International Game Theory Review
Vol. 21, No. 1 (2019) 1940002 (15 pages)

DOI: 10.1142/S0219198919400024



# Some Impossibilities of Ranking in Generalized Tournaments

László Csató

*Research Group of Operations Research and Decision Systems*
*Laboratory on Engineering and Management Intelligence*
*Institute for Computer Science and Control, Hungarian*
*Academy of Sciences, 1111 Budapest*
*Kende Street 13-17., Hungary*

*Department of Operations Research and Actuarial Sciences*
*Corvinus University of Budapest, 1093 Budapest*
*Fővám Square 13-15., Hungary*
*laszlo.csato@uni-corvinus.hu*

Received 13 October 2018
Revised 13 October 2018
Accepted 30 October 2018
Published 15 April 2019

In a generalized tournament, players may have an arbitrary number of matches against each other and the outcome of the games is measured on a cardinal scale with lower and upper bounds. An axiomatic approach is applied to the problem of ranking the competitors. Self-consistency (SC) requires assigning the same rank for players with equivalent results, while a player showing an obviously better performance than another should be ranked strictly higher. According to order preservation (OP), if two players have the same pairwise ranking in two tournaments where the same players have played the same number of matches, then their pairwise ranking is not allowed to change in the aggregated tournament. We reveal that these two properties cannot be satisfied simultaneously on this universal domain.

*Keywords*: Tournament ranking; paired comparison; axiomatic approach; impossibility.

Subject Classification: 91B14

## 1. Introduction

This paper addresses the problem of tournament ranking when players may have played an arbitrary number of matches against each other, from an axiomatic point of view. For instance, the matches among top tennis players lead to a set of similar data: *Andre Agassi* has played 14 matches with *Boris Becker*, but he has never played against *Björn Borg* [Bozóki *et al.*, 2016]. To be more specific, we show the incompatibility of some natural properties. Impossibility theorems are well-known in the classical theory of social choice [Arrow, 1950; Gibbard, 1973; Satterthwaite, 1975], but our setting has a crucial difference: the set of agents and the set of alternatives coincide, therefore the transitive effects of "voting" should be considered [Altman and Tennenholtz, 2008]. We also allow for cardinal and incomplete preferences as well as ties in the ranking derived.

Several characterizations of ranking methods have been suggested in the literature by providing a set of properties such that they uniquely determine a given method [Rubinstein, 1980; Bouyssou, 1992; Bouyssou and Perny, 1992; van den Brink and Gilles, 2003, 2009; Slutzki and Volij, 2005, 2006; Kitti, 2016]. There are some excellent axiomatic analyses, too [Chebotarev and Shamis, 1998; González-Díaz *et al.*, 2014].

However, apart from Csató [2019b], we know only one work discussing impossibility results for ranking the nodes of a directed graph [Altman and Tennenholtz, 2008], a domain covered by our concept of generalized tournament. We think these theorems are indispensable for a clear understanding of the axiomatic framework. For example, González-Díaz *et al.* [2014] have found that most ranking methods violate an axiom called order preservation (OP), but it is not known whether this negative result is caused by a theoretical impossibility or it is only due to some unhidden features of the procedures that have been considered.

It is especially a relevant issue because of the increasing popularity of sports rankings [Langville and Meyer, 2012], which is, in a sense, not an entirely new phenomenon, since sports tournaments have motivated some classical works of social choice and voting theory [Landau, 1895; Zermelo, 1929; Wei, 1952]. For instance, the ranking of tennis players has been addressed from at least three perspectives, with the use of methods from multicriteria decision-making [Bozóki *et al.*, 2016], network analysis [Radicchi, 2011], or statistics [Baker and McHale, 2014, 2017]. Consequently, the axiomatic approach can be fruitful in the choice of an appropriate sports ranking method. This issue has been discussed in some recent works [Berker, 2014; Pauly, 2014; Csató, 2017, 2019a,c,d,e, 2018, 2019f; Dagaev and Sonin, 2018; Vaziri *et al.*, 2018; Vong, 2017], but there is a great scope for future research.

For this purpose, we will place two properties, imported from the social choice literature, in the center of the discussion. *Self-consistency* (SC) [Chebotarev and Shamis, 1997] requires assigning the same rank for players with equivalent results, furthermore, a player showing an obviously better performance than another should be ranked strictly higher. OP[a] [González-Díaz *et al.*, 2014] excludes the possibility of rank reversal by demanding the preservation of players' pairwise ranking when two tournaments, where the same players have played the same number of matches, are aggregated. In other words, it is not allowed that

[a]The term OP may be a bit misleading since it can suggest that the sequence of matches does not influence the rankings (see Vaziri *et al.* [2018, Property III]). This requirement obviously holds in our setting.

player $A$ is judged better both in the first and second halves of the season than player $B$, but ranked lower on the basis of the whole season.

Our main result proves the incompatibility of SC and OP. This finding gives a theoretical foundation for the observation of González-Díaz *et al.* [2014] that most ranking methods do not satisfy OP. Another important message of the paper is that prospective users cannot avoid to take similar impossibilities into account and justify the choice between the properties involved.

The study is structured as follows. Section 2 presents the notion of ranking problem and scoring methods. Section 3 introduces the property called SC and proves that one type of scoring methods cannot satisfy it. Section 4 defines (strong) OP besides some other properties, addresses the compatibility of the axioms and derives a negative result by opposing SC and OP. Section 5 summarizes our main findings.

## 2. The Ranking Problem and Scoring Methods

Consider a *set of players* $N = \{X_1, X_2, \dots, X_n\}$, $n \in \mathbb{N}_+$ and a series of *tournament matrices* $T^{(1)}$, $T^{(2)}, \dots, T^{(m)}$ containing information on the paired comparisons of the players. Their entries are given such that $t_{ij}^{(p)} + t_{ji}^{(p)} = 1$ if players $X_i$ and $X_j$ have played in round $p$ $(1 \leq p \leq m)$ and $t_{ij}^{(p)} + t_{ji}^{(p)} = 0$ if they have not played against each other in round $p$. The simplest definition can be $t_{ij}^{(p)} = 1$ (implying $t_{ji}^{(p)} = 0$) if player $X_i$ has defeated player $X_j$, and $t_{ij}^{(p)} = 0$ (implying $t_{ji}^{(p)} = 1$) if player $X_i$ has lost against player $X_j$ in round $p$. A draw can be represented by $t_{ij}^{(p)} = t_{ji}^{(p)} = 0.5$. The entries may reflect the scores of the players, or other features of the match (e.g., an overtime win has less value than a normal time win), too.

The tuple $(N, T^{(1)}, T^{(2)}, \dots, T^{(m)})$, denoted shortly by $(N, \mathbf{T})$, is called a *general ranking problem*. The set of general ranking problems with $n$ players $(|N| = n)$ is denoted by $\mathcal{T}^n$.

The *aggregated tournament matrix* $A = \sum_{p=1}^{m} T^{(p)} = [a_{ij}] \in \mathbb{R}^{n \times n}$ combines the results of all rounds of the competition.

The pair $(N, A)$ is called a *ranking problem*. The set of ranking problems with $n$ players $(|N| = n)$ is denoted by $\mathcal{R}^n$. Note that every ranking problem can be associated with several general ranking problems, in this sense, ranking problem is a narrower notion.

Let $(N, A), (N, A') \in \mathcal{R}^n$ be two ranking problems with the same player set $N$. The *sum* of these ranking problems is $(N, A + A') \in \mathcal{R}^n$. For example, the ranking problems can contain the results of matches in the first and second halves of the season, respectively.

Any ranking problem $(N, A)$ has a skew-symmetric *results matrix* $R = A - A^\top = [r_{ij}] \in \mathbb{R}^{n \times n}$ and a symmetric *matches matrix* $M = A + A^\top = [m_{ij}] \in \mathbb{N}^{n \times n}$. $m_{ij}$ is the number of matches between players $X_i$ and $X_j$, whose outcome is given by $r_{ij}$. Matrices $R$ and $M$ also determine the aggregated tournament matrix through

$A = (R + M)/2$, so any ranking problem $(N, A) \in \mathcal{R}^n$ can be denoted analogously by $(N, R, M)$ with the restriction $|r_{ij}| \leq m_{ij}$ for all $X_i, X_j \in N$. Despite description with results and matches matrices is not parsimonious, this notation will turn out to be useful.

A *general scoring method* is a function $g : \mathcal{T}^n \to \mathbb{R}^n$. Several procedures have been suggested in the literature, see Chebotarev and Shamis [1998] for an overview of them. A special type of general scoring methods is the following.

**Definition 1 (Individual scoring method [Chebotarev and Shamis, 1999]).** A general scoring method $g : \mathcal{T}^n \to \mathbb{R}^n$ is called *individual scoring method* if it is based on individual scores, that is, there exist functions $\phi$ and $\delta$ such that for any general ranking problem $(N, \mathbf{T}) \in \mathcal{T}^n$, the corresponding score vector $\mathbf{s} = g(N, \mathbf{T})$ can be expressed as $\mathbf{s} = \delta(\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(m)})$, where the partial score vectors $\mathbf{s}^{(p)} = \phi(N, T^{(p)})$ depend solely on the tournament matrix $T^{(p)}$ of round $p$ for all $p = 1, 2, \ldots, m$.

A *scoring method* is a function $f : \mathcal{R}^n \to \mathbb{R}^n$. Any scoring method can also be regarded as a general scoring method — by using the aggregated tournament matrix instead of the whole series of tournament matrices — therefore some papers only consider scoring methods [Kitti, 2016; Slutzki and Volij, 2005]. González-Díaz *et al.* [2014] give a thorough axiomatic analysis of certain scoring methods.

In other words, scoring methods initially aggregate the tournament matrices and then rank the players by their scores, while individual scoring methods first give scores to the players in each round and then aggregate them.

## 3. An Argument Against the Use of Individual Scoring Methods

In this section, some properties of general scoring methods are presented, which will highlight an important failure of individual scoring methods.

### 3.1. *Universal invariance axioms*

**Axiom 1 (Anonymity (ANO)).** Let $(N, \mathbf{T}) \in \mathcal{T}^n$ be a general ranking problem, $\sigma : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, m\}$ be a permutation on the set of rounds, and $\sigma(N, \mathbf{T}) \in \mathcal{T}^n$ be the ranking problem obtained from $(N, \mathbf{T})$ by permutation $\sigma$. General scoring method $g : \mathcal{T}^n \to \mathbb{R}^n$ is *anonymous* if $g_i(N, \mathbf{T}) = g_i(\sigma(N, \mathbf{T}))$ for all $X_i \in N$.

ANO implies that any reindexing of the rounds (tournament matrices) preserves the scores of the players.

**Axiom 2 (Neutrality (NEU)).** Let $(N, \mathbf{T}) \in \mathcal{T}^n$ be a general ranking problem, $\sigma : N \to N$ be a permutation on the set of players, and $(\sigma(N), \mathbf{T}) \in \mathcal{T}^n$ be the ranking problem obtained from $(N, \mathbf{T})$ by permutation $\sigma$. General scoring method $g : \mathcal{T}^n \to \mathbb{R}^n$ is *neutral* if $g_i(N, \mathbf{T}) = g_{\sigma(i)}(\sigma(N), \mathbf{T})$ for all $X_i \in N$.

NEU means that the scores are independent of the labeling of the players.

### 3.2. *Self-consistency*

Now we want to formulate a further requirement on the ranking of the players by answering the following question: *When is player $X_i$ undeniably better than player $X_j$?* There are two such plausible cases: (1) if player $X_i$ has achieved better results against the same opponents; (2) if player $X_i$ has achieved the same results against stronger opponents. Consequently, player $X_i$ should also be judged better if he/she has achieved better results against stronger opponents than player $X_j$. Furthermore, since (general) scoring methods allow for ties in the ranking, player $X_i$ should have the same rank as player $X_j$ if he/she has achieved the same results against opponents with the same strength.

In order to apply these principles, both the results and strengths of the players should be measured. Results can be extracted from the tournament matrices $T^{(p)}$. Strengths of the players can be obtained from their scores according to the (general) scoring method used, hence the name of the implied axiom is *self-consistency*. It has been introduced in Chebotarev and Shamis [1997], and extensively discussed by Csató [2019b].

**Definition 2 (Opponent multiset).** Let $(N, \mathbf{T}) \in \mathcal{T}^n$ be a general ranking problem. The *opponent multiset*[b] of player $X_i$ is $O_i$, which contains $m_{ij}$ instances of $X_j$.

Players of the opponent multiset $O_i$ are called the *opponents* of player $X_i$.

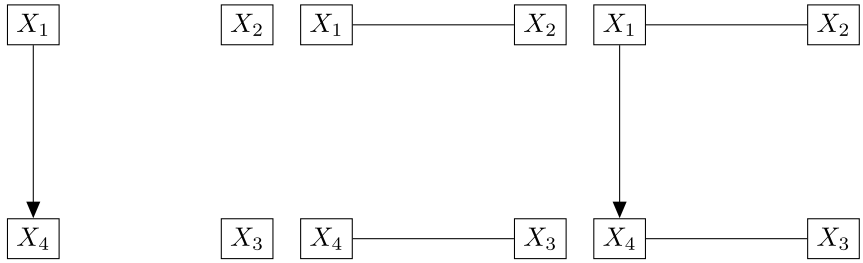
**Notation 1.** Consider the ranking problem $(N, T^{(p)}) \in \mathcal{T}^n$ given by restricting a general ranking problem to its $p$th round. Let $X_i, X_j \in N$ be two different players and $h^{(p)} : O_i^{(p)} \leftrightarrow O_j^{(p)}$ be a one-to-one correspondence between the opponents of $X_i$ and $X_j$ in round $p$, consequently, $|O_i^{(p)}| = |O_j^{(p)}|$. Then $\mathfrak{h}^{(p)} : \{k : X_k \in O_i^{(p)}\} \leftrightarrow \{\ell : X_\ell \in O_j^{(p)}\}$ is given by $X_{\mathfrak{h}^{(p)}(k)} = h^{(p)}(X_k)$.

**Axiom 3 (SC [Chebotarev and Shamis, 1997]).** A general scoring method $g : \mathcal{T}^n \to \mathbb{R}^n$ is called *self-consistent* if the following implication holds for any general ranking problem $(N, \mathbf{T}) \in \mathcal{T}^n$ and for any players $X_i, X_j \in N$: if there exists a one-to-one mapping $h^{(p)}$ from $O_i^{(p)}$ onto $O_j^{(p)}$ such that $t_{ik}^{(p)} \geq t_{j\mathfrak{h}^{(p)}(k)}^{(p)}$ and $g_k(N, \mathbf{T}) \geq g_{\mathfrak{h}^{(p)}(k)}(N, \mathbf{T})$ for all $p = 1, 2, \ldots, m$ and $X_k \in O_i^{(p)}$, then $f_i(N, R, M) \geq f_j(N, R, M)$, furthermore, $f_i(N, R, M) > f_j(N, R, M)$ if $t_{ik}^{(p)} > t_{j\mathfrak{h}^{(p)}(k)}^{(p)}$ or $g_k(N, \mathbf{T}) > g_{\mathfrak{h}^{(p)}(k)}(N, \mathbf{T})$ for at least one $1 \leq p \leq m$ and $X_k \in O_i^{(p)}$.

### 3.3. *Individual scoring methods and SC*

In this part, it will be proved that an anonymous and neutral individual scoring method cannot satisfy SC, which is a natural fairness requirement, thus it is enough

[b] *Multiset* is a generalization of the concept of set allowing for multiple instances of its elements.




(a) $(N, T^{(1)})$ (b) $(N, T^{(2)})$ (c) $(N, \mathbf{T})$

Fig. 1. The general ranking problem of Example 1.

to focus on ranking problems and scoring methods. For this purpose, the example below will be used.

**Example 1.** Let $(N, T^{(1)}, T^{(2)}) \in \mathcal{T}^4$ be a general ranking problem describing a tournament with two rounds.

It is shown in Fig. 1: A directed edge from node $X_i$ to $X_j$ indicates a win of player $X_i$ over $X_j$ (and a loss of $X_j$ against $X_i$), while an undirected edge from node $X_i$ to $X_j$ represents a drawn match between the two players. This representation will be used in further examples, too.

So, player $X_1$ has defeated $X_4$ in the first round (Fig. 1(a)), while players $X_2$ and $X_3$ have not played. In the second round, players $X_1$ and $X_2$, as well as players $X_3$ and $X_4$ have drawn (Fig. 1(b)). The whole tournament is shown in Fig. 1(c).

According to the following result, at least one property from the set of ANO, NEU and SC will be violated by any individual scoring method.

**Proposition 1.** *There exists no anonymous and neutral individual scoring method satisfying SC.*

**Proof.** Let $g : \mathcal{T}^n \to \mathbb{R}^n$ be an anonymous and neutral individual scoring method. Consider Example 1. ANO and NEU imply that $g_2(N, T^{(1)}) = g_3(N, T^{(1)})$ and $g_2(N, T^{(2)}) = g_3(N, T^{(2)})$, therefore

$$\begin{aligned} g_2(N, \mathbf{T}) &= \delta(g_2(N, T^{(1)}), g_2(N, T^{(2)})) \\ &= \delta(g_3(N, T^{(1)}), g_3(N, T^{(2)})) = g_3(N, \mathbf{T}). \end{aligned} \quad (1)$$

Note that $O_1^{(1)} = \{X_4\}$, $O_1^{(1)} = \{X_2\}$ and $O_4^{(1)} = \{X_1\}$, $O_4^{(2)} = \{X_3\}$. Take the one-to-one correspondences $h_{14}^{(1)} : O_1^{(1)} \leftrightarrow O_4^{(1)}$ such that $h_{14}^{(1)}(X_4) = X_1$ and $h_{14}^{(2)} : O_1^{(2)} \leftrightarrow O_4^{(2)}$ such that $h_{14}^{(2)}(X_2) = X_3$. Now $t_{12}^{(2)} = t_{43}^{(2)}$ since the corresponding matches resulted in draws. Furthermore, $t_{14}^{(1)} \neq t_{41}^{(1)}$ since the value of a win and a loss should be different. It can be assumed without loss of generality that $t_{14}^{(1)} > t_{41}^{(1)}$. Suppose that $g_1(N, \mathbf{T}) \leq g_4(N, \mathbf{T})$. Then players $X_1$ and $X_4$ have a draw against a player with the same strength ($X_2$ and $X_3$, respectively), but $X_1$ has defeated $X_4$, so it has a better result against a not weaker opponent. Therefore, SC (Axiom 3) implies $g_1(N, \mathbf{T}) > g_4(N, \mathbf{T})$, which is a contradiction, thus $g_1(N, \mathbf{T}) > g_4(N, \mathbf{T})$ holds.

However, $O_2^{(1)} = \emptyset$, $O_2^{(2)} = \{X_1\}$ and $O_3^{(1)} = \emptyset$, $O_3^{(2)} = \{X_4\}$. Consider the unique one-to-one correspondence $h_{14}^{(2)} : O_2^{(2)} \leftrightarrow O_3^{(2)}$, which — together with $t_{21}^{(2)} = t_{34}^{(2)}$ (the two draws should be represented by the same number) and $g_1(N, \mathbf{T}) > g_4(N, \mathbf{T})$ — leads to $g_2(N, \mathbf{T}) > g_3(N, \mathbf{T})$ because player $X_2$ has achieved the same result against a stronger opponent than player $X_3$. In other words, SC requires the draw of $X_2$ to be more valuable than the draw of $X_3$, but it cannot be reflected by any individual scoring method $g$ according to (1). □

## 4. The Case of Ranking Problems and Scoring Methods

According to Proposition 1, only the procedure underlying scoring methods can be compatible with self-consistency. Therefore, this section will focus on scoring methods.

### 4.1. *Axioms of invariance with respect to the results matrix*

Let $O \in \mathbb{R}^{n \times n}$ be the matrix with all of its entries being zero.

**Axiom 4 (Symmetry (SYM) [González-Díaz *et al.* (2014)]).** Let $(N, R, M) \in \mathcal{R}^n$ be a ranking problem such that $R = O$. Scoring method $f : \mathcal{R}^n \to \mathbb{R}^n$ is *symmetric* if $f_i(N, R, M) = f_j(N, R, M)$ for all $X_i, X_j \in N$.

According to SYM, if all paired comparisons (but not necessarily all matches in each round) between the players result in a draw, then all players will have the same score.

**Axiom 5 (Inversion (INV) [Chebotarev and Shamis, 1998]).** Let $(N, R, M) \in \mathcal{R}^n$ be a ranking problem. Scoring method $f : \mathcal{R}^n \to \mathbb{R}^n$ is *invertible* if $f_i(N, R, M) \geq f_j(N, R, M) \Leftrightarrow f_i(N, -R, M) \leq f_j(N, -R, M)$ for all $X_i, X_j \in N$.

INV means that taking the opposite of all results changes the ranking accordingly. It establishes a uniform treatment of victories and losses.

**Corollary 1.** *Let $f : \mathcal{R}^n \to \mathbb{R}^n$ be a scoring method satisfying INV. Then for all $X_i, X_j \in N$: $f_i(N, R, M) > f_j(N, R, M) \Leftrightarrow f_i(N, -R, M) < f_j(N, -R, M)$.*

The following result has been already mentioned by González-Díaz *et al.* [2014, p. 150].

**Corollary 2.** INV *implies* SYM.

It seems to be difficult to argue against SYM. However, scoring methods based on right eigenvectors [Wei, 1952; Slutzki and Volij, 2005, 2006; Kitti, 2016] violate INV.

### 4.2. *Properties of independence*

The next axiom deals with the effects of certain changes in the aggregated tournament matrix $A$.

**Axiom 6 (Independence of irrelevant matches (IIM) [González-Díaz *et al.* (2014)).** Let $(N, A), (N, A') \in \mathcal{R}^n$ be two ranking problems and $X_i, X_j, X_k, X_\ell \in N$ be four different players such that $(N, A)$ and $(N, A')$ are identical but $a_{k\ell} \neq a'_{k\ell}$. Scoring method $f : \mathcal{R}^n \to \mathbb{R}^n$ is called *independent of irrelevant matches* if $f_i(N, A) \geq f_j(N, A) \Rightarrow f_i(N, A') \geq f_j(N, A')$.

IIM means that "remote" matches — not involving players $X_i$ and $X_j$ — do not affect the pairwise ranking of players $X_i$ and $X_j$.

IIM seems to be a powerful property. González-Díaz *et al.* [2014] state that "*when players have different opponents* (*or face opponents with different intensities*), IIM *is a property one would rather not have*". Csató [2019b] argues on an axiomatic basis against IIM.

The rounds of a given tournament can be grouped arbitrarily. Therefore, the following property makes much sense.

**Axiom 7 (OP, [González-Díaz *et al.*, 2014]).** Let $(N, A), (N, A') \in \mathcal{R}^n$ be two ranking problems where all players have played $m$ matches and $X_i, X_j \in N$ be two different players. Let $f : \mathcal{R}^n \to \mathbb{R}^n$ be a scoring method such that $f_i(N, A) \geq f_j(N, A)$ and $f_i(N, A') \geq f_j(N, A')$.[c] $f$ satisfies OP if $f_i(N, A+A') \geq f_j(N, A+A')$, furthermore, $f_i(N, A + A') > f_j(N, A + A')$ if $f_i(N, A) > f_j(N, A)$ or $f_i(N, A') > f_j(N, A')$.

OP is a relatively restricted version of combining ranking problems, which implies that if player $X_i$ is not worse than player $X_j$ on the basis of some rounds as well as on the basis of another set of rounds such that all players have played in each round (so they have played the same number of matches altogether), then this pairwise ranking should hold after the two distinct set of rounds are considered jointly.

One can consider a stronger version of order preservation, too.

**Axiom 8 (Strong order preservation (SOP), [van den Brink and Gilles, 2009]).** Let $(N, A), (N, A') \in \mathcal{R}^n$ be two ranking problems and $X_i, X_j \in N$ be two players. Let $f : \mathcal{R}^n \to \mathbb{R}^n$ be a scoring method such that $f_i(N, A) \geq f_j(N, A)$ and $f_i(N, A') \geq f_j(N, A')$. $f$ satisfies SOP if $f_i(N, A+A') \geq f_j(N, A+A')$, furthermore, $f_i(N, A + A') > f_j(N, A + A')$ if $f_i(N, A) > f_j(N, A)$ or $f_i(N, A') > f_j(N, A')$.

In contrast to OP, SOP does not contain any restriction on the number of matches of the players in the ranking problems to be aggregated.

[c] González-Díaz *et al.* [2014] formally introduce a stronger version of this axiom since only $X_i$ and $X_j$ should have the same number of matches in the two ranking problems. However, in the counterexample of González-Díaz *et al.* [2014], which shows the violation of OP by several ranking methods, all players have played the same number of matches.

**Corollary 3.** SOP *implies* OP.

It will turn out that the weaker property, order preservation has still unfavourable implications.

### 4.3. *Relations among the axioms*

In this part, some links among SYM, INV, IIM, and (strong) OP will be revealed.

**Remark 1.** SYM and OP (SOP) imply INV.

**Proof.** Consider a ranking problem $(N, R, M) \in \mathcal{R}^n$ where $f_i(N, R, M) \geq f_j(N, R, M)$ for players $X_i, X_j \in N$. If $f_i(N, -R, M) > f_j(N, -R, M)$, then $f_i(N, O, 2M) > f_j(N, O, 2M)$ due to OP, which contradicts to SYM. So $f_i(N, -R, M) \leq f_j(N, -R, M)$ holds. □

It turns out that IIM is also closely connected to SOP.

**Proposition 2.** *A scoring method satisfying* NEU, SYM *and* SOP *meets* IIM.

**Proof.** Assume to the contrary, and let $(N, R, M) \in \mathcal{R}^n$ be a ranking problem, $f : \mathcal{R}^n \to \mathbb{R}^n$ be a scoring method satisfying NEU, SYM, and SOP, and $X_i, X_j, X_k, X_\ell \in N$ be four different players such that $f_i(N, R, M) \geq f_j(N, R, M)$, and $(N, R', M') \in \mathcal{R}^n$ is identical to $(N, R, M)$ except for the result $r'_{k\ell}$ and number of matches $m'_{k\ell}$ between players $X_k$ and $X_\ell$, where $f_i(N, R', M') < f_j(N, R', M')$.

According to Remark 1, $f$ satisfies INV, hence $f_i(N, -R, M) \leq f_j(N, -R, M)$. Denote by $\sigma : N \to N$ the permutation $\sigma(X_i) = X_j$, $\sigma(X_j) = X_i$, and $\sigma(X_k) = X_k$ for all $X_k \in N \setminus \{X_i, X_j\}$. NEU leads to $f_i[\sigma(N, R, M)] \leq f_j[\sigma(N, R, M)]$, and $f_i[\sigma(N, -R', M')] < f_j[\sigma(N, -R', M')]$ due to INV and Corollary 1. With the notations $R'' = \sigma(R) - \sigma(R') - R + R' = O$ and $M'' = \sigma(M) + \sigma(M') + M + M'$, we get

$$(N, R'', M'') = \sigma(N, R, M) + \sigma(N, -R', M') + (N, -R, M) + (N, R', M').$$

SYM implies $f_i(N, R'', M'') = f_j(N, R'', M'')$ since $R'' = O$, but $f_i(N, R'', M'') < f_j(N, R'', M'')$ from SOP, which is a contradiction. □

It remains to be seen whether NEU, SYM, and SOP are all necessary for Proposition 2.

**Lemma 1.** NEU, SYM, *and* SOP *are logically independent axioms with respect to the implication of* IIM.

**Proof.** It is shown that there exist scoring methods, which satisfy exactly two properties from the set NEU, SYM, and SOP, but violate the third and does not meet IIM, too:

(1) SYM and SOP: The sum of the results of the "previous" player, $f_i(N, R, M) = \sum_{j=1}^{n} r_{i-1,j}$ for all $X_i \in N \setminus \{X_1\}$ and $f_1(N, R, M) = \sum_{j=1}^{n} r_{n,j}$;

(2) NEU and SOP: Maximal number of matches of other players, $f_i(N,R,M) = \max\{\sum_{k=1}^{n} m_{jk} : X_j \neq X_i\}$[d];
(3) NEU and SYM: Aggregated sum of the results of opponents, $f_i(N,R,M) = \sum_{X_j \in O_i} \sum_{k=1}^{n} r_{jk}$. □

Proposition 2 helps in deriving another impossibility statement.

**Proposition 3.** *There exists no scoring method that satisfies* NEU, SYM, SOP *and* SC.

**Proof.** According to Proposition 2, NEU, SYM and SOP imply IIM. Csató [2019b, Theorem 3.1] has shown that IIM and SC cannot be met at the same time. □

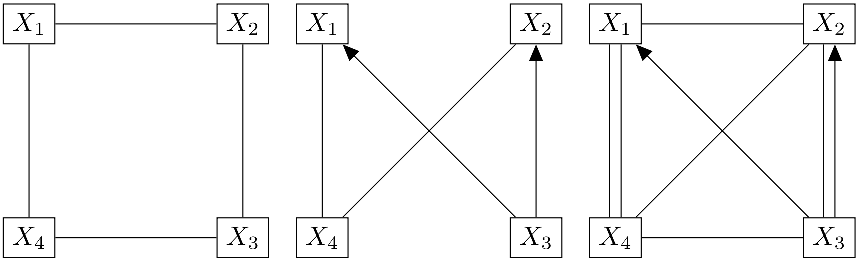
### 4.4. *A basic impossibility result*

The four axioms of Proposition 3 are not independent despite Lemma 1. However, a much stronger statement can be obtained by eliminating NEU and SYM, which also allows for a weakening of SOP by using OP. Note that substituting an axiom with a weaker one in an impossibility statement leads to a stronger result.

We will use a generalized tournament with four players for this purpose.

**Example 2.** Let $(N,R,M),(N,R',M') \in \mathcal{R}^4$ be two ranking problems. They are shown in Fig. 2: in the first tournament described by $(N,R,M)$, matches between players $X_1$ and $X_2$, $X_1$ and $X_4$, $X_2$ and $X_3$, $X_3$ and $X_4$ all resulted in draws (see Fig. 2(a)). On the other side, in the second tournament, described by $(N,R',M')$, players $X_1$ and $X_2$ have lost against $X_3$ and drawn against $X_4$ (see Fig. 2(b)). The two ranking problems can be summed in $(N,R'',M'') \in \mathcal{R}^4$ such that $R'' = R + R'$ and $M'' = M + M'$ (see Fig. 2(c)).

**Theorem 1.** *There exists no scoring method that satisfies* OP *and* SC.



(a) $(N,R,M)$ (b) $(N,R',M')$ (c) $(N,R+R',M+M')$

Fig. 2. The ranking problems of Example 2.

[d]It is worth to note that the maximal number of own matches satisfies NEU, SOP and IIM.

**Proof.** Assume to the contrary that there exists a self-consistent scoring method $f : \mathcal{R}^n \to \mathbb{R}^n$ satisfying OP. Consider Example 2.

(1) Take the ranking problem $(N, R, M)$. Note that $O_1 = O_3 = \{X_2, X_4\}$ and $O_2 = O_4 = \{X_1, X_3\}$.

  (a) Consider the identity one-to-one correspondences $h_{13} : O_1 \leftrightarrow O_3$ and $h_{31} : O_3 \leftrightarrow O_1$ such that $h_{13}(X_2) = h_{31}(X_2) = X_2$ and $h_{13}(X_4) = h_{31}(X_4) = X_4$. Since $r_{12} = r_{32} = 0$ and $r_{14} = r_{34} = 0$, players $X_1$ and $X_3$ have the same results against the same opponents, hence $f_1(N, R, M) = f_3(N, R, M)$ from SC.

  (b) Consider the identity one-to-one correspondences $h_{24} : O_2 \leftrightarrow O_4$ and $h_{42} : O_4 \leftrightarrow O_2$. Since $r_{21} = r_{41} = 0$ and $r_{23} = r_{43} = 0$, players $X_2$ and $X_4$ have the same results against the same opponents, hence $f_2(N, R, M) = f_4(N, R, M)$ from SC.

  (c) Suppose that $f_2(N, R, M) > f_1(N, R, M)$, which implies $f_4(N, R, M) > f_3(N, R, M)$. Consider the one-to-one mapping $h_{12} : O_1 \leftrightarrow O_2$, where $h_{12}(X_2) = X_1$ and $h_{12}(X_4) = X_3$. Since $r_{12} = r_{21} = 0$ and $r_{14} = r_{23} = 0$, player $X_1$ has the same results against stronger opponents compared to $X_2$, hence $f_1(N, R, M) > f_2(N, R, M)$ from SC, which is a contradiction.

  (d) An analogous argument shows that $f_1(N, R, M) > f_2(N, R, M)$ cannot hold.

  Therefore, SC leads to $f_1(N, R, M) = f_2(N, R, M) = f_3(N, R, M) = f_4(N, R, M)$ in the first ranking problem.

(2) Take the ranking problem $(N, R', M')$. Note that $O'_1 = O'_2 = \{X_3, X_4\}$ and $O'_3 = O'_4 = \{X_1, X_2\}$.

  (a) Consider the identity one-to-one correspondences $h'_{12} : O'_1 \leftrightarrow O'_2$ and $h'_{21} : O'_2 \leftrightarrow O'_1$. Since $r'_{13} = r'_{23} = -1$ and $r'_{14} = r'_{24} = 0$, players $X_1$ and $X_2$ have the same results against the same opponents, hence $f_1(N, R', M') = f_2(N, R', M')$ from SC.

  (b) Consider the identity one-to-one correspondence $h'_{34} : O'_3 \leftrightarrow O'_4$. Since $1 = r'_{31} > r'_{41} = 0$ and $1 = r'_{32} > r'_{42} = 0$, player $X_3$ has better results against the same opponents compared to $X_4$, hence $f_3(N, R', M) > f_4(N, R', M)$ from SC.

  So SC leads to $f_1(N, R', M') = f_2(N, R', M')$ and $f_3(N, R', M') > f_4(N, R', M')$ in the second ranking problem.

(3) Take the sum of these two ranking problems, the ranking problem $(N, R'', M'')$.

  Suppose that $f_1(N, R'', M'') \geq f_2(N, R'', M'')$. Consider the one-to-one mappings $g_{21} : O_2 \leftrightarrow O_1$ and $g'_{21} : O'_2 \leftrightarrow O'_1$ such that $g_{21}(X_1) = X_2$, $g_{21}(X_3) = X_4$ and $g'_{21}(X_3) = X_3$, $g'_{21}(X_4) = X_4$. Since $r_{21} = r_{12} = 0$, $r_{23} = r_{14} = 0$ and $r'_{23} = r'_{13} = -1$, $r'_{24} = r'_{14} = 0$, player $X_2$ has the same results against stronger opponents compared to $X_1$, hence $f_2(N, R'', M'') > f_1(N, R'', M'')$ from SC, which leads to a contradiction.

To summarize, SC results in $f_1(N, R'', M') < f_2(N, R'', M'')$, however, OP implies $f_1(N, R'', M'') = f_2(N, R'', M'')$ as all players have played two matches in $(N, R', M')$ and $(N, R', M')$, respectively, which is impossible.

Therefore, it has been derived that no scoring method can meet OP and SC simultaneously on the universal domain of $\mathcal{R}^n$. □

Theorem 1 is a serious negative result: by accepting SC, the ranking method cannot be required to preserve two players' pairwise ranking when some ranking problems, where all players have played the same number of matches, are aggregated.

**Example 3.** Let $(N, R, M) \in \mathcal{R}^4$ be the ranking problem in Fig. 3: $X_1$ has drawn against $X_2$, $X_2$ against $X_3$ and $X_3$ against $X_4$.

Theorem 1 would be more straightforward as a strengthening of Proposition 3 if SC implies NEU and/or SYM. However, it is not the case as the following result holds.

**Remark 2.** There exists a scoring method that is self-consistent, but not neutral and symmetric.

**Proof.** The statement can be verified by an example where an SC-compatible scoring method violates NEU and SYM.

Consider Example 3 with a scoring method $f$ such that $f_1(N, R, M) > f_2(N, R, M) > f_3(N, R, M) > f_4(N, R, M)$, for example, player $X_i$ gets the score $4 - i$. $f$ meets SC since $X_1$ has the same result against a stronger opponent compared to $X_4$, while there exists no correspondence between opponent sets $O_2$ and $O_3$ satisfying the conditions of SC.

Let $\sigma : N \to N$ be a permutation such that $\sigma(X_1) = X_4$, $\sigma(X_2) = X_3$, $\sigma(X_3) = X_2$, and $\sigma(X_4) = X_1$. Since $\sigma(N, R, M) = (N, R, M)$, NEU implies $f_4(N, R, M) > f_1(N, R, M)$ and $f_3(N, R, M) > f_2(N, R, M)$, a contradiction. Furthermore, SYM leads to $f_1(N, R, M) = f_2(N, R, M) = f_3(N, R, M) = f_4(N, R, M)$, another impossibility. Therefore, there exists a self-consistent scoring method, which is not neutral and symmetric. □

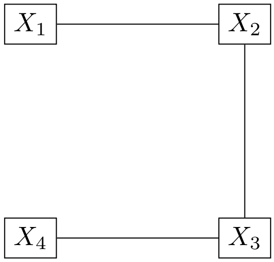



Fig. 3. The ranking problem of Example 3.

## 5. Conclusions

We have found some unexpected implications of different properties in the case of generalized tournaments where the players should be ranked on the basis of their match results against each other. First, SC prohibits the use of individual scoring methods, that is, scores cannot be derived before the aggregation of tournament rounds (Proposition 1). Second, independence of irrelevant matches (posing a kind of independence concerning the pairwise ranking of two players) follows from three axioms, NEU (independence of relabeling the players), SYM (implying a flat ranking if all aggregated comparisons are draws), and SOP (perhaps the most natural property concerning the aggregation of ranking problems). According to Csató [2019b], there exists no scoring method satisfying SC and IIM, hence Proposition 2 implies that NEU, SYM, SOP and SC cannot be met simultaneously (Proposition 3). It even turns out that SC and a weaker version of SOP are still enough to derive this negative result (Theorem 1), consequently, one should choose between these two natural fairness requirements.

What do our results say to practitioners who want to rank players or teams? First, SC does not allow to rank them in individual rounds, one has to wait until all tournament results are known and can be aggregated. Second, SC is not compatible with OP on this universal domain. It is not an unexpected and counter-intuitive result as, according to González-Díaz *et al.* [2014], a number of ranking methods violate OP. We have proved that there is no hope to find a reasonable scoring method with this property. From a more abstract point of view, breaking of OP in tournament ranking is a version of Simpson's paradox, a phenomenon in probability and statistics, in which a trend appears in different groups of data but disappears or reverses when these groups are combined.[e] This negative result holds despite SC is somewhat weaker than our intuition suggests: it does not imply NEU and SYM, so even a self-consistent ranking of players may depend on their names and without ties if all matches are drawn (Remark 2). Third, losing the simplicity provided by OP certainly does not facilitate the axiomatic construction of scoring methods.

Consequently, while sacrificing SC or OP seems to be unavoidable in our general setting, an obvious continuation of the current research is to get positive possibility results by some domain restrictions or further weakening of the axioms. It is also worth to note that the incompatibility of the two axioms does not imply that any scoring method is always going to work badly, but all can lead to problematic results at times.

## Acknowledgments

We are grateful to *Sándor Bozóki* for useful advice. Anonymous reviewers provided valuable comments and suggestions on earlier drafts. The research was supported

[e]We are grateful to an anonymous referee for this remark.

by OTKA grant K 111797 and by the MTA Premium Post Doctorate Research Program.

## References

Altman, A. and Tennenholtz, M. [2008] Axiomatic foundations for ranking systems, *J. Artif. Intell. Res.* **31**(1), 473–495.

Arrow, K. J. [1950] A difficulty in the concept of social welfare, *J. Politi. Econ.* **58**(4), 328–346.

Baker, R. D. and McHale, I. G. [2014] A dynamic paired comparisons model: Who is the greatest tennis player? *Europ. J. Operat. Res.* **236**(2), 677–684.

Baker, R. D. and McHale, I. G. [2017] An empirical Bayes model for time-varying paired comparisons ratings: Who is the greatest women's tennis player? *Europ. J. Operat. Res.* **258**(1), 328–333.

Berker, Y. [2014] Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations, *Europ. Sport Manag. Quart.* **14**(2), 194–210.

Bouyssou, D. [1992] Ranking methods based on valued preference relations: A characterization of the net flow method, *Europ. J. Operat. Res.* **60**(1), 61–67.

Bouyssou, D. and Perny, P. [1992] Ranking methods for valued preference relations: A characterization of a method based on leaving and entering flows, *Europ. J. Operat. Res.* **61**(1-2), 186–194.

Bozóki, S., Csató, L. and Temesi, J. [2016] An application of incomplete pairwise comparison matrices for ranking top tennis players, *Europ. J. Operat. Res.* **248**(1), 211–218.

Chebotarev, P. and Shamis, E. [1997] Constructing an objective function for aggregating incomplete preferences, in *Constructing Scalar-Valued Objective Functions*, Lecture Notes in Economics and Mathematical Systems, Vol. 453, eds. Tangian, A. & Gruber, J., (Springer, Berlin), pp. 100–124.

Chebotarev, P. and Shamis, E. [1998] Characterizations of scoring methods for preference aggregation, *Ann. Operat. Res.* **80**, 299–332.

Chebotarev, P. and Shamis, E. [1999] Preference fusion when the number of alternatives exceeds two: Indirect scoring procedures, *J. Franklin Inst.* **336**(2), 205–226.

Csató, L. [2017] 2018 FIFA World Cup qualification can be manipulated, http://unipub.lib.uni-corvinus.hu/3053/.

Csató, L. [2018] Was Zidane honest or well-informed? How UEFA barely avoided a serious scandal, *Econ. Bull.* **38**(1), 152–158.

Csató, L. [2019a] A fairer penalty shootout design in soccer, arXiv:1806.01114.

Csató, L. [2019b] An impossibility theorem for paired comparisons, *Central European Journal of Operational Research*, in press, doi:10.1007/s10100-018-0572-5.

Csató, L. [2019c] Incentive compatible designs for tournament qualifiers with round-robin groups and repechage, arXiv:1804.04422.

Csató, L. [2019d] Overcoming the incentive incompatibility of tournaments with multiple group stages, arXiv:1712.04183.

Csató, L. [2019e] UEFA Champions League entry has not satisfied strategy-proofness in three seasons, *Journal of Sports Economics*, in press, doi:10.1177/1527002519833091.

Csató, L. [2019f] When UEFA rules had inspired unfair behavior on the field, arXiv:1806.03978.

Dagaev, D. and Sonin, K. [2018] Winning by losing: Incentive incompatibility in multiple qualifiers, *J. Sports Econ.* **19**(8) 1122–1146.

Gibbard, A. [1973] Manipulation of voting schemes: A general result, *Econometrica* **41**(4) 587–601.

González-Díaz, J., Hendrickx, R. and Lohmann, E. [2014] Paired comparisons analysis: an axiomatic approach to ranking methods, *Social Choice Welf.* **42**(1), 139–169.

Kitti, M. [2016] Axioms for centrality scoring with principal eigenvectors, *Social Choice Welf.* **46**(3), 639–653.

Landau, E. [1895] Zur relativen Wertbemessung der Turnierresultate, *Deutsches Wochenschach* **11**, 366–369.

Langville, A. N. and Meyer, C. D. [2012] *Who's #1? : The Science of Rating and Ranking* (Princeton University Press, Princeton).

Pauly, M. [2014] Can strategizing in round-robin subtournaments be avoided? *Social Choice Welf.* **43**(1), 29–46.

Radicchi, F. [2011] Who is the best player ever? A complex network analysis of the history of professional tennis, *PloS one* **6**(2), e17249.

Rubinstein, A. [1980] Ranking the participants in a tournament, *SIAM J. Appl. Math.* **38**(1), 108–111.

Satterthwaite, M. A. [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.

Slutzki, G. and Volij, O. [2005] Ranking participants in generalized tournaments, *Int. J. Game Theory* **33**(2), 255–270.

Slutzki, G. and Volij, O. [2006] Scoring of web pages and tournaments–axiomatizations, *Social Choice Welf.* **26**(1), 75–92.

van den Brink, R. and Gilles, R. P. [2003] Ranking by outdegree for directed graphs, *Discrete Math.* **271**(1–3), 261–270.

van den Brink, R. and Gilles, R. P. [2009] The outflow ranking method for weighted directed graphs, *Europ. J. Operat. Res.* **193**(2), 484–491.

Vaziri, B., Dabadghao, S., Yih, Y. and Morin, T. L. [2018] Properties of sports ranking methods, *J. Operat. Res. Soc.* **69**(5), 776–787.

Vong, A. I. K. [2017] Strategic manipulation in tournament games, *Games Econ. Behav.* **102**, 562–567.

Wei, T. H. [1952] The algebraic foundations of ranking theory, PhD thesis, University of Cambridge.

Zermelo, E. [1929] Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung, *Math. Zeit.* **29**, 436–460.