European Sport Management Quarterly

ISSN: 1618-4742 (Print) 1746-031X (Online) Journal homepage: www.tandfonline.com/journals/resm20

Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations

Yannick Berker

To cite this article: Yannick Berker (2014) Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations, European Sport Management Quarterly, 14:2, 194-210, DOI: 10.1080/16184742.2014.884152

To link to this article: https://doi.org/10.1080/16184742.2014.884152

European Sport Management Quarterly, 2014
Vol. 14, No. 2, 194–210, http://dx.doi.org/10.1080/16184742.2014.884152

Routledge
Taylor & Francis Group

Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations

Yannick Berker*

Department of Physics of Molecular Imaging Systems, Institute for Experimental Molecular Imaging, RWTH Aachen University, Pauwelsstraße 30, Aachen 52074, Germany

(*Received 15 February 2013; accepted 20 September 2013*)

Research question: Many sports tournaments, or tournament stages, are held in round-robin format. In establishing standings, tie-breaker criteria are often required to assign unique ranks to competitors that are equal on points. During the 2012 Union of European Football Associations (UEFA) European Football Championship, the relative ranking of two teams could possibly be influenced by a match in which neither team was involved, which we refer to as heteronomous relative ranking (HRR). We seek to shed light on the origin and the practical relevance of HRR in soccer competitions.
Research methods: We trace the appearance of HRR back to particularities in the UEFA EURO 2012 tie-breaker criteria, which favor head-to-head records over goal difference, and relate the concept of autonomous relative ranking to Arrow's Independence of Irrelevant Alternatives. Using historical and Monte–Carlo simulated data, we compare HRR occurrence rates under EURO 2012 regulations to those under Fédération Internationale de Football Association (FIFA) 2010 regulations, the latter of which give less importance to head-to-head records.
Results and findings: HRR is well explained by tie-breaker criteria, namely the priorities of head-to-head records and goal difference. HRR occurs in more than 10% of four-team soccer groups under EURO 2012 regulations; this rate is further increased in round-robin groups of six teams. FIFA 2010 regulations lead to less than 0.1% HRR.
Implications: Head-to-head records may result in counterintuitive side effects that should be avoided when designing ranking systems.

Keywords: ranking; football; soccer; Independence of Irrelevant Alternatives; UEFA; FIFA

Introduction

Round-robin tournaments (or tournament stages), in which all participants face each other in turn, are popular in many sports. If tournament schedules allow, the necessary number of matches can be scheduled in order for complete rankings to be generated. In elimination systems, by contrast, participants eliminated before the semi-final stage can be assigned only a group of ranks such as 5–8 or 9–16, which depend on their last tournament stage.[1] Additionally, the achieved rank may depend heavily on the initial seeding.

The basic mechanism that renders round-robin systems preferable is the establishment of a *beating relation* (Rubinstein, 1980) between all pairs of participants. As a result, tournament planning is less complex since the matches can be scheduled independently

*Email: berker@pmi.rwth-aachen.de

of match results. Vice versa, scheduling does not necessarily have a significant effect on match results, as was recently shown for soccer (Goossens & Spieksma, 2012).

While round-robin systems have found application in individual sports, for example, the World Chess Championships, the majority of round-robin tournaments are held for team sports, such as soccer, football, basketball, or cricket. With the growing economic impact of many of these sports, the need for fair, consistent and, ideally, intuitive rules for participant ranking increases. League licensing and broadcast incomes may be distributed between participants as a function of their rank upon season completion. Thus, fair ranking affects the predictability of financial income for participants, especially professional teams. It also leads to higher potential supporter interest.

As an introductory example illustrating the focus of this paper, imagine a bet formulated as 'A finishes ahead of B' in a round-robin tournament. Intuitively, the outcome of this bet should be determined as soon as A and B have completed all their matches. However, situations may develop in which the outcome can be influenced – or even ultimately determined – by the result of matches in which neither A nor B is involved.

In this paper, we summarize axiomatic requirements that round-robin tournament ranking systems should fulfill. We focus on the 2012 Union of European Football Associations (UEFA) European Football Championship final tournament (UEFA EURO 2012) and describe the occurrence of artifactual cases where the relative rankings of two teams were influenced by matches in which neither was involved. To estimate the practical relevance of these cases, their rate of occurrence is estimated by applying EURO 2012 regulations to historical data and Monte–Carlo simulated data. We find that the origin of these artifactual cases lies within the EURO 2012 tie-breaker criteria, which we then compare with corresponding Fédération Internationale de Football Association (FIFA) regulations and comment on their practical relevance.

Participant ranking in literature

The issue of tournament rankings has attracted the interest of many authors, as can be concluded from the bibliography on college football ranking systems collected by Wilson (n.d.). Incentives for establishing rankings include both the prediction of results from intermediate rankings, as discussed by Fair and Oster (2007) and the purpose of fairly awarding tournament prizes or prestige. For examples of the latter, we refer to the works of Macmillan and Smith (2007) or Fainmesser, Fershtman, and Gandal (2009).

Ranking is straightforward if the beating relation can be formulated mathematically in a strict total order. This formulation, however, disregards the possibility of draws and implies transitivity. That is, if A beats B and B beats C, a necessary consequence is that A beats C. In this situation, the winner of each match can be theoretically predicted from a set of each participant's properties, and ranking the participants is reduced to sorting them by these properties (whether these may be observable or not). Intuitively, sports tournaments that are most attractive to spectators do not follow a strict total order and fair ranking is a more complex issue. For round-robin tournaments, Rubinstein (1980) has treated this basic problem by formulating the following axioms[2]:

(1) Anonymity

'[T]he ranking method does not discriminate against players because of their label' (Rubinstein, 1980). The mathematical formulation of this axiom can only be understood

ex post: based on a (partially) completed tournament and the corresponding standings, the ranking system must exchange the ranks of two given participants if their match results are exchanged.

(2) Positive responsiveness

If a participant turns a loss into a win, his rank must not decrease. In contrast to anonymity, this axiom also has *ex-ante* implications: before and during each match, a participant seeking to maximize his rank should never be motivated to lose. The *ex-post* interpretation implies that, at no time during a tournament, a result worse than the achieved one in any of the completed matches might have resulted in a better rank.[3]

(3) Autonomous relative ranking (ARR)[4]

'[T]he relative ranking of two players is independent of those matches in which neither is involved' (Rubinstein, 1980). Here, the *relative* ranking of two participants is the order in which they are ranked, irrespective of their actual ranks, or those of other participants. As above, this axiom has both *ex-ante* and *ex-post* implications. Before and during any match between A and B, no possible result should be able to influence the relative ranking of any other C and D (*ex ante*); and at no time during a tournament, changing a past result of any match between A and B may result in a different relative ranking of C and D (*ex post*). Violations of this axiom (be they *ex ante* or *ex post*) are what we term as heteronomous relative ranking (HRR).

Rubinstein proved that in a round-robin tournament, counting the number of opponents a participant has beaten is the only ranking system that fulfills these axioms. However, he did not consider draws as match results. His characterization was thus expanded by Henriet (1985), who showed that a point system considering draws fulfills three axioms very similar to the aforementioned ones. Moreover, it was also shown that fulfilling these axioms is not enough to characterize a unique ordering. In fact, in modern systems, the number of points awarded for wins and draws also has an influence on the ranking. Additionally, in a tournament in which all matches are tied, participants cannot be uniquely ranked. This reasoning does not touch upon the usefulness of the aforementioned axioms and they can still be seen as a minimum set of properties of an ideal ranking system.

Round-robin ranking in practice

In round-robin systems today, participants are typically ranked in close relation to the works of Rubinstein (1980) and Henriet (1985), that is, by the number of wins over other participants.[5] If draws are possible, wins, draws and losses are usually converted into points in a point system. Typical encoding schemes include two (or three) points for a win and one point for a draw; losses are not usually awarded any points.

Difficulties primarily arise in the case of point equality. Especially in early tournament stages such as qualifiers, a complete ranking (i.e., with no shared ranks) has to be established to determine and seed participants that are eligible for following tournament stages. Different criteria have been used in the past to resolve ties. Among these are the sum of scored in-match achievements (such as goals in the case of soccer), the in-match-achievement difference (scored minus conceded goals), positions in long-term rankings, fair play conduct during the tournament or, as a final criterion, the drawing

of lots (which reduces to flipping a coin in the case of two participants). Another criterion, common in leagues such as those organized by the National Basketball Association or the National Football League, are the head-to-head records of the tied participants. This criterion gained increased importance a few years ago in soccer competitions organized by the UEFA.

EURO 2012 tie-breaking regulations

The EURO 2012 group stage was held in groups of four teams each. The teams in each group faced each other in turn in two simultaneous matches on each of three match days. The groups' winners and runners-up proceeded to the elimination stage. In preparation for the final tournament, the importance of head-to-head records for group stage ranking was once again underlined by late changes to the regulations (UEFA, 2012a, amendment to p. 9). The amendments clarified that head-to-head records could be evaluated recursively for groups and subgroups of tied participants. Using head-to-head records for tie-breaking is controversial among supporters, and the EURO 2012 implementation of it provoked a high HRR occurrence rate as will be seen in the following. The final version of paragraph 8.07 of the official EURO 2012 regulations reads as follows (UEFA, 2012a, p. 9, amendment to p. 9)[6]:

> If two or more teams are equal on points on completion of the group matches, the following criteria are applied to determine the rankings:
>
> a) Higher number of points obtained in the matches played between the teams in question;
>
> b) Superior goal difference resulting from the matches played between the teams in question (if more than two teams finish equal on points);
>
> c) Higher number of goals scored in the matches played between the teams in question (if more than two teams finish equal on points);
>
> d) If, after having applied criteria a) to c), two teams still have an equal ranking, criteria a) to c) are reapplied exclusively to the matches between the two teams in question to determine the final rankings of the two teams. If this procedure does not lead to a decision, criteria e) to i) apply in the order given;
>
> e) superior goal difference in all group matches;
>
> f) higher number of goals scored in all group matches;
>
> g) position in the UEFA national team coefficient ranking system […];
>
> h) fair play conduct of the teams (final tournament);
>
> i) drawing of lots.

Rules a to c represent the application of head-to-head records, which have higher priority than goal differences (in e). After commenting on the importance of ARR, we will focus on how these tie-breaking criteria violate this axiom.[7]

On the importance of ARR

The importance of ARR can be explained with two primary arguments. First, we note the similarity between ARR and Independence of Irrelevant Alternatives (IIA) as introduced by Arrow (1963). Arrow's IIA, and alternative formulations such as ranked IIA (Altman & Tennenholtz, 2008), have been extensively studied in many settings, and strong rationale has been found for the former in social choice theory (Blin, 1976).

The main difference between IIA and HRR is that the definition of IIA requires the relative ranking of any set of (in particular, two) choices to be independent of any choice outside this set (in particular, any *third* choice). In that sense, IIA is stronger than ARR, which only requires the relative ranking of two participants to be independent of any match result between arbitrary *third and fourth* participants. This difference can be understood from the field of application for IIA ranking, which is the aggregation of consistent preferences (referred to as *orderings*) of a set of alternatives given by multiple voters, which may be inconsistent with each other. An example is competitive dancing, in which all participants are ordered by each of several judges and the IIA property of ranking systems has been studied (Mora, 2002; Munoz, 2003). Here, instead, ARR is examined in a setting of a single 'judge' (the given match results) with possibly non-consistent preferences (e.g., intransitive or cyclic results). In this setting, IIA cannot be fulfilled, since the relative ranking of two teams would reduce to the single match result, and cycles in the results would prevent the establishment of an ordering from these relative rankings. ARR, on the other hand, can be fulfilled by the points system. To summarize, we find that ARR is a relaxed variant of the desirable IIA property, tailored for the points systems considered by Rubinstein (1980) and Henriet (1985).

Second, the aforementioned ranking systems are defined (even uniquely, in Rubinstein's case) by fulfilling anonymity, positive responsiveness, and ARR, but can yield shared ranks (ties). The addition of tie-breaking criteria as in the EURO 2012 regulations maintains positive responsiveness, but violates anonymity and ARR. Focusing on the latter, it seems reasonable to minimize the rate of HRR to obtain ranking system properties as close as possible to the fundamental properties of the points system before addition of any tie-breaking criteria.

EURO 2012 HRR *ex-ante* examples

To motivate our approach and clarify the impact of HRR we present two examples from the EURO 2012. Since we assume that intermediate rankings (after one or two out of three match days) are of limited interest, we restrict our examples to those with relevant third match day events. These examples are thus *ex ante* by definition.

Group A: actual situation

Table 1 (left) summarizes intermediate (as of half time of the third match day) and final match results in Group A (UEFA, 2013b). Here, we are especially interested in the influence of the match between the Czech Republic and Poland on the relative ranking of Greece and Russia. Table 2 lists intermediate standings, and it is seen that the first three teams are level on points and are, thus, being ranked according to head-to-head records.[8] During half time it could be concluded that, current results standing, Russia (as group winner) and Greece (as runner-up) would proceed to the elimination stage. However, it

Table 1. EURO 2012 intermediate match results during third match day.

Day	Team 1	Group A	Team 2	Team 1	Group B	Team 2
1	Poland	1–1	Greece	Netherlands	0–1	Denmark
	Russia	4–1	Czech Republic	Germany	1–0	Portugal
2	Greece	1–2	Czech Republic	Denmark	2–3	Portugal
	Poland	1–1	Russia	Netherlands	1–2	Germany
3	Greece	(1) 1–0 (0)	Russia	Portugal	(0) 0–1 (1)	Netherlands
	Czech Republic	**(1) 0–0 (0)**	**Poland**	**Denmark**	**(0) 0–0 (1)**	**Germany**

Note: Bold lines represent matches with different intermediate and final results.
Group A: at half-time (actual final results in parentheses); Group B: at 15th minute (hypothetical final results in parentheses).

Table 2. Intermediate standings of EURO 2012 Group A.

#	Team	W	D	L	+	–	Δ	P	W*	D*	L*	+*	–*	Δ*	P*
1	**Russia**	**1**	**1**	**1**	**5**	**3**	**+2**	**4**	**1**	**0**	**1**	**4**	**2**	**+2**	**3**
2	**Greece**	**1**	**1**	**1**	**3**	**3**	**±0**	**4**	**1**	**0**	**1**	**2**	**2**	**±0**	**3**
3	Czech Republic	1	1	1	3	5	−2	4	1	0	1	3	5	−2	3
4	Poland	0	3	0	2	2	±0	3							

Note: #, Rank; W, wins; D, draws; L, losses; +, scored; –, conceded; Δ, difference; P, points; asterisks denote head-to-head-record values between teams level on points. Bold lines represent teams subject to HRR.

Table 3. Actual final standings of EURO 2012 Group A.

#	Team	W	D	L	+	–	Δ	P	W*	D*	L*	+*	–*	Δ*	P*
1	Czech Republic	2	0	1	4	5	−1	6							
2	**Greece**	**1**	**1**	**1**	**3**	**3**	**±0**	**4**	**1**	**0**	**0**	**1**	**0**	**+1**	**3**
3	**Russia**	**1**	**1**	**1**	**5**	**3**	**+2**	**4**	**0**	**0**	**1**	**0**	**1**	**−1**	**0**
4	Poland	0	2	1	2	3	−1	2							

Note: #, Rank; W, wins; D, draws; L, losses; +, scored; –, conceded; Δ, difference; P, points; asterisks denote head-to-head-record values between teams level on points. Bold lines represent teams subject to HRR.

was also evident that by scoring a single goal against Poland, the Czech Republic could replace Russia as group winner, while Greece would remain runner-up.

This indeed happened, yielding the final standings in Table 3. By comparison with Table 2, we observe HRR of Greece and Russia: not only has the Czech Republic surpassed both Greece and Russia, but also the relative ranking of Greece and Russia is reversed, without any of their match results having changed.

Group B: hypothetical situation

While HRR did not manifest in the final standings of Group B, the group was in an intermediate state prone to HRR by the scoring of a single goal on the third match day. Table 1 (right) also summarizes intermediate and hypothetical match results of Group B (UEFA, 2013c), and Table 4 lists Group B's intermediate standings. Had Germany scored against Denmark, which is the hypothetical case treated in Table 5, the relative ranking of the Netherlands and Portugal would have been reversed, directly violating ARR. Note

Table 4. Intermediate standings of EURO 2012 Group B.

#	Team	W	D	L	+	–	Δ	P	W*	D*	L*	+*	–*	Δ*	P*
1	Germany	2	1	0	3	1	+2	7							
2	Denmark	1	1	1	3	3	±0	4							
3	**The Netherlands**	**1**	**0**	**2**	**2**	**3**	**−1**	**3**	**1**	**0**	**0**	**1**	**0**	**+1**	**3**
4	**Portugal**	**1**	**0**	**2**	**3**	**4**	**−1**	**3**	**0**	**0**	**1**	**0**	**1**	**−1**	**0**

Note: #, Rank; W, wins; D, draws; L, losses; +, scored; –, conceded; Δ, difference; P, points; asterisks denote head-to-head-record values between teams level on points. Bold lines represent teams subject to HRR.

Table 5. Hypothetical final standings of EURO 2012 Group B (2nd-level head-to-head records between Portugal and Denmark omitted for clarity).

#	Team	W	D	L	+	–	Δ	P	W*	D*	L*	+*	–*	Δ*	P*
1	Germany	3	0	0	4	1	+3	9							
2	**Portugal**	**1**	**0**	**2**	**3**	**4**	**−1**	**3**	**1**	**0**	**1**	**3**	**3**	**±0**	**3**
3	Denmark	1	0	2	3	4	−1	3	1	0	1	3	3	±0	3
4	**The Netherlands**	**1**	**0**	**2**	**2**	**3**	**−1**	**3**	**1**	**0**	**1**	**1**	**1**	**±0**	**3**

Note: #, Rank; W, wins; D, draws; L, losses; +, scored; –, conceded; Δ, difference; P, points; asterisks denote head-to-head-record values between teams level on points. Bold lines represent teams subject to HRR.

that this special case requires repeated calculations of head-to-head records at different levels, according to paragraph 8.07 d.

Common properties and hypothesis

Comparison of the described situations yields some insight: intermediate Group A standings involve a three-team head-to-head record, while the final standings involve a two-team head-to-head record. Here, the reason for HRR is that the Czech Republic left the head-to-head group by scoring against Poland. Of course the inverse is also possible, for example, by Poland equalizing again and forcing the Czech Republic back into the head-to-head group. In fact, Group B's hypothetical standings show this inverse effect, demonstrated by Denmark being forced into the head-to-head group of Portugal and the Netherlands as a consequence of Germany scoring.

It seems reasonable to hypothesize that under EURO 2012 regulations, HRR only occurs when transitioning from two-team head-to-head groups to more-than-two-team head-to-head groups (or vice versa). This seems plausible, as can be clarified in a two-step argument. First, any head-to-head group including at least two teams requires these teams to be level on points, and since we consider HRR, these teams' points cannot change. Second, in order to influence these teams' relative ranking without contribution of either of the two, another team has to join (or leave) their head-to-head group to cause HRR.

Quantitative evaluations

To test our aforementioned hypothesis and to determine HRR occurrence rates, we implemented the EURO 2012 ranking system (sections 'a' to 'f') in a computer program (MATLAB, The MathWorks, Natick, MA) to perform systematic evaluations. To compare the EURO 2012 ranking system with a system that puts less focus on head-to-head records, we also implemented the ranking system according to 2010 FIFA World

Cup regulations, article 39, section 5 (FIFA, 2007, p. 41). Both implementations were thoroughly validated not only, but above all, by comparing the standings calculated from official final match results of the respective tournament group stages (EURO 2012 and 2010 FIFA World Cup) with the official standings. Data-sets will now be described, and results and their analyses will follow.

Historical data

For the empirical data-set we used match results from past UEFA European Football Championships and FIFA World Cups. Matches were selected from all group stages that were held in a four-team round-robin format with six matches per group. This includes 28 groups that played at UEFA EURO events between 1980 and 2012 (2 groups per tournament until 1992, and 4 thereafter; UEFA, 2013a) as well as 84 groups at FIFA World Cups held between 1958 and 2010 (4 groups per tournament until 1970, 4 + 2 groups in 1974 and in 1978, 6 groups per tournament from 1982 to 1994, and 8 thereafter; FIFA, 2013).

Monte–Carlo simulations

A second data-set was artificially created to complement the preliminary findings from historical data, which are hardly convincing by themselves due the comparably low number of groups. As a template, we chose a group with four teams and six matches, thus providing 12 free parameters (numbers of goals per team and match). In each of the 100,000 × 6 simulation runs, we created a random group state by drawing 12 random integers from one of six discrete probability distributions and attributing these numbers as goals to match results.[9] The sampled probability distributions are discrete uniform, Poisson or negative binomial distributions. Uniform distributions have been chosen since they allow for simple validation of some results, for example, the fraction of drawn matches, which should equal 1 over the number of possible numbers of goals. We further selected Poisson and negative binomial distributions, which have been shown to model the number of football goals rather well (Greenhough, Birch, Chapman, & Rowlands, 2002). These were parameterized using two sets of parameters each, all inspired by results reported by Bittner, Nußbaumer, Janke, and Weigel (2009). The complete list of distributions is as follows:

(1) U_3: discrete uniform probability distribution defined on $\{0, \ldots, 3\}$,
(2) U_5: discrete uniform probability distribution defined on $\{0, \ldots, 5\}$,
(3) P_{Const}: Poisson distribution with mean value $\lambda = 1.5$ for all teams,
(4) P_{Var}: Poisson distribution with separate λ_i per team, where the values $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.75, 1.25, 1.75, 2.25)$ are randomly permuted and assigned to the teams in every new group state,
(5) NBD_{Const}: negative binomial distribution with number of successes $r = 10$ and probability of failure in a single trial $p = 0.15$, and
(6) NBD_{Var}: negative binomial distribution with $p = 0.15$ and a separate r_i per team, $(r_1, r_2, r_3, r_4) = (8.5, 9.5, 10.5, 11.5)$.

Thus, using six probabilistic models, we created 6 × 100,000 = 600,000 reference group states in total. No effort was made to avoid duplicate or equivalent group states.

HRR occurrence rates

From the above *ex-ante* examples, it is clear that HRR can only be detected by comparing two group states. Therefore, for each of the 112 historical and 600,000 simulated group states, the standings were established following sections 'a' to 'f' of EURO 2012 regulations and used as a reference.[10] Then, separately for each of the 12 free parameters, we tested whether increasing the number of goals represented by this parameter by one would reveal HRR in the modified standings. Thus, in total, we had (112 + 600,000) reference states and 12 × (112 + 600,000) = 7,201,344 modified group states at our disposal.

After each free parameter had been modified, the contestants of the affected match, A and B, were determined. Then the relative ranking in the modified standings of the other two 'unaffected' participants, C and D, was compared to their relative ranking in the reference standings. If they were different, that is, if C was ahead of D in the reference standings, but the inverse was true in the modified standings, the reference group state was counted as an '*ex-post* HRR group state.' In the special case where the modified parameter belonged to a match on the third match day, the group state was also counted as an '*ex-ante* HRR group state.' A group state was not counted twice in the same category, and was excluded if the relative ranking could not be determined; for example, in cases where the ranking system is incomplete or in more degenerate cases such as six goalless draws.

Using the exact same input, we repeated the same procedure following FIFA regulations, which essentially rank teams by a) points, b) goal difference, and c) scored goals in all matches, followed by d) points, e) goal difference, and f) scored goals in matches between the teams in question if the rules in sections 'a' to 'c' do not lead to a decision.[11] We compare UEFA and FIFA tie-breaking criteria in Table 6 and note that the

Table 6. Comparison of tie-breaking criteria in EURO 2012 and FIFA 2010 regulations.

EURO 2012	Recursive	FIFA 2010	Recursive
Points		Points (a)	
Head-to-head points (a)	(d)	–	
Head-to-head goal difference (b)		–	
Head-to-head goals scored (c)		–	
Goal difference (e)		Goal difference (b)	
Goals scored (f)		Goals scored (c)	
–		Head-to-head points (d)	(Note 11)
–		Head-to-head goal difference (e)	
–		Head-to-head goals scored (f)	
UEFA national team coefficient (g)		–	
Fair play conduct (h)		–	
Lots (i)		Lots (g)	

Note: In parenthesis, references to the subsections of the respective rules. Some groups of criteria are applied recursively to groups of participants equally ranked using the former criteria.

main difference is the order of head-to-head records and global goal difference and goals scored. It can therefore be expected that, if HRR only occurs as a result from the application of head-to-head records, occurrence rates are lower under FIFA regulations. This can be explained by participants required to be equal not only on points, but also on global goal difference and goals scored, the probability of which is lower than for being equal on points alone.

Finally, in order to evaluate the influence of the number of participants per group, we repeated all simulation experiments with groups of six teams and appropriately modified UEFA- and FIFA-like ranking systems (according to Table 6). While it can be shown that the maximum required level of head-to-head-record evaluations is two in groups of four teams,[12] this does not hold in the case of six teams. Nevertheless, to allow for unbiased comparison, we maintained the structure of the program and ignored the potential incompleteness of group states requiring three (or more) levels of head-to-head records. In our 6 × 100,000 × (1 + 30) = 18.6 million reference and modified group states, this occurred in 39 group states related to 21 distinct reference group states under EURO 2012 regulations.

Results

Table 7 summarizes the HRR occurrence rates for the data described above. Relevant *ex-post* occurrence rates were found following EURO 2012 regulations (about 8% in historical and 11% in simulated data). Due to the low number of historical groups, the Clopper–Pearson confidence interval (CI, for $\alpha = 5\%$; Clopper & Pearson, 1934) covers 11% (3.7%, 14.7%), while all CI for simulated data cover less than 0.5%. We therefore conclude that all simulations are compatible with the findings from historical data.

To further strengthen this argument, we determined the rate of drawn games for reference states in each data-set, which is shown in Table 7. From the comparison of U_3 and U_5, we note that a higher draw rate correlates with a higher HRR rate, which indicates that the rate of drawn games is another useful measure to validate simulations. With the exception of U_5, all drawn game rates are within 6 percentage points from the one for historical data. Overall, the influence of the different goal probability distributions is marginal, as also confirmed by the comparison of standard deviations of the mean values with the mean values.

As a first conclusion, in a tournament with 10 groups of 4 teams, HRR can be expected in the final standings of about one group. *Ex-ante* occurrence rates are significantly higher than one-third of *ex-post* ones, signaling that an HRR group state (even if counted only once) can be affected *ex post* by events on different match days

Table 7. Rates of HRR occurrence, rates of drawn games, and rates of differences between UEFA and FIFA standings (in %) for historical data (H) and different goal probability distributions.

Analysis		H	U_3	U_5	P_{Const}	P_{Var}	NBD_{Const}	NBD_{Var}	Mean ± SD
HRR *ex post*	UEFA	8.04	11.3	10.8	11.8	9.17	11.9	11.6	10.7 ± 1.4
	FIFA	0	0.046	0.019	0.045	0.025	0.031	0.027	0.028 ± 0.015
HRR *ex ante*	UEFA	4.46	4.06	3.90	4.35	3.35	4.29	4.23	4.09 ± 0.35
	FIFA	0	0.016	0.006	0.018	0.006	0.010	0.004	0.009 ± 0.006
Draws		26.8	25.0	16.7	24.3	22.2	21.3	21.1	22.5 ± 3.1
UEFA ≠ FIFA		9.82	17.1	21.6	17.8	13.6	19.4	18.9	16.9 ± 3.7

Table 8. *ex-post* HRR occurrence rates (in %, including incomplete rankings) by class of head-to-head-record transition.

Transition		H	U_3	U_5	P_{Const}	P_{Var}	NBD_{Const}	NBD_{Var}	Mean ± SD
2-to-3	UEFA	6.25	10.1	10.2	10.5	8.43	10.8	10.6	9.53 ± 1.52
	FIFA	0	0.109	0.023	0.122	0.070	0.069	0.061	0.065 ± 0.040
2-to-4	UEFA	1.79	1.49	0.661	1.50	0.853	1.17	1.10	1.22 ± 0.37
	FIFA	0	0.015	0	0.014	0.004	0.009	0.010	0.007 ± 0.005
4-to-4	UEFA	0	0.008	0	0.008	0.003	0.007	0.006	0.005 ± 0.003
	FIFA	0	0	0	0	0	0	0	0 ± 0

Note: Without incomplete rankings, rates are slightly lower: UEFA 2-to-3, ≤0.3 percentage points (pp) lower; FIFA 2-to-3, zero (historical data) or about one-third of values reported (simulated data); UEFA 2-to-4, up to 0.03 pp lower; FIFA 2-to-4, zero; UEFA 4-to-4, zero.

under EURO 2012 regulations. The FIFA system, by contrast, leads to less than 0.1% HRR.

Table 7 also provides rates of differences between UEFA standings and FIFA standings in reference states. Comparing the difference rates with HRR occurrence rates, we see that the FIFA system shows small but significant differences to the UEFA system, apart from preventing HRR.[13] Since the ranking differences are due to the different priority given to head-to-head records, this priority difference may also be the main reason for the difference in HRR rates: intuitively, head-to-head records are much less likely to apply if the teams in question need not only have the same number of points but also the same goal difference.

Contrary to our hypothesis, not all HRR occurred at transitions between three-or-four-team head-to-head groups and two-team head-to-head groups (or vice versa), as shown in Table 8. In addition to the former, the special case of four-team head-to-head groups with transitions to different four-team head-to-head groups occurred at very low rates, when a complete ranking cannot be determined. These situations arise as a merge (or a separation) of a two-team head-to-head group with (or from) two other teams, respectively. When only regarding complete rankings, these situations do not arise with our data.

It should be noted that we tested for HRR by adding a single goal to a reference group state to obtain a modified group state. This test can be formulated as 'Could HRR have occurred with respect to final standings if another single goal had been scored?' Only slightly different results are to be expected by removing a single goal ('…if a single goal had *not* been scored'): in technical terms, both options involve only directly neighboring group states. Even larger occurrence rates are to be expected when testing for all possible match outcomes, both for *ex-post* and *ex-ante* evaluations.

Finally, HRR occurrence rates under UEFA-like regulations for groups with six teams are about a factor of 2.4 (*ex-post* results) or 2.2 (*ex-ante* results) higher than in four-team groups. At the same time, FIFA-like regulations only yield HRR occurrence rates about 1.5 times *lower* than in four-team groups (Table 9). One mechanism increasing HRR occurrence rates can be seen by increases in both the number of matches to modify (to obtain modified group states) and the number of teams *not* involved in a specific match (and thus potentially subject to HRR). With regard to the former, the number of matches modified is increased from 6 to 15 in *ex-post* results and from 2 to 3 in *ex-ante* results; the latter number doubles from 2 to 4 in both cases. On the other hand, the probability of ties leading to the application of head-to-head records is decreased as the number of matches (and thus the spread of points [UEFA] or points, goal differences, and goals [FIFA])

Table 9. Rates as in Table 7 for simulated data of six-team groups.

Analysis		U_3	U_5	P_{Const}	P_{Var}	NBD_{Const}	NBD_{Var}	Mean ± SD
HRR *ex post*	UEFA	27.0	26.0	27.6	20.0	27.2	25.5	25.6 ± 2.6
	FIFA	0.034	0.012	0.018	0.01	0.015	0.011	0.017 ± 0.008
HRR *ex ante*	UEFA	10.0	8.35	10.4	6.93	9.77	8.91	9.07 ± 1.18
	FIFA	0.014	0.006	0.007	0.004	0.004	0.002	0.006 ± 0.004

Table 10. Match results yielding incomplete ranking under EURO 2012 regulations.

Day 1			Day 2			Day 3		
A	2–1	B	A	1–1	C	A	0–1	D
C	1–0	D	B	0–0	D	B	2–1	C

Table 11. Standings with incomplete ranking (head-to-head records in the right panel do not apply when interpreting EURO 2012 regulations strictly).

#	Team	W	D	L	+	–	Δ	P	W*	D*	L*	+*	–*	Δ*	P*
1	**A**	**1**	**1**	**1**	**3**	**3**	**±0**	**4**	1	1	0	3	2	+1	4
1	**B**	**1**	**1**	**1**	**3**	**3**	**±0**	**4**	1	0	1	3	3	±0	3
1	**C**	**1**	**1**	**1**	**3**	**3**	**±0**	**4**	0	1	1	2	3	−1	1
4	D	1	1	1	1	1	±0	4							

Note: #, Rank; W, wins; D, draws; L, losses; +, scored; –, conceded; Δ, difference; P, points; asterisks denote head-to-head-record values between teams level on points. Bold lines highlight teams with equal ranks.

Table 12. Match results yielding incomplete ranking under strictly interpreted FIFA 2010 regulations.

	Day 1			Day 2			Day 3	
A	2–1	B	A	0–1	C	A	0–0	D
C	1–1	D	B	0–0	D	B	1–0	C

increases. This effect is stronger under FIFA-like regulations since the definition of a tie is so much stricter.

Conclusion

We presented how EURO 2012 regulations violate the axioms of anonymity, scheduling invariance and autonomous relative ranking: in the case of heteronomous relative ranking, the order of two participants in a ranking depends on the results of a match in which neither participant is involved. We found that the occurrence rate of heteronomous relative ranking is relevant in both historical and simulated data, and showed that

Table 13. Standings with incomplete ranking under strictly interpreted FIFA 2010 regulations if head-to-head records in the right-most panel do not apply.

#	T	W	D	L	+	−	Δ	P	W*	D*	L*	+*	−*	Δ*	P*	W**	D**	L**	+C	−**	Δ**	P**
1	**A**	**1**	**1**	**1**	**2**	**2**	**±0**	**4**	**1**	**0**	**1**	**2**	**2**	**±0**	**3**	1	0	0	2	1	+1	3
1	**B**	**1**	**1**	**1**	**2**	**2**	**±0**	**4**	**1**	**0**	**1**	**2**	**2**	**±0**	**3**	0	0	1	1	2	−1	0
3	**C**	**1**	**1**	**1**	**2**	**2**	**±0**	**4**	1	0	1	1	1	±0	3							
4	D	0	3	0	1	1	±0	3														

Note: Bold lines highlight teams subject to head-to-head record ranking.
**Denotes 2nd level head-to-head record values between teams with equal 1st level head-to-head record values.

heteronomous relative ranking occurs primarily when transitioning from more-than-two-team to two-team head-to-head records or vice versa.

Anonymity violations are primarily academic in nature, and as discussed, violations of scheduling invariance may be well justified after all. In terms of heteronomous relative ranking, using head-to-head records only in case of exactly *two* teams equal on points seems like an easy way of preventing head-to-head-record transitions; however, as can be easily verified, this does not resolve heteronomous relative ranking in most cases. Using the FIFA ranking system would be an improvement over the current system, but it is still not ideal. The possibility of heteronomous relative ranking should thus be considered when designing future ranking systems.

Notes

1. In alternative systems, such as double elimination tournaments participants join a *losers' bracket* after being eliminated from the default *winners' bracket*, thus, allowing more detailed ranks to be assigned, but requiring a higher number of matches.
2. We note that a fourth axiom could also prove useful in a round-robin tournament, which we tentatively term as scheduling invariance: the ranking should be invariant with respect to the order in which individual match results are determined.
3. In some competitions, participants may have incentives to lose a match if there is an elimination stage that follows, for example, in order to prevent having to face a specific opponent too early in the following stage. A recent example is the badminton women's doubles tournament at the 2012 Olympic Games, during which several participants have been disqualified as a consequence of attempting to lose their last group stage match. As this specific example shows, such behavior is considered detrimental to the sport and regulations are usually designed to prevent it, underlining the axiom's validity.
4. While the names of the first two axioms have been coined by Rubinstein (1980), we chose the third axiom's name.
5. This is in spite of some results showing that the goal difference may be a better measure of team fitness, for example, in soccer (Heuer & Rubner, 2009).
6. Section 'd' only appears in the amendments; see also UEFA (2012b). It is the only significant change regarding tie-breaking, compared to the preceding tournament's regulations (UEFA, 2005, p. 9).
7. For completeness, we note further peculiarities of the EURO 2012 regulations, which are less relevant than HRR and therefore outside the scope of this paper.

- *Anonymity violations*: Paragraph 8.07 g violates anonymity. It can easily be seen that all teams are ranked according to their UEFA national team coefficient in case of six goalless draws. Exchanging match results between any pair of participants does not influence the ranking, contrary to what is postulated by the anonymity axiom. (By contrast, the fair play conduct of the teams fulfills anonymity, as it is explicitly constrained to the final tournament and can be regarded part of the match results. Similarly, the drawing of lots is compatible with anonymity, as the average result does not depend on the teams' labels and the random variable behaves identical for both parties.)
- *Scheduling variance*: The rules laid down in paragraph 8.08 of EURO 2012 regulation read (UEFA, 2012a):

> If two teams which have the same number of points, the same number of goals scored and conceded play their last group match against each other and are still equal at the end of that match, the ranking of the two teams in question is determined by kicks from the penalty mark […], provided no other teams within the group have the same number of points on completion of all group matches. Should more than two teams have the same number of points, the criteria listed under paragraph 8.07 apply.

These rules may be understood as an attempt to obtain a decision on the playing field wherever possible, and this should be positively recognized. However, a consequence is that the order of matches (and not only their results) has an impact on the ranking: this can be seen by considering that two teams equal on points should decide their relative ranking by a penalty shoot-out in case they draw on their last group match, but by other rules if the order of match days is changed and the draw is a result of a different match day.

- *Incompleteness*: Table 10 shows possible match results with the EURO 2012 ranking displayed in Table 11. By parsing paragraph 8.07, we observe that a complete ranking cannot be established: rules 'a' to 'c' apply to teams A–D but do not break the tie between teams A–C. Consequently, rule 'd' does not apply, since it references 'two teams,' which is not defined in this case. Consequently, since rules 'e' and 'f' do not lead to a decision, teams A–C are ranked according to rule 'g' and their UEFA national team coefficient. This is despite availability of a unique ranking that can be established using three-team head-to-head records after four-team head-to-head records, as can be seen from the right part of Table 11. Since the intention of rule 'd' is clear by the rankings in the examples to follow, this problem is easily fixable by changing the restrictive formulation. This has been considered in more recent UEFA regulations, such as those for the 2013 UEFA European Women's Championship ('two or more teams'; UEFA, 2012c, paragraph 8.05, second sentence) or for the 2013/2014 UEFA Champions League ('teams'; UEFA, 2013d, paragraph 7.06 e). Our implementation uses this corrected formulation: head-to-head records are reapplied if two teams are equally ranked within a group of three or four teams equal on points.

8. In this situation, the same ranking would result if head-to-head records had not been applied, and the teams had simply been ranked by points and goal difference.
9. For EURO and FIFA final tournaments, held in one tournament region, all participants except the hosts can be considered 'away team,' so no discrimination between home and away teams was made.
10. For historical data, current regulations are essentially different from the ones used when the respective tournaments were held; for example, wins were awards two points only in tournaments before 1996. Therefore, standings for historical match results are not necessarily identical to original standings.
11. We note that these rules give room for interpretation since, unlike UEFA rules, they do not explicitly state whether rules 'd' to 'f' shall be applied repeatedly. As this is necessary to completely rank participants in situations such as the one exemplified by Tables 12 and 13, we implemented the FIFA ranking with up to twofold evaluation of head-to-head rankings.
12. Repeated evaluation of head-to-head records is only required if the number of tied teams in a head-to-head group is lower than the size of that group and greater than 1. Hence, 4-to-3 following by 3-to-2 transitions are the only way that might require a third evaluation of head-to-head records between the remaining two teams. We show that under the points system awarding three points for a win and one point for a tie in a group of four teams, these two teams necessarily need to draw their match.Each match awards 3 + 0, 1 + 1, or 0 + 3 points to the participants, that is, either two or three. The final total number of points in the group (after six matches) is thus bounded by 12 and 18. For four teams to be equal on (integer) points, 16 and 12 points are the only remaining possibilities. With 16 points, which equals 4 per team, each team must have won, drawn, and lost one game, respectively. Hence, when removing any one participant from a group of four teams equal on points to establish a group of three teams, these three teams cannot be equal on points; 4 points per team are thus impossible. Hence, the group needs to count 12 points in total, which is the minimum possible; as a consequence, each match only generates the minimum number of points (two) by ending in a draw.
13. This comprises, for example, the case of the 1974 FIFA World Cup Group 2, where Scotland failed to proceed to the elimination stage under (both historical and current) FIFA regulations, but would have done so under EURO 2012 regulations. However, despite the differences, the UEFA ranking is consistent in that it does not manifest HRR in this example.

References

Altman, A., & Tennenholtz, M. (2008). Axiomatic foundations for ranking systems. *Journal of Artificial Intelligence Research*, *31*, 473–495. doi:10.1613/jair.2306

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New York, NY: John Wiley & Sons.

Bittner, E., Nußbaumer, A., Janke, W., & Weigel, M. (2009). Football fever: Goal distributions and non-Gaussian statistics. *The European Physical Journal B*, *67*, 459–471. doi:10.1140/epjb/e2008-00396-1

Blin, J.-M. (1976). How relevant are 'irrelevant' alternatives? *Theory and Decision*, *7*, 95–105. doi:10.1007/BF00141104

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*, 404–413. doi:10.1093/biomet/26.4.404

Fainmesser, I., Fershtman, C., & Gandal, N. (2009). A consistent weighted ranking scheme with an application to NCAA college football rankings. *Journal of Sports Economics*, *10*, 582–600. doi:10.1177/1527002509336891

Fair, R. C., & Oster, J. F. (2007). College football rankings and market efficiency. *Journal of Sports Economics*, *8*, 3–18. doi:10.1177/1527002505276724

FIFA. (2007). *Regulations 2010 FIFA World Cup South Africa*™. Retrieved from www.fifa.com/mm/document/tournament/competition/fifa_wc_south_africa_2010_regulations_en_14123.pdf

FIFA. (2013). *Previous FIFA World Cups*™. FIFA.com. Retrieved from http://www.fifa.com/worldcup/archive/

Goossens, D. R., & Spieksma, F. C. R. (2012). The carryover effect does not influence football results. *Journal of Sports Economics*, *13*, 288–305. doi:10.1177/1527002511402932

Greenhough, J., Birch, P. C., Chapman, S. C., & Rowlands, G. (2002). Football goal distributions and extremal statistics. *Physica A: Statistical Mechanics and its Applications*, *316*, 615–624. doi:10.1016/S0378-4371(02)01030-0

Henriet, D. (1985). The Copeland choice function: An axiomatic characterization. *Social Choice and Welfare*, *2*, 49–63. doi:10.1007/BF00433767

Heuer, A., & Rubner, O. (2009). Fitness, chance, and myths: An objective view on soccer results. *The European Physical Journal B*, *67*, 445–458. doi:10.1140/epjb/e2009-00024-8

Macmillan, P., & Smith, I. (2007). Explaining international soccer rankings. *Journal of Sports Economics*, *8*, 202–213. doi:10.1177/1527002505279344

Mora, X. (2002). Puntuació en competicions amb diversos jutges [Scoring in competitions with several judges]. In E. Fossas & J. Solà-Morales (Eds.), *Grups d'estudi de Matemàtica i Tecnologia* [Study groups of mathematics and technology] (pp. 13–16). Barcelona: Facultat de Matemàtiques i Estadística, Universitat Politècnica de Catalunya. Retrieved from http://plus.upc.es/gemt/Docs/2002/gemtBCN_2002.pdf

Munoz, X. (2003). Fair ranking in dancesport competitions. In M. H. Albert (Ed.), *Proceedings, permutation patterns 2003* (Technical Report OUCS-2003-02) (pp. 39–48). Dunedin: Department of Computer Science, University of Otago. Retrieved from http://www.cs.otago.ac.nz/research/publications/oucs-2003-02.pdf

Rubinstein, A. (1980). Ranking the participants in a tournament. *SIAM Journal on Applied Mathematics*, *38*, 108–111. doi:10.1137/0138009

UEFA. (2005). *Regulations of the UEFA European Football Championship 2006/08*. Retrieved from http://www.uefa.com/newsfiles/19079.pdf

UEFA. (2012a). *Regulations of the UEFA European Football Championship 2010-12*. Retrieved from http://www.uefa.com/MultimediaFiles/Download/Regulations/competitions/Regulations/91/48/36/914836_DOWNLOAD.pdf

UEFA. (2012b). *Key EURO regulation changes approved – UEFA.com*. Retrieved from http://www.uefa.com/uefa/management/newsid=1800152.html

UEFA. (2012c). *Regulations of the UEFA European Women's Championship 2011-13*. Retrieved from http://www.uefa.com/MultimediaFiles/Download/Regulations/competitions/Regulations/01/78/71/58/1787158_DOWNLOAD.pdf

UEFA. (2013a). *UEFA EURO – History – UEFA.com*. Retrieved from http://www.uefa.com/uefaeuro/finals/history/index.html

UEFA. (2013b). *UEFA EURO – Standings – UEFA.com*. Retrieved from http://www.uefa.com/uefaeuro/season=2012/standings/round=15172/group=700964/index.html

UEFA. (2013c). *UEFA EURO – Standings – UEFA.com*. Retrieved from http://www.uefa.com/uefaeuro/season=2012/standings/round=15172/group=700965/index.html

UEFA. (2013d). *Regulations of the UEFA Champions League 2012-15 cycle 2013/14 season*. Retrieved from http://www.uefa.com/MultimediaFiles/Download/Regulations/competitions/Regulations/01/94/62/34/1946234_DOWNLOAD.pdf

Wilson, D. (n.d.). *Bibliography on college football ranking systems*. Retrieved from http://www.cae.wisc.edu/~dwilson/rsfc/rate/biblio.html