate  $c$  in the linear order  $v$ . The candidates having

the highest score are the winners. Prominent examples include the plurality rule, where only the first-placed candidate in each vote gets one point and thus has the scoring vector  $(1, 0, \dots, 0)$ , or the Borda rule defined through the vector  $(m-1, m-2, \dots, 0)$ . Beside usual elections, scoring systems are also used in competitions like the Eurovision Song Contest or the Formula 1. In Section 5 the data from Formula 1 races will be considered in detail with respect to the problems studied in this paper.

The score values are often restricted to natural numbers, however, we also consider the more general case of real numbers or rational numbers. We will show that this choice is crucial for determining the computational complexity. To facilitate our analysis we will use the following lemma that characterizes equivalent scoring rules.

**Lemma 1.** *A scoring system  $\vec{\alpha} \in \mathbb{R}_{\geq 0}^m$  is equivalent to a scoring system  $\vec{\alpha}' \in \mathbb{R}_{\geq 0}^m$  in terms of winner determination if and only if  $\vec{\alpha}' = a \cdot \vec{\alpha} + b$  with  $a \in \mathbb{R}_{> 0}$  and  $b \in \mathbb{R}$  holds.*

Lemma 1 implies that every scoring system can be transformed into an equivalent scoring system with  $\alpha_m = 0$ . Therefore, we will generally assume that  $\alpha_m = 0$  holds. In our proofs we need to construct profiles that satisfy certain criteria on the scores of the candidates. This can be achieved by the following lemma. This is a slightly adapted version of the lemma by Betzler and Dorn [2010] (see also [Dey and Misra, 2017]), which is formulated over score differences rather than absolute scores. For a given profile  $V$  over  $C$  and a scoring vector  $\vec{\alpha} \in \mathbb{R}_{\geq 0}^{|C|}$  we define the score difference between two candidates  $a, b \in C$  as  $\text{sd}_{(V,C)}^{\vec{\alpha}}(a, b) = \text{score}_{(V,C)}^{\vec{\alpha}}(a) - \text{score}_{(V,C)}^{\vec{\alpha}}(b)$ .

**Lemma 2.** *Given a set of candidates  $C = \{c_1, c_2, \dots, c_{m-2}, p, e\}$  and an integer matrix  $H = (h_{i,j})_{i=1, \dots, m-2, j=1, \dots, m-1}$  with  $H \in \mathbb{Z}^{(m-2) \times (m-1)}$ . One can construct a profile  $V = (v_1, \dots, v_n)$  such that evaluated by any scoring system  $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{m-1}, 0)$  is satisfying  $\text{sd}_{(V,C)}^{\vec{\alpha}}(e, p) < 0$  and  $\text{sd}_{(V,C)}^{\vec{\alpha}}(c_i, p) = [h_{i,1} \cdot \alpha_1 + \dots + h_{i,m-1} \cdot \alpha_{m-1}]$  for  $i \in \{1, \dots, m-2\}$ , with  $n \in \mathcal{O}(\hat{h} \cdot m^3)$  and  $\hat{h} = \max_{i \in \{1, \dots, m-2\}, j \in \{1, \dots, m-1\}} |h_{i,j}|$ .*

We assume that the reader is familiar with the basics of computational complexity, such as the classes P and NP. For further information we refer to the textbooks by Papadimitriou [1994] and Arora and Barak [2009]. In addition to classical complexity, we also study the parameterized complexity and obtain FPT results and W[2]-hardness. For further reading we refer to the textbooks by Downey and Fellows [1999] and by Flum and Grohe [2006].

In our constructions, we will make use of the NP-complete problem DOMINATING SET. An instance consists of an undirected graph  $G = (X, E)$  and an integer  $s \leq |X|$ . The question is whether there exists a subset  $X' \subseteq X$  with  $|X'| \leq s$  such that each  $x_j \in X$  is either a member of  $X'$  or shares an edge with a member of  $X'$ . This problem is W[2]-hard with respect to the parameter  $s$ , the size of the dominating set (see [Downey and Fellows, 1999]). For our reductions, we will construct ILPs that are similar to the following direct ILP formulation of DOMINATING SET. Given a DOMINATING SET

instance consisting of  $G = (X, E)$  with  $X = \{x_1, \dots, x_\ell\}$  and integer  $s \leq |X|$ . Let  $N[x_j]$  denote the closed neighborhood of  $x_j$ . An equivalent ILP is given as follows.

$$\forall x_j \in X : y_j \in \{0, 1\} \quad (1)$$

$$\forall x_j \in X : \sum_{x_i \in N[x_j]} y_i \geq 1 \quad (2)$$

$$\sum_{x_j \in X} y_j \leq s \quad (3)$$

### 3 Scoring System Existence

As an election problem, we model the situation where the chair of the election seeks an alternative scoring system that guarantees the victory of a distinguished candidate for different possible profiles. This is, for example, the case if different opinion polls are known but there is some uncertainty about the real votes. A different application is the evaluation of (online) surveys, where the voting system used is initially unknown to the participants. The initiator may afterwards try to adjust the scores in order to ensure a victory for her preferred candidate or to support her hypothesis. To formally capture the idea of manipulation through the choice of the scoring vector we introduce the problem SCORING SYSTEM EXISTENCE.

SCORING SYSTEM EXISTENCE	
<b>Given:</b>	A set of candidates $C$ with $ C  = m$ , a list of profiles $V_1, V_2, \dots, V_N$ over $C$ , and a candidate $p \in C$ .
<b>Question:</b>	Is there a scoring vector $\vec{\alpha} \in \mathbb{R}_{\geq 0}^m$ with $\alpha_m = 0$ , such that $p$ is the unique winner of election $(C, V_j)$ , $1 \leq j \leq N$ , with respect to $\vec{\alpha}$ ?

We focus on the unique winner case. The non-unique winner case is less interesting, as vectors that assign nearly the same value to all positions are often a solution in the unconstrained case. In general, the approaches in the constrained case are technically more interesting for the unique winner case and can easily be broken down for the non-unique case. This decision problem is contained in NP for all variants considered here since for a given scoring vector the election can be evaluated efficiently. Therefore, we will only refer to the hardness in the later NP-completeness proofs.

Even though the formal definition uses the notions of elections, this problem also captures completely different settings, for example the Formula 1 or the Eurovision Song Contest. The task is again to figure out whether a change of the system, based on suspected placements or a subsequent change based on the actual placements, can guarantee certain outcomes. This question is particularly interesting if people in charge might have an interest in the success of a certain team because of preferences, arrangements, or bets. In Section 5 we will use real-world data to show that it is not unlikely in regular competitions to be able to replace the unique winner by changing the system only slightly. In all these cases, one has to keep in mind that arbitrary systems, which follow no idea or intuition, are suspicious and make a targeted modification of the system very obvious. Therefore, we consider certain restrictions of the scoring systems.

**Related work.** Our problem belongs to the group of election problems studied in computational social choice. Comparing it to the existing election problems, it is much closer to the different variants of control (see Bartholdi et al. [1992]), than to manipulation (see Bartholdi et al. [1989]) or bribery (see Faliszewski et al. [2009]), since the influence is on the chair and the election itself rather than on the voters or candidates. Baumeister et al. [2011] considered a special case of our problem, showing that SCORING SYSTEM EXISTENCE for one profile is NP-complete for  $(\alpha_1, \dots, \alpha_{m-4}, x_1, x_2, x_3, 0)$  with  $x_i = 1$  for at least one  $i \in \{1, 2, 3\}$ , if we assume that the votes are given in succinct representation, i.e., instead of listing each vote separately the distinct votes are given with their corresponding multiplicity. We will strengthen this result by showing that even without the requirement of succinct representation, the problem is NP-complete for a variety of slightly constrained vectors. On the other hand, our results imply that their problem for unconstrained vectors can be solved efficiently. Elkind and Erdélyi [2012] considered a related problem, in which the voting rule to be used will be chosen from a given set of rules, and the aim of the manipulator is to find a vote that ensures the victory of some distinguished candidate for all of the possible voting rules. Shiryayev et al. [2013] and Bredereck et al. [2017] studied the robustness of an election taking into account the minimum degree of change necessary regarding the votes to alter the outcome of the election. In contrast, we measure the amount of change in the scoring vector needed by using distances.

## Complexity

We briefly outline the scenarios examined in the following. First, we consider the unrestricted case. Then, we investigate the complexity of vectors restricted to natural numbers using one of the simplest constraints: requiring a certain value at a given rear/front position in the vector. Our hardness results are formulated in the most restricted case and therefore also show hardness for the most general cases, e.g. where either the value, the position, or both are part of the input. For this case we also show that the problem is FPT regarding the number of candidates or regarding the total number of voters across all profiles. The main results for SCORING SYSTEM EXISTENCE are summarized in Table 1.

The first result is that without restrictions it can be checked in polynomial-time whether an alternative scoring system exists that makes the distinguished candidate the unique winner of all the elections.

**Theorem 1.** SCORING SYSTEM EXISTENCE is in P for  $\vec{\alpha} \in \mathbb{R}_{\geq 0}^m, \mathbb{Q}_{\geq 0}^m$ , and  $\mathbb{N}_{\geq 0}^m$ .

*Proof.* Let the profiles be  $V_j = (v_{j,1}, v_{j,2}, \dots, v_{j,n_j})$  for  $1 \leq j \leq N$ . From the definition of a scoring system and the properties of a scoring vector, the following linear program (LP) can be easily derived. This LP is solvable, if and only if there exists a system in which  $p$  becomes the unique winner with a minimum advantage of  $\epsilon > 0$  in each of the

elections.

$$\forall j \in \{1, \dots, N\} : \forall q \in C \setminus \{p\} :$$

$$\sum_{k=1}^m [T_{(C,V_j)}(p, k) - T_{(C,V_j)}(q, k)] \cdot \alpha_k \geq \epsilon \quad (1)$$

$$\text{with } T_{(C,V_j)}(c, k) = |\{i \in \{1, \dots, n_j\} \mid \text{pos}(v_{j,i}, c) = k\}|.$$

$$\forall k \in \{1, \dots, m-1\} : \alpha_k - \alpha_{k+1} \geq 0 \quad (2)$$

$$\alpha_m = 0 \quad (3)$$

The choice of  $\epsilon$  restricts the possible scoring vectors, but nevertheless, in order to check whether any vector exists at all, the actual value of  $\epsilon$  is irrelevant: As the advantage is proportional to the scaling of the vector, by Lemma 1 we can always obtain an equivalent vector guaranteeing a minimum advantage of  $\epsilon$ , if one exists at all. Therefore,  $\epsilon$  can be chosen arbitrarily and we choose  $\epsilon = 1$  by default.

LPs can be solved efficiently depending on the number of variables and on the length of the input (see [Karmarkar, 1984] and [Adler et al., 1989]). The output of these algorithms is always rational for rational inputs, whereby a rational scoring vector exists if there exists one at all. Therefore, by applying Lemma 1, we can also receive an equivalent integer scoring vector by scaling if there exists one at all.  $\square$

Through Lemma 1, rational vectors can always be reshaped to contain a certain value at a given position. However this does not hold for integer vectors, and this is what we now focus on. In this case, SCORING SYSTEM EXISTENCE is already getting hard if one specific value on some fixed rear position is required in the vector. Note that the position, as well as the specific value, are independent of the actual input.

**Theorem 2.** Let  $k \in \mathbb{N}_{\geq 1}$  and  $\gamma \in \mathbb{N}_{\geq 1}$  be some fixed values. SCORING SYSTEM EXISTENCE is NP-complete if the scoring vector  $\vec{\alpha} \in \mathbb{N}_{\geq 0}^m$  for  $m > k$  candidates has to satisfy  $\alpha_{m-k} = \gamma$ .

*Proof.* NP-hardness will be shown by a reduction from DOMINATING SET. Assume we are given a DOMINATING SET instance  $G = (X, E)$  with  $X = \{x_1, x_2, \dots, x_\ell\}$  and an integer  $s \leq |X|$ . We construct an election with the set of candidates  $C = \{p, e, g\} \cup A \cup B \cup D$ , where  $A = \{a_1, \dots, a_\ell\}$ ,  $B = \{b_1, \dots, b_\ell\}$ , and dummy candidates  $D = \{d_1, \dots, d_{k+1}\}$ . The preferred candidate is  $p$ . According to Lemma 2 we construct the profile  $V$  that ensures the following score differences, where we choose  $p$  and  $e$  respectively. For each  $a_j$ ,  $1 \leq j \leq \ell$ :  $\text{sd}_{(C,V)}^{\vec{\alpha}}(a_j, p) = [\gamma] \cdot \alpha_{(m-k)-j} - [\gamma] \cdot \alpha_{(m-k)-(j-1)} - [2] \cdot \alpha_{m-k}$ , for each  $b_j$ ,  $1 \leq j \leq \ell$ :  $\text{sd}_{(C,V)}^{\vec{\alpha}}(b_j, p) = [\chi_j(x_1)] \cdot \alpha_{(m-k)} - [\chi_j(x_\ell)] \cdot \alpha_{(m-k-\ell)} + \sum_{i=1}^{\ell-1} [\chi_j(x_{i+1}) - \chi_j(x_i)] \cdot \alpha_{(m-k)-i}$  with  $\chi_j$  denoting the characteristic function of  $N[x_j]$ , and  $\text{sd}_{(C,V)}^{\vec{\alpha}}(g, p) = [\gamma] \cdot \alpha_{(m-k)-\ell} - [s+1+\gamma] \cdot \alpha_{m-k}$ .

For each  $d_j \in D$  we set  $\text{sd}_{(C,V)}^{\vec{\alpha}}(d_j, p)$  to be negative. Therefore, for the candidates in  $\{e\} \cup D$  the score difference to  $p$  is less than zero, i.e., they have fewer points than  $p$  irrespective of the entries in the scoring vector.

We will now show that a scoring vector of the required form making  $p$  the unique winner exists for the constructed

$\vec{\alpha} \in \mathbb{N}_{\geq 0}^m$	$\alpha_{m-k} = \gamma$ , fixed rear position	$\alpha_k = \gamma$ , fixed front position
$k$ fixed, $\gamma$ fixed	NP-complete, Thm. 2	$\in \text{P}$ , Thm. 3
$k$ fixed, $\gamma$ input	NP-complete, Thm. 2	NP-c. / W[2]-hard wrt. $\gamma$ , Thm. 3
$k$ input, $\gamma$ fixed / input		NP-complete, Thm. 2
for all cases	FPT wrt. number of candidates or total number of voters, Thm. 4	

Table 1: Complexity results for SCORING SYSTEM EXISTENCE.

instance if and only if a dominating set with at most  $s$  vertices exists for the given instance. We begin with equivalences regarding the solvability of the instance. If  $p$  is the unique winner of the election, the score difference  $\text{sd}_{(C,V)}^{\vec{\alpha}}(c,p)$  for all candidates  $c \in C \setminus \{p\}$  has to be negative. This already holds for  $e$  and the candidates in  $D$ , hence we only have to focus on the score differences of the candidates in  $A \cup B \cup \{g\}$ . Recall that  $\gamma = \alpha_{m-k}$  holds. For each  $a_j$ ,  $1 \leq j \leq \ell$ , we have:  $0 > [\gamma] \cdot \alpha_{(m-k)-j} - [\gamma] \cdot \alpha_{(m-k)-(j-1)} - [2] \cdot \alpha_{m-k} \Leftrightarrow \alpha_{(m-k)-(j-1)} + 1 \geq \alpha_{(m-k)-j}$ . Taking into account the monotonicity, each  $\alpha_{(m-k)-j}$  can have exactly two distinct values:  $\alpha_{(m-k)-(j-1)}$  or  $\alpha_{(m-k)-(j-1)} + 1$ . For each  $b_j$ ,  $1 \leq j \leq \ell$ , we have:  $0 > [\chi_j(x_1)] \cdot \alpha_{(m-k)} - [\chi_j(x_\ell)] \cdot \alpha_{(m-k-l)} + \sum_{i=1}^{\ell-1} [\chi_j(x_{i+1}) - \chi_j(x_i)] \cdot \alpha_{(m-k)-i} \Leftrightarrow 1 \leq \sum_{x_i \in N[x_j]} (\alpha_{(m-k)-i} - \alpha_{(m-k)-(i-1)})$ . And for  $g$  we have:  $0 > [\gamma] \cdot \alpha_{(m-k)-\ell} - [s+1+\gamma] \cdot \alpha_{m-k} \Leftrightarrow s \geq \alpha_{(m-k)-\ell} - \alpha_{m-k}$ . To summarize, a scoring vector of the required form making  $p$  the unique winner exists for the constructed instance if and only if there exists a scoring vector fulfilling the following restrictions.

$$\forall j \in \{1, \dots, \ell\} : \alpha_{(m-k)-j} \in \{\alpha_{(m-k)-(j-1)}, \alpha_{(m-k)-(j-1)} + 1\} \quad (1)$$

$$\forall j \in \{1, \dots, \ell\} : 1 \leq \sum_{x_i \in N[x_j]} (\alpha_{(m-k)-i} - \alpha_{(m-k)-(i-1)}) \quad (2)$$

$$s \geq \alpha_{(m-k)-\ell} - \alpha_{m-k} \quad (3)$$

The values not considered so far, namely  $\alpha_1, \alpha_2, \dots, \alpha_{(m-k)-\ell-1}$  and  $\alpha_{(m-k)+1}, \alpha_{(m-k)+2}, \dots, \alpha_m$ , do not have an influence on the score differences between  $p$  and the candidates that could prevent  $p$  from being the unique winner, hence their actual value is insignificant as long as they comply with the monotonicity of the scoring vector. Note that the above set of inequalities is equivalent to the ILP formulation of DOMINATING SET given previously regarding feasibility: (1) ensures that every  $\alpha_{(m-k)-j}$  can have two different values depending on the previous values, namely  $\alpha_{(m-k)-(j-1)}$  indicating that  $x_j \notin X'$  or  $\alpha_{(m-k)-(j-1)} + 1$  indicating that  $x_j \in X'$ . (2) ensures for each vertex  $x_j$  either  $x_j$  or at least one of its neighbors is covered by  $X'$ . (3) with  $\alpha_{(m-k)-\ell} - \alpha_{m-k}$  equaling the number of values set to their upper values ensures that  $|X'| \leq s$  holds.

Therefore, a scoring vector of the required form making  $p$  the unique winner exists for the SCORING SYSTEM EXISTENCE instance if and only if a dominating set with at most  $s$  vertices exists for the DOMINATING SET instance.  $\square$

Note that in this proof only one profile is needed to establish NP-hardness of the problem. In sharp contrast to the previous result, we now show that if some value at some front position is fixed in advance the problem remains solvable in polynomial time. This includes, for example, the case where the maximum number of points to be given has been limited from the outset. However if the value is not fixed in advance but part of the input, this problem is W[2]-hard.

**Theorem 3.** *Let  $k \in \mathbb{N}_{\geq 1}$  and  $\gamma \in \mathbb{N}_{\geq 1}$  be some fixed values. SCORING SYSTEM EXISTENCE is in P if the scoring vector  $\vec{\alpha} \in \mathbb{N}_{\geq 0}^m$  for  $m \geq k$  candidates has to satisfy  $\alpha_k = \gamma$ . If  $\gamma$  is not fixed in advance, the problem is NP-complete and W[2]-hard with respect to  $\gamma$ .*

The proof is omitted due to space restrictions. For the hardness, the idea is to use additional candidates to enforce the conditions in which one can apply the reduction of Theorem 2. For this, we set  $\gamma$  to be  $s+1$  whereby the NP- and W[2]-hardness follows from the respective hardness of DOMINATING SET. Again, only one profile is necessary to establish the NP- and W[2]-hardness. Now, we will consider the parameterized complexity of the problem with respect to the number of candidates and voters. For this, we consider the more general case of the problem where a specific value and its position itself may be part of the input. Note that an efficiency result for this case also applies to the restricted cases studied before where  $k$  and  $\alpha_k$  or  $\alpha_{m-k}$  are fixed.

**Theorem 4.** SCORING SYSTEM EXISTENCE with  $\vec{\alpha} \in \mathbb{N}_{\geq 0}^m$ ,  $\alpha_k = \gamma \in \mathbb{N}_{\geq 1}$  and  $k \in \{1, \dots, m-1\}$  as part of the input is FPT with respect to the number of candidates or with respect to the total number of voters.

*Proof sketch.* Regarding the number of candidates, the corresponding ILP (see proof of Theorem 1, with  $\gamma$  as part of the input and  $\alpha_m = 0$ , contains only  $m-2$  variables to be determined.

Regarding the total number of voters, the proof is based on the fact that for the solvability of the instance, one only has to consider the values of the vector on which positions  $p$  has been placed in at least one of the votes. All other values can be decreased to  $\gamma, 0$ , or the next of those values, as this only reduces the score of candidates other than  $p$ . Therefore, we have to consider an ILP with at most  $n$  variables, with  $n$  being the total number of voters.

In both cases, ILPs are FPT regarding the number of variables (see [Lenstra, 1983]), the problem is FPT regarding the number of candidates or the total number of voters.  $\square$

Interestingly, one can show that in the case where the number of profiles is variable, but the number of voters per pro-

file is bounded by a constant depending on  $\gamma$ , the problem is again hard once you fix only one value, even if it is constant at a constant position from the bottom. This special case is motivated by the fact that in (sports) competitions the number of jurors per event or races per season is mostly constant.

## 4 Closest Scoring System

So far we considered only very mild restrictions on the scoring vectors. However, the manipulator may wish to limit the changes to the scores. Formally this can be modeled by measuring the difference between the original and the new scoring system by some appropriate distance. We will use the following *Minkowski distances* to compare scoring systems.

**Definition 1.** *The Manhattan distance  $\mathcal{D}^1$ , the Euclidean distance  $\mathcal{D}^2$ , and the Chebyshev distance  $\mathcal{D}^\infty$  for two vectors  $\vec{x}, \vec{y} \in \mathbb{R}^m$  with  $\vec{x} = (x_1, x_2, \dots, x_m)$  and  $\vec{y} = (y_1, y_2, \dots, y_m)$  are defined as follows:*

$$\mathcal{D}^1(\vec{x}, \vec{y}) = \sum_{j=1}^m |x_j - y_j|, \quad \mathcal{D}^2(\vec{x}, \vec{y}) = \sqrt{\sum_{j=1}^m (x_j - y_j)^2},$$

$$\text{and } \mathcal{D}^\infty(\vec{x}, \vec{y}) = \max_{j=1}^m |x_j - y_j|.$$

In terms of scoring systems, the Manhattan distance measures the total change at given points per position and thereby the overall adjustments to the vector. The Chebyshev distance takes into account only the largest deviation. With regard to this distance, it is advantageous to distribute the adjustments to the scoring vector. The Euclidean distance combines both approaches, measuring the total change while heavily weighting larger deviations of individual entries. This motivates again the minimization of the overall adjustments to the vector while distributing the adjustments to the entire vector.

The problem we will focus on asks whether there is an alternative scoring system making the distinguished candidate the unique winner, while the distance between the two vectors is bounded. For a distance  $\mathcal{D}$  on arbitrary sized vectors, the problem is formally defined as follows.

D-CLOSE SCORING SYSTEM	
<b>Given:</b>	A set of candidates $C$ with $ C  = m$ , a list of profiles $V_1, V_2, \dots, V_N$ over $C$ , a scoring system $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m) \in \mathbb{R}_{\geq 0}^m$ with $\alpha_m = 0$ , a distance limit $K \in \mathbb{R}_{> 0}$ , and a candidate $p \in C$ .
<b>Question:</b>	Is there a scoring system $\vec{\alpha}' \in \mathbb{R}_{\geq 0}^m$ with $\alpha'_m = 0$ , such that $p$ is the unique winner of election $(C, V_j)$ , $1 \leq j \leq N$ with respect to $\vec{\alpha}'$ and $\mathcal{D}(\vec{\alpha}, \vec{\alpha}') \leq K$ ?

Again, the problem  $\mathcal{D}$ -CLOSE SCORING SYSTEM is contained in NP, for all distances considered here. In the previous section we showed that the unrestricted variant of SCORING SYSTEM EXISTENCE can be solved in polynomial time. For  $\mathcal{D}$ -CLOSE SCORING SYSTEM this still holds if the vector consists of real numbers or rational numbers.

**Theorem 5.**  *$\mathcal{D}$ -CLOSE SCORING SYSTEM for  $\vec{\alpha}, \vec{\alpha}' \in \mathbb{Q}_{\geq 0}^m$  or  $\mathbb{R}_{\geq 0}^m$ , and Minkowski distances  $\mathcal{D}^1$ ,  $\mathcal{D}^2$ , and  $\mathcal{D}^\infty$  is in P.*

The proof modifies the LP from the proof of Theorem 1 but is omitted due to space restrictions. In the following, we will show that the problem is hard when the scoring vectors

are restricted to natural numbers. Nevertheless, if we want to solve the problem in practice, we can interpret the LP together with the distance optimization used in the proof of Theorem 5 as an ILP.

**Theorem 6.**  *$\mathcal{D}$ -CLOSE SCORING SYSTEM for  $\vec{\alpha}, \vec{\alpha}' \in \mathbb{N}_{\geq 0}^m$  and Minkowski distances  $\mathcal{D}^1$ ,  $\mathcal{D}^2$ , and  $\mathcal{D}^\infty$  is NP-complete.*

The proof is omitted due to space restrictions and its similarity to the proof of Theorem 2. For the instances constructed in the proof of Theorem 6 there does always exist a scoring vector making  $p$  the unique winner. We have shown that if the instance is not solvable with  $\alpha_{m-1} = 1$ , the necessary distance increases at least by a factor of  $\ell$ . Based on these two arguments, it can be shown that for the problem for  $\mathcal{D}^1$ ,  $\mathcal{D}^2$ , and  $\mathcal{D}^\infty$ , no efficient constant-factor approximation algorithm for the distance can exist, unless P = NP.

Now, we present an analysis of the parameterized complexity of the problem with respect to the distance limit. Note that the proof of Theorem 6 includes that the problem with respect to  $\mathcal{D}^\infty$  is NP-complete even for a constant distance limit of  $K = 1$ . Therefore we will now focus on the parameterized complexity of the problem regarding  $\mathcal{D}^1$  and  $\mathcal{D}^2$ .

**Theorem 7.** *Let  $K \in \mathbb{R}_{> 0}$  be some fixed value.  $\mathcal{D}$ -CLOSE SCORING SYSTEM for  $\vec{\alpha}, \vec{\alpha}' \in \mathbb{N}_{\geq 0}^m$  and Minkowski distances  $\mathcal{D}^1$  and  $\mathcal{D}^2$  is in P. If  $K$  is not fixed in advance, the problem is W[2]-hard with respect to  $K$ .*

*Proof sketch.* The proof of W[2]-hardness is already included in the proof of Theorem 6, as DOMINATING SET is W[2]-hard with respect to  $s$  and we used  $K = s$  for  $\mathcal{D}^1$  and  $\mathcal{D}^2$ . It remains to prove that the problem can be solved efficiently for a constant  $K$ . It can be shown that for a constant  $K$ , the number of vectors to be considered within  $\mathcal{D}^1$  or  $\mathcal{D}^2$  distance  $K$  is bounded by a polynomial depending on  $m$ . Since those vectors can be efficiently generated, one can check for all the vectors, whether they are monotone and non-negative, and if so whether  $p$  becomes the unique winner by one of them.  $\square$

## 5 Experiments

In addition to the theoretical results presented so far, we investigated the manipulative design of scoring systems in a case study. Our aim was to identify how relevant the problem presented here is, in other words, how likely it is that with small changes to the system, one can change the outcome of an election or competition.

We consider real-life competition data from the Formula 1 from the seasons 1961 to 2008. To be precise, we use the version of the data available at PrefLib.org [Mattei and Walsh, 2013]. There exists a list of race results for each season, with each race result consisting of a strict order of drivers who have completed the race with all other drivers being tied on the last position. Since we examine fixed results here, not predictions, each instance contains exactly one profile. In the provided seasons the Formula 1 used the following scoring systems:  $\vec{\alpha} = (9, 6, 4, 3, 2, 1, 0, \dots, 0)$  in 1961–1990,  $(10, 6, 4, 3, 2, 1, 0, \dots, 0)$  in 1991–2002, and  $(10, 8, 6, 5, 4, 3, 2, 1, 0, \dots, 0)$  in 2003–2008.

Until 1991, not a stand-alone scoring system was used, as only a certain number of the best results of a driver were

	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
(I.)	12	-	-	2	2	12	2	8	21	3	19	11	20	2	34	2	10	5	3	10	1	4	2	2
(II.)	-	-	-	2	2	-	2	-	-	3	-	-	-	2	-	2	10	5	3	10	1	4	2	2
(III.)	-	-	-	8	4	-	5	-	-	-	-	-	-	4	-	2	22	-	16	-	6	9	2	-

  

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
(I.)	25	2	3	4	7	5	-	22	-	1	17	6	2	12	2	4	29	-	1	7	40	40	1	1
(II.)	-	2	3	4	7	5	-	-	-	1	-	6	2	-	2	4	-	-	1	7	-	-	1	1
(III.)	-	2	4	9	10	-	-	-	-	1	-	-	2	-	4	-	-	-	3	-	-	-	3	2

Table 2: Results for the case study of  $\mathcal{D}^1$ -CLOSE SCORING SYSTEM for Formula 1 data from 1961–2008. The table lists the minimum distance to a scoring vector that results in a different unique winner with respect to the restrictions (I.), (II.), and (III.).

counted. To keep the results consistent, we will however count all races of a season. In addition, we will also not consider special events such as exclusions or penalties. Thus, we treat the data as it is. The data-set contains 48 seasons, with the number of drivers (candidates) varying between 22 and 62, and the number of races (voters) between 8 and 19.

We used the LP formulation presented in the proof of Theorem 1 to test whether it is possible to change the current winner by using a different integer scoring system. We proposed several natural restrictions on the possible vector while minimizing its Manhattan distance  $\mathcal{D}^1$  to the original system.

In Table 2 we list the minimum  $\mathcal{D}^1$  distance from  $\vec{\alpha}$  to a vector  $\vec{\alpha}'$  with another unique winner, and “-” denotes that it is not possible. For this we consider three different strong natural restrictions on the vector.

- (I.)  $\alpha'_j = 0$  for each position  $j \in \{1, \dots, m\}$  that is tied with the last position at least once in the given profile.
- (II.)  $\alpha'_j = 0$  for each position  $j \in \{1, \dots, m\}$  with  $\alpha_j = 0$ .
- (III.)  $\alpha'_j - \alpha'_{j+1} \geq \alpha'_{j+1} - \alpha'_{j+2}$  for each position  $j \in \{1, \dots, m-2\}$  and (II.).

Restriction (I.) is a prerequisite and only affects the last positions. All drivers who are tied on the last position must receive 0 points. Restriction (II.) ensures that candidates who get no points in the original system should not get points in the new system. Restriction (III.) enforces that the point differences in between the top positions are non-decreasing. All of these properties are derived from scoring systems used in practice and are basic properties that an appropriate scoring system should fulfill. For the exact solution, we have implemented the LP from the proof of Theorem 1 as an ILP with  $\epsilon = 1$  including the  $\mathcal{D}^1$  formulation from Theorem 5 and additionally added requirement (I.), (II.), or (III.). The ILP was then solved using CPLEX, with the solution requiring only a fraction of a second per profile on a standard machine.

Considering almost unrestricted scoring systems by only requiring property (I.), the unique winner could be replaced in 43 out of 48 seasons. The maximum number of possible other unique winners was 6. Often, drivers who had no chance of winning in the original system could become the unique winner in a system with widely distributed points. Interestingly, even in this unrestricted case, in some years, the winner is

determined solely by the decision to use a scoring system, regardless of what the actual system is. For example, in 1991, 1993 and 2002, there was no other system with a unique winner apart from Senna, Prost, and Schumacher respectively. We call these winners *perfect winners* because they do not depend on the system itself, but only on the minimal requirement from restriction (I.). A weakening of this property is the perfect winner regarding certain restrictions. For example, in 1998, as soon as one demands that only the first 6 drivers get points at all, i.e., we consider restriction (II.), Häkkinen was the only possible unique winner. If restriction (III.) is demanded, in 27 of the 48 seasons the unique winners are already fixed without considering the actual values. Nevertheless, some of the results were highly dependent on the system used. In 2008, Massa would have been the unique winner if  $(12, 8, 6, 5, 4, 3, 2, 1, 0, \dots, 0)$  would have been used instead of  $(10, 8, 6, 5, 4, 3, 2, 1, 0, \dots, 0)$  highlighting the first place just slightly more, similar to as it has been done in the previous systems. On the other hand, in 2003 and 2007 the unique winner would have been Räikkönen and Alonso respectively if  $(8, 7, 6, 5, 4, 3, 2, 1, 0, \dots, 0)$  would have been used. To summarize, the experiments show that there is a potential for the manipulative design of scoring systems one should be aware of.

## 6 Conclusions

In this paper, we provide a general framework to study the manipulative design of scoring systems in diverse situations. A direct extension of the problem would be to consider distributions of profiles based on predictions in which we want to maximize the chances of success. Furthermore, an axiomatic and game theoretical investigation of the problem is very interesting. For the above-mentioned possibilities, the basic complexity usually follows directly from the results presented here. Finally, we assume that with the increasing convergence of performances in professional sports, the influence of the actual scoring system on the outcome of the competitions grows (see [Berthelot *et al.*, 2015]). This also makes the competitions more vulnerable.

## Acknowledgments

This work is supported by the DFG-grant BA6270/1-1.

## References

- [Adler *et al.*, 1989] I. Adler, M. G. C. Resende, G. Veiga, and N. Karmarkar. An implementation of Karmarkar’s algorithm for linear programming. *Mathematical Programming*, 44(1-3):297–335, 1989.
- [Arora and Barak, 2009] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
- [Bartholdi *et al.*, 1989] J. Bartholdi, C. Tovey, and M. Trick. The Computational Difficulty of Manipulating an Election. *Social Choice and Welfare*, 6(3):227–241, 1989.
- [Bartholdi III *et al.*, 1992] J. Bartholdi III, C. Tovey, and M. Trick. How Hard is it to Control an Election? *Mathematical Comput. Modelling*, 16(8/9):27–40, 1992.
- [Baumeister *et al.*, 2011] D. Baumeister, M. Roos, and J. Rothe. Computational Complexity of Two Variants of the Possible Winner Problem. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 853–860. IFAAMAS, 2011.
- [Berthelot *et al.*, 2015] G. Berthelot, A. Sedeaud, A. Marck, J. Antero-Jacquemin, J. Schipman, G. Saulière, A. Marc, F.-D. Desgorces, and J.-F. Toussaint. Has athletic performance reached its peak? *Sports Medicine*, 45(9):1263–1271, 2015.
- [Betzler and Dorn, 2010] N. Betzler and B. Dorn. Towards a Dichotomy of Finding Possible Winners in Elections Based on Scoring Rules. *Journal of Computer and System Sciences*, 76(8):812–836, 2010.
- [Bredereck *et al.*, 2017] R. Bredereck, P. Faliszewski, A. Kaczmarczyk, R. Niedermeier, P. Skowron, and N. Talmon. Robustness Among Multiwinner Voting Rules. In *Proceedings of the 10th International Symposium on Algorithmic Game Theory*, pages 80–92. Springer International Publishing, 2017.
- [Dey and Misra, 2017] P. Dey and N. Misra. On the Exact Amount of Missing Information that Makes Finding Possible Winners Hard. In *42nd International Symposium on Mathematical Foundations of Computer Science*, 2017.
- [Downey and Fellows, 1999] R. Downey and M. Fellows. *Parameterized Complexity*. Springer-Verlag, 1999.
- [Elkind and Erdélyi, 2012] E. Elkind and G. Erdélyi. Manipulation Under Voting Rule Uncertainty. In *Proceedings of the 11th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 627–634. IFAAMAS, 2012.
- [Faliszewski *et al.*, 2009] P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. How Hard Is Bribery in Elections? *Journal of Artificial Intelligence Research*, 35:485–532, 2009.
- [Flum and Grohe, 2006] J. Flum and M. Grohe. *Parameterized Complexity Theory*. EATCS Texts in Theoretical Computer Science. Springer-Verlag, 2006.
- [Karmarkar, 1984] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.
- [Lenstra, 1983] H. W. Lenstra. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):538–548, 1983.
- [Mattei and Walsh, 2013] N. Mattei and T. Walsh. PrefLib: A Library for Preferences <http://www.preflib.org>. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory*, pages 259–270. Springer-Verlag Lecture Notes in Artificial Intelligence #8176, 2013.
- [Papadimitriou, 1994] C. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.
- [Shiryaev *et al.*, 2013] D. Shiryaev, L. Yu, and E. Elkind. On elections with robust winners. In *Proceedings of the 13th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 415–422. IFAAMAS, 2013.