Mathematical Social Sciences 71 (2014) 20–29

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Social choice violations in rank sum scoring: A formalization of conditions and corrective probability computations

James Boudreau [a], Justin Ehrlich [b], Shane Sanders [c,*], Adam Winn [d]

[a] Department of Economics and Finance, College of Business Administration, University of Texas - Pan American, 1201 West University Drive Edinburg, TX 78539, United States

[b] Department of Computer Science, College of Business and Technology, Western Illinois University, 447P Stipes Hall, Macomb, IL 61455, United States

[c] Department of Economics and Decision Sciences, College of Business and Technology, Western Illinois University, 429 Stipes Hall, Macomb, IL 61455, United States

[d] Global Consulting Program, University of Illinois at Urbana-Champaign, United States

HIGHLIGHTS

- Formally defines and discusses applications of rank sum scoring (RSS).
- Considers the prevalence of rank sum scoring in ordinal group comparisons.
- Derives conditions under which RSS leads to violations of social choice principles.
- Uses computational methods to correct and extend prior probability calculations.
- Discusses the salience of social choice violations in RSS.

ARTICLE INFO

Article history:
Received 16 November 2012
Received in revised form
27 December 2013
Accepted 31 March 2014
Available online 12 April 2014

ABSTRACT

Rank sum scoring is a popular manner by which to obtain group rankings from events that are individual in nature. For example, rank sum scoring is employed in cross country team competition, the Putnam Mathematics Team Competition, snowboard cross team competition, snowboard half-pipe team competition, and snowboard slopestyle team competition. It also forms the basis for non-parametric statistical tests (the Wilcoxon rank-sum test and the related Mann–Whitney U test) that serve as an ordinal substitute for the two-sample t-test. Moreover, a close variant of rank sum scoring is used in team rowing competitions, alpine ski competitions, as well as in multiple event cross-fit contests. Given its applicability, we formally define rank sum scoring, derive conditions under which the methodology leads to violations of major social choice principles (ranking cycles and violations of "independence from irrelevant groups"), and calculate the probability that an "outcome sequence" leads to a given social choice violation. These probability calculations provide sizable corrections to prior calculations on the topic and alter our view as to the social choice properties of rank sum scoring. Contrary to prior results on the topic, we find that the two violations considered are not mutually inclusive under rank sum scoring.

1. Introduction

Rank sum scoring is a popular manner by which to obtain group rankings from events that are individual in nature. For example, rank sum scoring is employed in cross country running team competition, the famous Putnam Mathematics Team Competition, snowboard cross team competition, snowboard half-pipe team competition, and snowboard slopestyle team competition. It also forms the basis for two popular non-parametric statistical tests (the Wilcoxon rank-sum test and the related Mann–Whitney U test) that serve as an ordinal substitute for the two-sample t-test. Moreover, a close variant of rank sum scoring is used in team rowing competitions, alpine ski competitions, committee voting, as well as in multiple event cross-fit contests.[1] Given its

* Corresponding author. Tel.: +1 7853171116.
E-mail addresses: boudreaujw@utpa.edu (J. Boudreau), ja-ehrlich@wiu.edu (J. Ehrlich), sd-sanders@wiu.edu (S. Sanders), awinn4@illinois.edu (A. Winn).

http://dx.doi.org/10.1016/j.mathsocsci.2014.03.004
0165-4896/

[1] In this variant form, multiple event rankings are aggregated such that a given rank is not unique. We might call this variant "rank sum scoring with rank

broadening applicability and importance in certain allocation processes, we formally define rank sum scoring, derive conditions under which the methodology leads to violations of major social choice principles (ranking cycles and violations of "independence from irrelevant groups or IIG"), and calculate the probability that an "outcome sequence" leads to a given social choice violation.

These probability calculations provide sizable corrections and extensions to prior calculations on the topic and alter our view as to the "social choice fallibility" of rank sum scoring. Contrary to prior results on the topic, we find that the two violations considered are not mutually inclusive under rank sum scoring. There are several studies that consider the social choice properties of other aggregation rules in sport. Truchon (1998, 2004, 2008) and Gordon and Truchon (2008) show that social choice violations have plagued aggregate scoring in sports such as figure skating. Saari (2001) finds that team scoring in track and field can also violate the weak axiom of revealed preference. MacKay (1980) demonstrates problems with preference aggregation (event scoring aggregation) in the Olympic decathlon.

The previous literature on rank sum scoring, as applied to sport or other environments, is relatively sparse. Huntington (1938) shows by example that rank sum scoring can lead to violations of IIG. Huntington mentions cross country running and academic competition as applications of rank sum scoring. In a *Public Choice* piece on cross country running team competition, Hammond (2007) shows by example that rank sum scoring can violate two major social choice principles: transitivity and IIG.[2] He further examines the prevalence of these violations in rank sum team scoring and shows that violations of transitivity and independence are always possible in pairwise comparisons among three or more groups, each of which possesses three or more elements. Mixon and King (2012) find evidence of social choice violations in the 2008 and 2009 NCAA Cross Country Team Championships and conclude that "salaries, bonuses and budgets at the collegiate (cross country) level can be substantial, so that any inconsistencies and ambiguities in the scoring mechanism can be quite costly for individuals and institutions" (32). Wilcoxon (1945) and Wilcoxon et al. (1970) develop what is often referred to as the Wilcoxon Rank Sum Test. In these works, the authors suggest that individual observations arising from two balanced groups can be ranked individually. From these individual rankings, group quality can be compared according to the sum of ranks in each group. Mann and Whitney (1947) formulate a similar non-parametric statistical test. Essentially, these works advocate using pairwise group comparisons of rank sum as a statistical estimator. These previous works, though important, recognize only one application of rank sum scoring. Moreover, previous works fail to formally define rank sum scoring or to provide rigorous conditions under which social choice violations occur under rank sum scoring.

The remainder of the paper is laid out as follows. Section 2 develops the theory of rank sum scoring. Section 3 considers the likelihood that an outcome sequence generates a given social choice violation in rank sum scoring. Section 4 concludes by discussing the salience of social choice violations in rank sum scoring. This discussion is based upon a comparison of practitioner beliefs and calculated probabilities.

replacement". In terms of outcomes generated, this variant is identical to the Borda Count method of voting.

2 Recognition of these social choice violations derives largely from Arrow's (1951) seminal work on social choice theory.

2. The theory of rank sum scoring

Let us formally define rank sum scoring of groups. Consider three groups: X, Y, and Z. Each group is defined as a rank-ordered sequence of n individual elements, where n is some integer greater than 1.[3] For example, X might be defined as follows.

$$X = (x_1, x_2, x_3, \dots, x_n),$$

where the element x_i ($i \in \{1, 2, \dots, n\}$) represents the ith ranked element $\in X$. For simplicity, we assume that a team's rank-order sequence is invariant. Let us define an event as an objective process of comparison that generates a complete rank-order sequence of individuals across more than one group. In this sense, an event might be defined as a competition or as a statistical test. An event allows us to generate outcome sequences from team rank-order sequences by allowing for all possible between-team comparisons of individuals. Let us consider an event in which elements of X and Y are compared. If X and Y are each composed of n elements, for example, then the event generates a rank-ordered outcome sequence of $2n$ elements. For example, one possible outcome sequence, F_{XY} (for the case in which $n = 5$) is listed below.[4]

$$F_{XY} = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, x_5, y_4, y_5).$$

If x_1 precedes y_1 in the outcome sequence F_{XY}, we say $x_1 \succ y_1$ (x_1 ranks higher or better than y_1). For simplicity, we assume that rank-order equality between two elements is not possible, an outcome that would obtain given continuous (time) measurement in a race or continuous (weight, height, ...) measurement in a biological sample (to which the Wilcoxon rank sum test is applied).[5] For any x_i and y_j, we have that $x_i \succ y_j$ or $y_j \succ x_i$.

In rank sum scoring, each element of an outcome sequence is assigned an individual score equivalent to its position (its *rank*) in the outcome sequence F. In other words, each element is ranked pairwise against each other element in the event to form a set of $\frac{2n(2n-1)}{2}$ relations in an event featuring $2n$ elements (n elements per group). For each element, $(2n - 1)$ such relations are considered in generating the element's rank (score).

Formally, we denote the *rank* of an element $x_i \in X$ in the outcome sequence F_{XY} as $r(x_i|F_{XY})$. If $x_i^+(F_{XY}) = \{a \in F_{XY} | a \succ x_i\}$ is the set of elements that rank better than x_i in the outcome sequence F_{XY}, then $r(x_i|F_{XY}) = |x_i^+(F_{XY})| + 1$.

From elemental rankings, we generate a score for each team in an event as follows. The scores for teams X and Y based on the outcome sequence F_{XY} are

$$S(X|F_{XY}) = \sum_{x_j \in X} r(x_j|F_{XY})$$

and

$$S(Y|F_{XY}) = \sum_{y_j \in Y} r(y_j|F_{XY})$$

respectively. In outcome sequence $F'_{XY} = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, x_5, y_4, y_5)$, for example, it can be confirmed that $S(Y|F'_{XY}) = 37$ and $S(X|F'_{XY}) = 18$. We then map from group scores to group rankings

3 By the convention of cross country races, snowboard competitions, or the Putnam Math Team Competition, we will typically treat n as an odd integer greater than or equal to three when considering particular cases.

4 We use subscripts to indicate the specific teams included in an outcome sequence whenever relevant. When the specific teams (or the number of teams) involved in an outcome sequence is arbitrary, subscripts are omitted.

5 Many distance running races, equipped with finish line cameras, do not allow for ties.

as follows.

If $S(X|F_{XY}) < S(Y|F_{XY})$, then $X \succ Y \equiv$ If $S(X|F_{XY}) < S(Y|F_{XY})$, then X ranks higher than Y.
If $S(X|F_{XY}) = S(Y|F_{XY})$, then $X \sim Y \equiv$ If $S(X|F_{XY}) = S(Y|F_{XY})$, then X and Y rank equally.
If $S(X|F_{XY}) > S(Y|F_{XY})$, then $X \prec Y \equiv$ If $S(X|F_{XY}) < S(Y|F_{XY})$, then X ranks lower than Y.

In a three-group event, group scoring is analogously defined but competition is then between three groups. And, as Hammond (2007) notes, a given third group can influence the scores of two primary groups asymmetrically. Below is a possible outcome from a three-group event (where $n = 5$) featuring X, Y, and Z.

$$F_{XYZ} = (x_1, x_2, z_1, z_2, z_3, x_3, x_4, y_1, y_2, y_3, x_5, y_4, y_5, z_4, z_5).$$

In a three-group event, we similarly use elemental rankings to determine event scores for each team,

$$S(X|F_{XYZ}) = \sum_{x_j \in X} r(x_j|F_{XYZ})$$
$$S(Y|F_{XYZ}) = \sum_{y_j \in Y} r(y_j|F_{XYZ})$$
$$S(Z|F_{XYZ}) = \sum_{z_j \in Z} r(z_j|F_{XYZ})$$

and team scores are similarly mapped to team rankings (If $S(X|F_{XYZ}) < S(Z|F_{XYZ})$), then X ranks higher than Z. However, there are now three pairwise group rankings to consider (i.e., $\{X, Y\}$, $\{X, Z\}$, $\{Y, Z\}$). We now consider conditions under which violations of social choice principles might occur somewhere in the set of mappings from individual outcomes to team rankings.

Again, events are defined such that individual performances are invariant across event. Therefore, whether or not the third team, Z, is included in the outcome sequence for an event does not influence direct orderings between *elements of* X and *elements of* Y. Even under this assumption of independence of primitive element outcomes, however, we find that the inclusion of Z in an event may change the pairwise (team) ranking between X and Y. In the context of cross country running, Hammond (2007) calls such influence a violation of "independence from irrelevant teams". In the present study, we term this a violation of "independence from irrelevant groups" (IIG).

Definition 1 (*Violation of IIG under Rank Sum Scoring*). Given an outcome sequence F, a weak violation of IIG occurs between X and Y if[6]:

$S(X|F) < S(Y|F)$ but $S(X|F \setminus Z) = S(Y|F \setminus Z)$, or
$S(X|F) > S(Y|F)$ but $S(X|F \setminus Z) = S(Y|F \setminus Z)$, or
$S(X|F) = S(Y|F)$ but $S(X|F \setminus Z) > S(Y|F \setminus Z)$, or
$S(X|F) = S(Y|F)$ but $S(X|F \setminus Z) < S(Y|F \setminus Z)$.

A Strict Violation of IIG occurs between X and Y if:

$S(X|F) < S(Y|F)$ but $S(X|F) > S(Y|F \setminus Z)$, or if
$S(X|F) > S(Y|F)$ but $S(X|F) < S(Y|F \setminus Z)$.

Note that the rank of an individual element from either of two competing groups in an event can only be weakly increased by the addition of a third group. That is, $r(a|F) \geq r(a|F \setminus Z)$ for all $a \in X \cup Y$. Not all elements will have their rankings affected equally by the addition of a third group, however, and thus the scores of any two teams will not necessarily be altered by the same magnitude. The following statement formalizing this result is straightforward, but worth stating explicitly as it helps to structure our later analysis.

[6] Here we use the set difference operator to indicate an outcome sequence that is unchanged except for the exclusion of a specific group's elements.

Proposition 1. *For any outcome sequence F containing groups X, Y, and Z,*

$S(X|F) - S(X|F \setminus Z)$ *is not generally equal to* $S(Y|F) - S(Y|F \setminus Z)$.

This statement is verified by examining the empirical observations of Hammond (2007) or Mixon and King (2012) or by considering the following constructed outcome set:

$$F'' = (x_1, x_2, z_1, z_2, z_3, x_3, x_4, y_1, y_2, y_3, x_5, y_4, y_5, z_4, z_5).$$

In this case, $S(X|F'') - S(X|F'' \setminus Z) = 9$, since x_3, x_4, and x_5 each have their rankings increased by three due to z_1, z_2, and z_3. Meanwhile, $S(Y|F'') - S(Y|F'' \setminus Z) = 15$, as all five elements of group Y have their rankings increased by z_1, z_2, and z_3.

Proposition 2 is similarly straightforward, but also similarly useful in that it further clarifies how the scores of two groups can be impacted by the presence of a third. Specifically, it states that a higher ranking by X (as compared to Y) when Z is not included in an event's outcome sequence does not imply that the score of X will increase by less than the score of Y when Z is included.

Proposition 2. *If $S(X|F \setminus Z) < S(Y|F \setminus Z)$, it does not generally follow that*

$$(S(X|F) - S(X|F \setminus Z)) < (S(Y|F) - S(Y|F \setminus Z)).$$

This statement is verified by examining the empirical observations of Hammond (2007) or Mixon and King (2012) or by considering the following constructed outcome set:

$$F''' = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4, y_5, z_1, z_2, z_3, z_4, z_5, x_5).$$

In this case, $S(X|F''' \setminus Z) = 20$ and $S(Y|F''' \setminus Z) = 35$. However, $S(X|F''') - S(X|F''' \setminus Z) = 5$ while $S(Y|F''') - S(Y|F''' \setminus Z) = 0$. Practically, if one group (in this case, X) has an exceptionally weak element (here the lowest-ranked element across the three groups), this characteristic can become dramatically more costly as more groups are included in an event's outcome sequence.

Our next observation focuses on the relationship between how the scores of two groups change when a third team is added to an event's outcome sequence, and how those two teams compare pairwise with the added group. Note that

$$\begin{aligned} S(X|F) - S(X|F \setminus Z) &= \sum_{x_j \in X} r(x_j|F) - \sum_{x_j \in X} r(x_j|F \setminus Z) \\ &= \sum_{x_j \in X} (|x_j^+(F)| + 1) - \sum_{x_j \in X} (|x_j^+(F \setminus Z)| + 1) \\ &= \sum_{x_j \in X} |x_j^+(Z)|, \end{aligned}$$

where $|x_j^+(Z)|$ is the number of elements in Z that rank better than x_j. In other words, the increase in group X's score when group Z is added to an event is determined by the number of elements of Z that rank better than each element of X. Rearranging,

$$\begin{aligned} &(S(X|F) - S(X|F \setminus Z)) - (S(Y|F) - S(Y|F \setminus Z)) \\ &= \sum_{x_j \in X} |x_j^+(Z)| - \sum_{y_j \in Y} |y_j^+(Z)| \\ &= \sum_{x_j \in X} r(x_j|F_{XZ}) - \sum_{y_j \in Y} r(y_j|F_{YZ}) \\ &= S(X|F_{XZ}) - S(Y|F_{YZ}). \end{aligned}$$

That is, the difference between how the scores of any two teams change when a third team is added to an event depends upon how those two teams score pairwise against the third team.

Finally, and with that relationship in mind, we formalize the condition under which the IIG condition is violated under rank sum scoring.

Proposition 3. *When ordering two or three teams $\{X, Y, Z\}$ under rank sum scoring, the ordering between X and Y depends upon the presence (absence) of Z iff:*

$$[S(X|F_{XY}) - S(Y|F_{XY})] * [S(X|F_{XZ}) - S(Y|F_{YZ})] < 0$$

and

$$|S(X|F_{XZ}) - S(Y|F_{YZ})| \geq |S(X|F_{XY}) - S(Y|F_{XY})|.$$

The first condition says that the superior group in a pairwise comparison between X and Y, $X(Y)$, would score more points in a pairwise comparison with Z than would $Y(X)$ in the same pairwise comparison. The second condition says that the magnitude by which $X(Y)$ scores more points in a pairwise comparison with Z than does $Y(X)$ in the same comparison is greater than or equal to the magnitude by which $X(Y)$ scores fewer points than $Y(X)$ in a pairwise comparison between X and Y. Note that a strict violation of IIG occurs iff both inequalities in Proposition 3 hold strictly. An example of such a violation is provided in the outcome sequence below.

$$\hat{F} = (x_1, y_1, x_2, y_2, y_3, x_3, x_4, y_4, y_5, z_1, z_2, z_3, z_4, z_5, x_5).$$

In this case, $S(X|\hat{F} \setminus Z) = 27$ and $S(Y|\hat{F} \setminus Z) = 28$, while $S(X|\hat{F}) = 32$ and $S(Y|\hat{F}) = 28$. Worth noting is that the violation occurs due to the difference in the performance by the weakest elements of group X and Y versus the elements of Z. While $S(X|\hat{F} \setminus Y) = 20$, $S(Y|\hat{F} \setminus X) = 15$.

The relationship between how two teams compare pairwise with a third team has two further implications. First, it relates to violations of another social choice principle, transitivity, which we explore in the next section. Also, however, the relationship sheds additional light upon when violations of IIG can occur.

Broadly speaking, the larger the score disparity between any two teams, X and Y, the larger the disparity necessary in their pairwise performances against a third team, Z, in order for a violation to occur. In our preceding example, for instance, the disparity between the performance of X and Y against Z is relatively large, but it need not be for a violation to occur. To see this, simply consider the alternative example,

$$\hat{F}' = (z_1, z_2, z_3, x_1, y_1, x_2, y_2, y_3, x_3, x_4, y_4, y_5, z_4, z_5, x_5).$$

As before, X and Y are closely matched when Z is absent from the event, but the disparity between teams X and Y in terms of their pairwise performance versus Z is much smaller in this case, with $S(X|\hat{F}' \setminus Y) = 32$ and $S(Y|\hat{F}' \setminus X) = 30$. The disparity is still sufficient to meet the criteria outlined in Proposition 3, and a violation of IIG still occurs: $S(X|\hat{F}' \setminus Z) = 27$ and $S(Y|\hat{F}' \setminus Z) = 28$, while $S(X|\hat{F}') = 44$ and $S(Y|\hat{F}') = 43$. The second inequality in Proposition 3 thus illustrates how two closely (pairwise) matched groups present a greater potential for order disruption by a third group.

It is not entirely coincidence, then, that Hammond's (2007) empirical observations of IIG violations feature pairwise matches that are all very close in terms of their score, with nearly identical scores for each team in each pairwise match.[7] Specifically, Hammond (2007, p. 364) reports results from a cross country event that yield the following outcome sequence:

$$F^H = (x_1, y_1, x_2, y_2, z_1, z_2, y_3, z_3, z_4, x_3, y_4, x_4, z_5, x_5, y_5).$$

That sequence results in the following results.

$S(X|F^H \setminus Z) = 27$, $S(Y|F^H \setminus Z) = 28$, so $X \succ Y$ in the absence of Z;
$S(Y|F^H \setminus X) = 26$, $S(Z|F^H \setminus X) = 29$, so $Y \succ Z$ in the absence of X;
$S(Z|F^H \setminus Y) = 27$, $S(X|F^H \setminus Y) = 28$, so $Z \succ X$ in the absence of Y; and
$S(Y|F^H) = 39$, $S(X|F^H) = 40$, $S(Z|F^H) = 41$, so $Y \succ X \succ Z$ when all three compete.

Hammond's (real world) example thus entails two violations of IIG, as the orderings of X and Y and of Z and X are both flipped when all three teams compete. This is possible because both the X–Y and Z–X scores are separated by just one point, while there is just enough disparity in the Y–Z score to allow the second part of the condition in Proposition 3 to be satisfied. Increasing the disparity between Y and Z, for example by switching the order of elements y_3 and z_2, would allow all violations (and the above orderings) to remain. Increasing the disparity between X and Y, on the other hand, by switching the elements x_2 and y_1 (for example), would eliminate one of the violations, switching the result to $X \succ Y \succ Z$ when all three compete.

Extreme closeness in all pairwise scores is therefore not a necessary condition for violations of IIG to occur; rather, the right mix of closeness in some pairwise scores and (some) disparity in others is the recipe to ensure that a violation occurs. The disparity between teams cannot be too exaggerated, however, and yawning gaps in pairwise scores can eliminate the possibility of a violation of IIG for any two teams. We close this section by characterizing those limits.

How much disparity between any two teams is too much depends delicately on how the individual elements of two groups rank. To see how, note that the insertion of any element, $z \in Z$, into an outcome sequence will increase the rank of every subsequent element by exactly 1. The difference in the impact of that element on the scores of any two teams is therefore equal to the difference in the number of elements from each team that z precedes. This concept leads to the following result.

Proposition 4. *For any two teams, X and Y, such that $X \succ Y$ in pairwise competition for a given outcome sequence F, let ω be the largest integer such that $y_{i+\omega-1} \succ x_i$ in F. A strict violation of IIG by the inclusion of a third team Z is impossible if and only if $S(Y|F \setminus Z) - S(X|F \setminus Z) \geq n\omega$. Weak violations are impossible if and only if $S(Y|F \setminus Z) - S(X|F \setminus Z) > n\omega$.*

Proof. For the "if" part, note that the maximum differential impact the third team, Z, can have on the two teams' scores occurs when all n elements of Z rank better than x_i, but worse than $y_{i+\omega-1}$. That is, $y_{i+\omega-1} \succ z_1, \ldots, z_n \succ x_i$. In that case the difference between the two groups' scores is altered in favor of Y by precisely $n\omega$ when Z is included in the outcome sequence, since the ranks of elements y_i through $y_{i+\omega-1}$ are not affected while the rank of each element x_i through $x_{i+\omega-1}$ is increased by n. If the inequality in the statement holds even weakly, then at most a weak cycle is possible. If it holds strictly, there is no possibility that $S(X|F) > S(Y|F)$.

For the "only if" part, suppose $S(Y|F \setminus Z) - S(X|F \setminus Z) < n\omega$. In that case, a violation of IIG is always possible if all n elements of Z rank just in between $y_{i+\omega-1}$ and x_i. ■

Remark 1. Proposition 4 essentially means that violations of IIG can only be ruled out when one team is "too far" behind another. In fact, it is worth noting that if $X \succ Y$, ω can be at most $\frac{n-1}{2}$. This then provides an upper bound for the maximum score differential necessary to rule out strict violations of IIG. Similarly, if $X \sim Y$, ω can be at most $\frac{n}{2}$ (though attaining that bound requires n to be even), which provides the maximum lead necessary to rule out weak violations.

7 We thank an anonymous referee for pointing out this link between our results and Hammond's.

As an illustration of Proposition 4's implications, note that in the case of $\hat{F}$, it will always be the case that $X \succ Z$, regardless of the location of any elements of a third team. Similarly, in the case of $\hat{F}'$, it will always be the case that $Z \succ X$. In Hammond's observed example, on the other hand, ω is a mere 1, but so is $S(X|F \setminus Z) - S(Y|F \setminus Z)$. The presence of z_3 and z_4 after y_3 but before x_3 is then sufficient to reverse the relationship between Y and X in the presence of Z.

Remark 2. The mechanics of Proposition 4 can also be used to demonstrate violations of the Condorcet Rule, which specifies that if one group beats all others in pairwise competition, it should also defeat all others when all alternatives are considered together. More formally, $S(X|F_{xy}) < S(Y|F_{xy})$ and $S(X|F_{xz}) < S(Z|F_{xz})$ should imply that $S(X|F_{xyz}) < S(Y|F_{xyz})$ and $S(X|F_{xyz}) < S(Z|F_{xyz})$. If the Condorcet Rule does not hold then clearly IIG is violated, but just because IIG is violated does not necessarily mean that the Condorcet Rule does not hold. To show that the Condorcet rule can be violated in instances of rank sum scoring, and also to get an idea of when such violations may occur, let ω be defined as in the statement of Proposition 4 and let n be odd. Then suppose that $X \succ Y$ pairwise, that $S(Y|F_{xy}) - S(X|F_{xy}) < n\omega$, and that the maximum i such that $y_{i+\omega-1} \succ x_i$ is greater than $\frac{n}{2}$. Then if all n elements of Z rank ahead of x_i but after x_{i-1} and after $y_{i+\omega-1}$ then $X \succ Z$ pairwise, but $S(Y|F_{xyz}) < S(X|F_{xyz})$. An example of such a violation is illustrated by $\hat{F}$ above, where $\omega = 1$.

Although Hammond's (2007) example does not display the (perhaps) more egregious, Condorcet variety of IIG violations, it does illustrate violations of another major social choice principle: transitivity. The topic of transitivity is discussed next, and we will begin by defining violations of transitivity formally.

2.1. Transitivity in rank sum scoring

We begin this sub-section with a definition of a violation of transitivity in rank sum scoring.

Definition 2a (*Strict Violation of Transitivity under Rank Sum Scoring*)**.** Without loss of generality, we assume in the definition that $X \succ Y$ and $Y \succ Z$. Under this restriction, a strict violation of transitivity occurs under the following condition.

If $S(X|F \setminus Z) > S(Y|F \setminus Z)$, $S(Y|F \setminus X) > S(Z|F \setminus X)$, and $S(Z|F \setminus Y) > S(X|F \setminus Y)$, then a (strict) ranking cycle of the type $X \succ Y, Y \succ Z, Z \succ X$ exists in pairwise rank sum scoring.

Definition 2b (*Weak Violation of Transitivity under Rank Sum Scoring*)**.** Without loss of generality, we assume in the definition that $X \succcurlyeq Y$ and $Y \succcurlyeq Z$ (X is at least as highly ranked as Y and Y is at least as highly ranked as Z in pairwise scoring). When $X \succcurlyeq Y$ and $Y \succcurlyeq Z$, there are six types of weak violation, as listed below.

$$\{\{X \succ Y, Y \succ Z, Z \sim X\}, \{X \succ Y, Y \sim Z, Z \succ X\},$$
$$\{X \sim Y, Y \succ Z, Z \succ X\}, \{X \succ Y, Y \sim Z, Z \sim X\},$$
$$\{X \sim Y, Y \succ Z, Z \sim X\}, \{X \sim Y, Y \sim Z, Z \succ X\}\}.$$

The conditions for the first weak violation type are given as follows.

If $S(X|F \setminus Z) < S(Y|F \setminus Z)$, $S(Y|F \setminus X) < S(Z|F \setminus X)$, and $S(Z|F \setminus Y) = S(X|F \setminus Y)$, then a (weak) ranking cycle of the type $X \succ Y, Y \succ Z, Z \sim X$ exists in pairwise rank sum scoring.

We now identify the structural condition in outcome sequences that makes violations of transitivity possible, just as we previously did for the IIG condition. Before doing so, however, it is worth noting that violations of transitivity are intimately connected with violations of IIG. Hammond's example, F^H, serves to illustrate that violations of transitivity are indeed possible. The example also illustrates a case in which IIG is violated, however, which begs the question as to whether or not the two types of violation must occur together. Indeed, Hammond (2007) himself shows that as long as teams have more than 3 components ($n \geq 3$), both types of violations are possible. We now show that the connection goes further.

Proposition 5. *Any violation of transitivity under rank sum scoring must also lead to at least a weak violation of IIG.*

Proof. When team scores are based upon a complete outcome sequence, the corresponding ranking of teams that they generate cannot cycle, as the ranking of teams is simply based upon a weakly decreasing sequence of numbers. Thus, any time a cycle exists in pairwise scoring, it must be the case that at least one of the pairwise rankings is at least weakly reversed when teams are ordered based on their complete outcome sequence scores. ■

A corollary to Proposition 5 is that if an outcome sequence does satisfy the IIG property, then it must be acyclic. The reverse is not, true, however.

Proposition 6. *For $n \geq 3$, an outcome sequence may yield a violation of IIG under rank sum scoring and still satisfy transitivity.*

Proof. The proof is by construction. Consider the outcome sequence

$$\breve{F} = (x_1, z_1, z_2, \ldots, z_n, y_1, y_2, x_2, y_3, x_3, \ldots, y_{n-1}, x_{n-1}, y_n, x_n),$$

where $z_1, z_2, \ldots, z_n$ indicates the uninterrupted sequence of components from group Z, and $y_2, x_2, y_3, x_3, \ldots, y_{n-1}, x_{n-1}, y_n, x_n$ indicates an alternating pattern of elements from X and Y, with each component of Y coming ahead of its counterpart on Y. That is, $y_i \succ x_i \, \forall i > 1$.

Under rank sum scoring, $S(X|\breve{F} \setminus Z) - S(Y|\breve{F} \setminus Z) = n - 2$, so $Y \succ X$. Since all n components of Z rank after x_1 and before y_1, however, $S(X|\breve{F}) < S(Y|\breve{F})$. The inclusion of Z increases Y's score by n more than that of X, and thus there is a violation of IIG. The outcome sequence is free of any cycles, however, as group Z beats both X and Y in pairwise rank sum scoring.

Finally, note that additional groups could be included in this example without altering its results by simply assuming that all of their elements rank after x_n. ■

A violation of transitivity therefore necessarily implies at least a weak violation of IIG, but a violation of IIG does not necessarily imply a violation of transitivity. These results will prove to be important for interpreting our probability calculations in the next sections, but first we build upon them with our next theoretical result, which characterizes when violations of transitivity are and are not possible.

Proposition 7. *For $n \geq 3$ and odd, and for any two groups X and Y such that $X \succ Y$ in pairwise rank sum scoring, a strict violation of transitivity involving a third team, Z, is possible if and only if there exists at least one component $y_i \in Y$ such that $y_i \succ x_i$ in the outcome sequence F_{xyz}.*

Proof. We begin by showing that no matter which component $y_i \in Y$ defeats its counterpart, x_i, an outcome sequence can be constructed that violates transitivity. Worth noting is that the violations we construct hold even if y_i is the only component from group Y that defeats its counterpart on group X.

Throughout our explanation, it is helpful to keep in mind that, in rank sum scoring, any time a set of components x_j through $x_{j+\alpha}$ all rank ahead of y_j (and thus also y_{j+1} through $y_{j+\alpha}$), the difference between the two team scores based only on their components j through $j+\alpha$ is $(\alpha+1)^2$. Also, relating to this idea, we adopt the following notational choice for brevity: any time a sequence of elements x_j through $x_{j+\alpha}$ is uninterrupted by elements from another group in an outcome sequence, we indicate that by $x_j, \ldots, x_{j+\alpha}$.

There are five relevant cases. To differentiate them, let i^* indicate the index of the first component of group Y such that $y_i \succ x_i$. That is, for all $j < i^*$, $x_j \succ y_j$.

Case 1. $i = 1$. In this case simply consider F_{xyz} such that elements $z_1, \ldots, z_{\frac{n+1}{2}}$ all rank after y_1 but before x_1, and $z_{\frac{n-1}{2}}, \ldots, z_n$ all rank after the last-ranking element between the two groups (either x_n or y_n). Since $z_1, \ldots, z_{\frac{n+1}{2}} \succ x_1, \ldots, x_{\frac{n+1}{2}}$, group Z accumulates an insurmountable lead and $Z \succ X$ in pairwise competition. But even though $z_2, \ldots, z_{\frac{n+1}{2}} \succ y_2, \ldots, y_{\frac{n+1}{2}}$, that lead is eliminated (or "evened out") because $y_{\frac{n-1}{2}}, \ldots, y_n \succ z_{\frac{n-1}{2}}, \ldots, z_n$, and the fact that $y_1 \succ x_1$ means $Y \succ Z$ in pairwise competition. Since $X \succ Y$ by assumption, this yields the strict cycle $X \succ Y \succ Z \succ X$.

Case 2. $1 < i^* < \frac{n+1}{2}$. As in the first case, the components of Z can be arranged so that a cycle exists. To characterize that arrangement, we first identify a critical threshold index, λ, which is the minimum integer that satisfies $i^{*2} + \lambda^2 \geq (n - i^* - \lambda)^2$. Then consider F_{xyz} such that elements $z_1, \ldots, z_{n-\lambda}$ all rank just after y_{i^*} but before x_{i^*}, and $z_{n-\lambda+1}, \ldots, z_n$ all rank after the last-ranking element between the two groups (either x_n or y_n). This means that $y_1, \ldots, y_{i^*} \succ z_1, \ldots, z_{i^*}$ and $y_{n-\lambda+1}, \ldots, y_n \succ z_{n-\lambda+1}, \ldots, z_n$, thereby giving Y a "lead" of $i^{*2} + \lambda^2$ over Z based on those components alone. Thus, even though $z_{i^*+1}, \ldots, z_{n-\lambda} \succ y_{i^*+1}, \ldots, y_{n-\lambda}$, by the definition of λ it must be the case that $Y \succ Z$ in pairwise competition. In the case of X, however, only $x_1, \ldots, x_{i^*-1} \succ z_1, \ldots, z_{i^*-1}$ while $z_{i^*}, \ldots, z_{n-\lambda} \succ y_{i^*}, \ldots, y_{n-\lambda}$. Since the definition of λ means that $(i^*-1)^2 + \lambda^2 < (n-i^*-\lambda+1)^2$, it must be that $Z \succ X$ in pairwise competition, thereby producing the strict cycle $X \succ Y \succ Z \succ X$.

Case 3. $i = \frac{n+1}{2}$. In this case simply consider F_{xyz} such that all of the elements of Z, $z_1, \ldots, z_n$ rank after y_{i^*} but before x_{i^*}. Since $y_1, \ldots, y_{\frac{n+1}{2}} \succ z_1, \ldots, z_{\frac{n+1}{2}}$, that guarantees $Y \succ Z$ in pairwise competition; since $z_{\frac{n+1}{2}}, \ldots, z_n \succ x_{\frac{n+1}{2}}, \ldots, x_n$, that guarantees $Z \succ X$ in pairwise competition and thus produces the strict cycle $X \succ Y \succ Z \succ X$.

Case 4. $\frac{n+1}{2} < i^* < n$. For this case we must consider two subcases, though both are similar. First, the case when i^* is even.

If i^* is even, consider F_{xyz} such that $z_1, \ldots, z_{\frac{i^*}{2}}$ rank ahead of the first components of both X and Y. Then let $z_{\frac{i^*}{2}+1}, \ldots, z_{i^*+\frac{n-i^*-1}{2}}$ rank just after y_{i^*} but before x_{i^*}; then $z_{i^*+\frac{n-i^*+1}{2}}, \ldots, z_n$ rank last after both x_n and y_n. Since $z_1, \ldots, z_{\frac{i^*}{2}} \succ y_1, \ldots, y_{\frac{i^*}{2}}$ but then $y_{\frac{i^*}{2}+1}, \ldots, y_{i^*} \succ z_{\frac{i^*}{2}+1}, \ldots, z_{i^*}$, the score between the two teams becomes even up to that point. Then $z_{i^*+1}, \ldots, z_{i^*+\frac{n-i^*-1}{2}} \succ y_{i^*+1}, \ldots, y_{i^*+\frac{n-i^*-1}{2}}$, but $y_{i^*+\frac{n-i^*+1}{2}}, \ldots, y_n \succ z_{i^*+\frac{n-i^*+1}{2}}, \ldots, z_n$ which means $Y \succ Z$ pairwise (since $n - i^*$ is odd).

To see that $Z \succ X$ pairwise, note that $z_1, \ldots, z_{\frac{i^*}{2}} \succ x_1, \ldots, x_{\frac{i^*}{2}}$ but then $x_{\frac{i^*}{2}+1}, \ldots, x_{i^*-1} \succ z_{\frac{i^*}{2}+1}, \ldots, z_{i^*-1}$, so the score of X will exceed that of Z by $(\frac{i^*}{2})^2 - \left(\frac{i^*}{2} - 1\right)^2 = i^* - 1$. The remaining elements then simply keep the score differential steady, since $z_{i^*}, \ldots, z_{i^*+\frac{n-i^*-1}{2}} \succ x_{i^*}, \ldots, x_{i^*+\frac{n-i^*-1}{2}}$, is just offset by $x_{i^*+\frac{n-i^*+1}{2}}, \ldots, x_n \succ z_{i^*+\frac{n-i^*+1}{2}}, \ldots, z_n$. Again, $Z \succ X$ and the strict cycle $X \succ Y \succ Z \succ X$ is produced.

If i^* is odd, consider F_{xyz} such that $z_1, \ldots, z_{\frac{i^*-1}{2}}$ rank ahead of the first components of both X and Y. Then let $z_{\frac{i^*+1}{2}}, \ldots, z_{i^*+\frac{n-i^*}{2}}$ rank just after y_{i^*} but before x_{i^*}; then $z_{i^*+\frac{n-i^*}{2}+1}, \ldots, z_n$ rank last after both x_n and y_n. Since $z_1, \ldots, z_{\frac{i^*-1}{2}} \succ x_1, \ldots, x_{\frac{i^*-1}{2}}$ but then $x_{\frac{i^*+1}{2}}, \ldots, x_{i^*-1} \succ z_{\frac{i^*+1}{2}}, \ldots, z_{i^*-1}$, the score between X and Z is even up to that point. Then $z_{i^*}, \ldots, z_{i^*+\frac{n-i^*}{2}} \succ x_{i^*}, \ldots, y_{i^*+\frac{n-i^*}{2}}$, which more than offsets $x_{i^*+\frac{n-i^*}{2}+1}, \ldots, x_n \succ z_{i^*+\frac{n-i^*}{2}+1}, \ldots, z_n$ so that $X \succ Z$ in pairwise competition (since $n - i^*$ is even).

To see that $Y \succ Z$ pairwise, note that $z_1, \ldots, z_{\frac{i^*-1}{2}} \succ y_1, \ldots, y_{\frac{i^*-1}{2}}$ but then $y_{\frac{i^*+1}{2}}, \ldots, y_{i^*} \succ z_{\frac{i^*}{2}+1}, \ldots, z_{i^*}$, so the score of Z will exceed that of Y by $\left(\frac{i^*+1}{2}\right)^2 - \left(\frac{i^*-1}{2}\right)^2 = i^*$. The remaining elements then simply keep the score differential steady, since $z_{i^*+1}, \ldots, z_{i^*+\frac{n-i^*}{2}} \succ x_{i^*+1}, \ldots, x_{i^*+\frac{n-i^*}{2}}$, is just offset by $x_{i^*+\frac{n-i^*}{2}+1}, \ldots, x_n \succ z_{i^*+\frac{n-i^*}{2}+1}, \ldots, z_n$. Again, $Y \succ Z$ and the strict cycle $X \succ Y \succ Z \succ X$ is produced.

Case 5. $i^* = n$. In this case consider F_{xyz} such that $z_1, \ldots, z_{\frac{n-1}{2}}$ rank first ahead of all elements of X and Y. Then $z_{\frac{n+1}{2}}, \ldots, z_n$ rank just after y_n but before x_n. This outcome means that $z_{\frac{n+1}{2}}, \ldots, z_n \succ x_{\frac{n+1}{2}}, \ldots, x_n$, guaranteeing that $Z \succ X$ in pairwise competition. To see that $Y \succ Z$ pairwise, note that the teams' (pairwise) scores remain tied through their first $n - 1$ components, then $y_n \succ x_n$. Thus, $X \succ Y \succ Z \succ X$.

We have now shown exhaustively that, for any two teams with odd $n \geq 3$ such that $X \succ Y$ in pairwise rank sum scoring, cycles are always possible so long as at least one element from Y defeats its counterpart with the same index from X. This result holds no matter which element that is and regardless of the relative rankings of any other elements. To see that cycles are impossible otherwise, simply note that if every element $x_i \in X$ defeats its counterpart $y_i \in Y$, any and all elements from a third team, Z, which defeat components of X must also defeat the same components of Y. Any outcome sequence that leads to $Z \succ X$ in pairwise competition must also lead to $Z \succ Y$. ■

Several comments are in order regarding Proposition 7. First, in line with Proposition 5, it can be verified that all of the outcome sequences we construct in the proof of Proposition 7 that lead to violations of transitivity also lead to violations of IIG. Which specific pairwise comparison will be altered when all three teams compete at once, however, depends on additional details of the outcome sequences in question. Conversely, if there does not exist any $y_i \in Y$ such that $y_i \succ x_i$, this corresponds with the case of $\omega = 0$ and neither violations of transitivity or IIG are possible.

Second, though Proposition 7 specifies n odd, our proof can easily be adapted to provide a similar result for the case of n even. The caveat, however, is that when n is even the cycles our process finds may only be weak in nature. For example, consider the following outcome sequence with $n = 6$ and $i^* = 3$, $\{x_1, x_2, x_3, y_1, y_2, y_3, y_4, x_4, x_5, x_6, y_5, y_6\}$. In this case our treatment for Case 3 above can easily be adapted: simply insert $z_1, \ldots, z_6$ between y_4 and x_4. That outcome sequence yields a cycle, but a weak one: $X \succ Y \succ Z \sim X$. This does not mean that strict cycles are not possible with n even, of course; indeed, they most certainly are. But it does provide one explanation as to why our computational results show that, for the cases we consider, strict cycles are less likely when n is even.

Finally, although Proposition 7 reveals that the conditions under which violations of transitivity can occur are surprisingly broad, just because they can occur does not mean they are necessarily likely.

3. Probability calculations

We now consider the likelihood that an outcome sequence generates a (strict) social choice violation. We first consider the

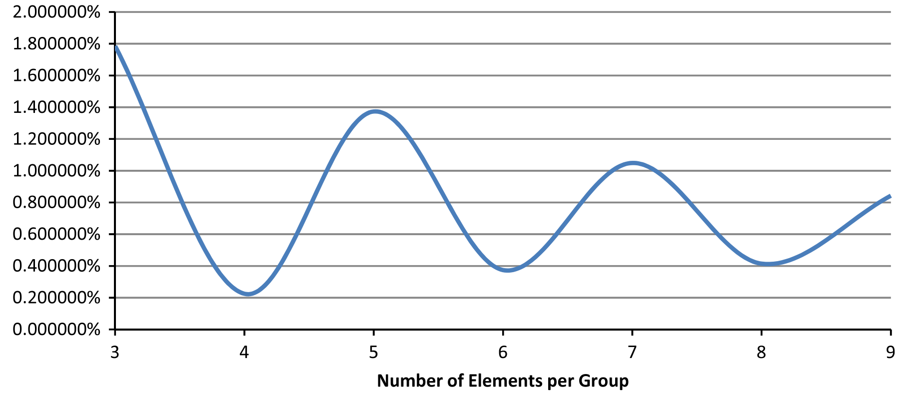

Fig. 1. Likelihood of cycles.

accuracy of Hammond's (2007) calculations on the topic—the only previous such calculations. Hammond (2007) manually calculates the likelihoods of the considered social choice violations in the case of rank sum scoring for a subset of outcome sequences. It is important to note that pairwise scoring "ties" are impossible in the cases considered by Hammond. Hence, the likelihood of a weak social choice violation is zero in such cases. He states, "…a complete manual count was made of all possible orders of finish in which team A beats team B and team B beats team C when (there are three teams and three runners per team). There are 290 of these orders of finish, and it turns out that for only 15 of these 290 orders of finish (5.17%) does team C beat team A, thereby generating a cycle and a violation of independence".[8] Hammond considers three teams: $\{A, B, C\}$. By "team A beats team B and team B beats team C", Hammond means that $S(A|F \setminus C) < S(B|F \setminus C)$ and $S(B|F \setminus A) < S(C|F \setminus A)$. This is apparent because three-team cycles are not possible when three groups are scored simultaneously. In light of Proposition 6 of the present study, it is curious that Hammond treats the two violations as mutually inclusive.

To check his result, we generated a Java program (see Appendices A and B) to determine the likelihood that a sequence generates a cycle or a violation of IIG. For the subset of sequences considered by Hammond, we confirm the likelihood of a (strict) cycle to be 5.17% (15 of 290 sequences). However, results differ substantially when the analysis is not conditioned (i.e., over the entire set of outcome sequences). There are $(9!/(3!3!3!)) = 1680$ possible orderings in the 3 group, 3 element per group case.[9] Of these orderings, only 30 produce a cycle. In other words, cycles are found in only 1.79% (rather than 5.17%) of all sequences in this case. Hammond does not calculate the probability of a cycle in the 3 group, 5 element per group case.[10] He states, "The greater degrees-of-freedom for the freedom for the $m \geq 3$ and $k = 5$ case, compared to the $m = 3$ and $k = 3$ case, might yield a substantially higher probability of cycles (though that is a matter for future determination)".[11] There are $(15!/(3!3!3!)) = 756,756$ orderings in this latter case. We find that 10,392 of these orderings (1.37%) produce cycles. Thus, the likelihood of a cycle actually decreases from the 3 group, 3 element per group case to the 3 group, 5 element per group case. Table 1 provides a fairly extensive listing of cycle probabilities by case, including some cases that are highly computationally intensive.

Table 1
Likelihood of a 3-group cycle in a 3-group comparison.

Number of groups	Elements per group	Likelihood of a cycle
3	3	$30/1680 = 1.79\%$
3	4	$78/34,650 = 0.23\%$
3	5	$10,392/756,756 = 1.37\%$
3	6	$64,230/17,153,136 = 0.37\%$
3	7	$4,186,398/399,072,960 = 1.05\%$
3	8	$39,236,706/9,465,511,770 = 0.41\%$
3	9	$1,920,331,578/227,873,431,500 = 0.84\%$

Table 2
Likelihood of a 3-group cycle in a 4-group comparison.

Number of groups	Elements per group	Likelihood of 3-group cycle
4	3	$22,320/369,600 = 6.04\%$
4	4	$535,152/63,063,000 = 0.85\%$
4	5	$560,653,608/11,732,745,024 = 4.78\%$

We note that the trend of declining likelihood continues as long as n (the number of elements per group) is odd, into the 3 group, 7 element case and the 3 group, 9 element case. When n is even, however, the reverse happens. With precisely 4 elements cycles only occur in 78 out of 34,650 possible orderings. As n increases by two to 6 and 8, strict cycles appear more and more, though they remain significantly less likely than in the cases we consider with n odd. As per our second comment following Proposition 7, we believe this is at least in part due to similar outcome sequences resulting in weak cycles when n is even as opposed to strict cycles when n is odd. The wave-like pattern (see Fig. 1) of decreasing likelihood of cycles for n odd, increasing likelihood for n even, may suggest that the likelihood converges as the number of elements per group becomes large (perhaps because weak cycles become relatively less likely), but that is only a conjecture at this point since exhaustive computational analysis becomes prohibitively taxing for cases when n is large.

We now briefly explore the likelihood that a sequence generates at least one three-team cycle when four teams are compared in a pair-wise manner. These calculations, which are presented in Table 2 and which involve more computations per outcome sequence, suggest that the likelihood of at least one three-group cycle may become quite high in likelihood as the number of groups rises. For the specific case of four groups, however, we find once again that the likelihood of a three-group cycle declines, albeit with a wave pattern as n switches from odd to even, as we move to a higher number of elements per group.

Why are Hammond's probability calculations incorrect? When he restricts the analysis to all sequences such that $S(A|F \setminus C) < S(B|F \setminus C)$ and $S(B|F \setminus A) < S(C|F \setminus A)$, he omits sequences

8 A cycle is not a necessary condition for a violation of IIG. Therefore, we treat this calculation as reflecting only the prevalence of a cycle.

9 The Putnam Mathematics Team Competition involves 3 scoring members per team.

10 Cross country races typically involve 5 scoring members per team.

11 Hammond (2007) uses m to refer to the number of teams (or groups), and k to refer to the number of elements on each team.

Table 3
Likelihood of a violation of IIG with 3 groups.

Number of groups	Elements per group	Likelihood of 3-group cycle
3	3	$78/1680 = 4.64\%$
3	5	$62,106/756,756 = 8.21\%$

that cannot produce cycles. By instead imposing the restriction that $S(A|F \setminus C) < S(B|F \setminus C)$ and $S(B|F \setminus A) > S(C|F \setminus A)$, one could conclude that the likelihood of a cycle is 0 (i.e., there can be no cycles in sequences satisfying this restriction). In the 3 group, 3 element per group case, we conclude that cycles are expected in roughly 1 of every 56 orderings rather than "once in every 20 meets" (Hammond, 370). In the 3 group, 5 element per group case, cycles are expected in roughly 1 of every 73 orderings. In other words, the generalization and expansion of the set computation are crucial (see Tables 1–3).

In light of our Proposition 5, which states that a (strict) cycle implies a (strict) violation of IIG but a violation of IIG does not imply a cycle, we expect strict violations of IIG to be potentially more prevalent than strict violations of transitivity. For comparison, therefore, we calculate the likelihood of a (strict) violation of IIG whether or not a violation of transitivity also occurs. We find that at least one strict violation of IIG occurs in 78 of 1680 outcome sequences in the 3 group, 3 element per group case (4.64% of sequences) and in 62,106 of 756,756 possible outcome sequences in the 3 group, 5 element per group case (8.21% of sequences). Thus, violations of IIG and violations of transitivity are quite distinct in the case of rank sum scoring. Violations of IIG occur in 2.6 times as many outcome sequences in the 3 group, 3 element per group case, and in roughly 5.98 times as many outcome sequences in the 3 group, 5 element per group case. Thus, there is not even a stable relationship between the likelihood of one violation relative to the other.

4. Conclusion

This analysis informs us that violations of independence from irrelevant groups are much more prevalent than violations of transitivity in considered cases of rank sum scoring. Moreover, our theoretical analysis demonstrates that the two violations arise from distinct (though not entirely unrelated) sets of conditions.

Hammond (2007) closes his study by discussing the salience of social choice violations in rank sum scoring. Hammond compares practitioner beliefs as to the likelihood of social choice violations in rank sum scoring to their calculated likelihood. The results of this comparison provide a commentary upon the salience of social choice violations. Namely, the comparison informs us as to whether practitioners accept a flawed method of outcome aggregation (a) because they do not know it is flawed or (b) because they understand it is flawed but lack alternatives that are clearly Pareto-improving.[12] Cross country coaches interviewed by Hammond largely took both violations as exceedingly rare. More than one veteran coach had never perceived the occurrence of either social choice violation in rank sum scoring. None of the interviewed coaches suggested an understanding that the two violations may be distinct in any manner.

In support of Hammond (2007), therefore, the results of this study provide additional evidence that social choice violations in rank sum scoring may be accepted at least partly due to a lack of salience. Indeed, Wilcoxon (1945), Mann and Whitney (1947), and Wilcoxon et al. (1970) fail to mention the possibility of a ranking cycle in presenting the Wilcoxon rank sum test. Moreover, the Mathematical Association of America, a society of mathematicians, accepts rank sum scoring in administering the annual Putnam Mathematics Team Competition without apparent discussion of potential social choice violations. Indeed, an exhaustive study of rank sum scoring may be necessary to distinguish between the two violations.

We close by acknowledging that other scoring or group-comparison methods might avoid the types of social choice violations that we highlight in this paper. Hammond (2007) notes in particular that, in the case of cross country team competitions, simply summing the actual times of runners as opposed to their order of finish would avoid violations of IIG or transitivity. That alternative entails its own drawbacks, however. Most prominently, it would be vulnerable to violations of a "no-dictators" criterion, since a substantially good or bad time by one runner could by itself determine the ranking of any two teams. Indeed, that vulnerability is specifically mentioned by cross country coaches interviewed by Hammond (2007, p. 368) as a reason to prefer the rank sum method over a sum of times method.

More generally, in light of Arrow's (1951) well-known results, it is very unlikely that any one scoring method will be completely free of social choice violations. Szydlik (2010), for example, establishes a variety of impossibility theorem for cross country competitions involving three or more teams, showing that no cross country scoring method can satisfy IIG and a strong Pareto criterion (if $x_i \succ y_i \; \forall i$ then $X \succ Y$) unless it is vulnerable to dictatorship (one runner's finish completely determining the ranking of two teams). Thus, while more general considerations of alternative scoring methods and social choice criteria are certainly worthy avenues for future research (e.g. Boudreau and Sanders, 2013), they are beyond the scope of this paper.

Appendix A. Java code supporting violation of transitivity probability calculations

To calculate the number of cycles possible, we implemented a Java program that makes a complete enumeration of every possible generic outcome, e.g., $A_1A_2A_3B_1B_2B_3C_1C_2C_3 = A_3A_2A_1 B_3B_2B_1C_3C_2C_1$, and then checks for a strong cycle. These generic outcomes are generated with a tree, where each outcome is a path, and this path is valid if and only if its height is the height of the tree. See Fig. A.1 for two valid generic outcomes (the slashes represent trimmed branches that would start invalid outcomes). A recursive algorithm was implemented to concurrently create every possible valid path and test this path for a cycle for both three and four member group. The valid paths are generated by recursively branching every possible continuation at every node, where the edges represent each possible participant. A string recording the path is appended by a letter representing each participant and is passed to the recursive call. Each recursive call repeats this step unless one of two events terminates the recursion: (1) the string represents an invalid subtree, (2) the string contains a full, valid branch. In the first scenario, there are too many participants of one of the teams and the branch is removed. The second scenario happens at the terminating node, where the string represents a full generic outcome. This outcome is checked for a cycle and the result is recorded. Fig. A.2 demonstrates the Java code executed at each node (ASCII character 97 is the letter 'a').

Checking for an outcome where three teams are involved is straightforward and is accomplished in two steps. First, the program independently compares each of the three possible pairwise scores and accumulates each win for each group. The win is determined by checking the relative placement of each player of the two groups independently of the rest of the players (Fig. A.2). Then the number of wins is checked for each group. If any of them

[12] We assume that practitioners have incentives to seek the "best" outcome aggregation method in terms of social choice characteristics.

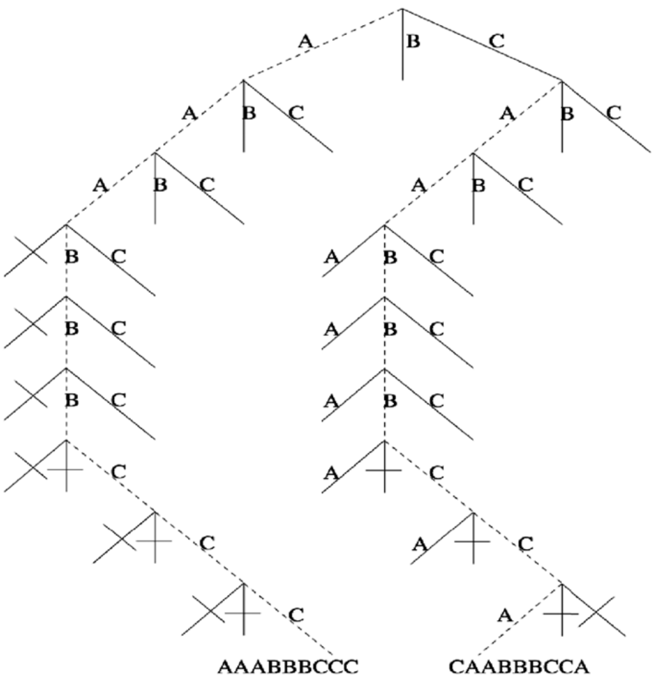

Fig. A.1. Partial tree showing two valid outcomes: AAABBBCCC and CAABBBCCA.

```
private void generate(String runners)
{
    for(int i=0; i<numTeams; i++)
    {
        if(countOccurrences(runners,(char)(i+97)) > numRunners)
        {
            return;
        }
    }
    if (runners.length() < numTeams * numRunners)
    {
        for(int i=0; i<numTeams; i++)
        {
            generate(runners+(char)(i+97));
        }
    }
    else
    {
        findCycles(runners);
    }
}
```

Fig. A.2. The tree generator used to enumerate every possible generic outcome.

```
private boolean isWinner(char x, char y, String str)
{
    int xScore=0;
    int yScore=0;
    int yPos=0;
    for(int xPos=0; xPos<str.length(); xPos++)
    {
        if(str.charAt(xPos)==x)
        {
            while((yPos<str.length())&&(str.charAt(yPos)!=y))
            {
                yPos++;
            }
            if(yPos<xPos)
            {
                yScore++;
            }
            else
            {
                xScore++;
            }
            yPos++;
        }
    }
    return (xScore>yScore);
}
```

Fig. A.3. Checking for a win by relative placements of individuals from two groups.

```
if(isIndependentWinner('a','b',tempStr))//a>b
    {
        if(isDependentWinner('b','a',tempStr))
        {
            numStrongViolations++;
        }
        else if(isDependenteEqual('b','a',tempStr))
        {
            numWeakViolations++;
        }
    }
```

Fig. B.1. Test for the violation of independence.

```
private boolean isIndependentWinner(char x, char y, String str)
{
    int xScore=0;
    int yScore=0;
    int counter=1;
    for(int i=0; i<str.length(); i++)
    {
        if(str.charAt(i)==x)
        {
            xScore+=counter;
            counter++;
        }
        else if(str.charAt(i)==y)
        {
            yScore+=counter;
            counter++;
        }
    }
    return (xScore<yScore);
}
```

Fig. B.2. Test for win when ranked without other team(s).

```
private boolean isDependentWinner(char x, char y, String str)
{
    int xScore=0;
    int yScore=0;
    int counter=1;
    for(int i=0; i<str.length(); i++)
    {
        counter++;
        if(str.charAt(i)==x)
        {
            xScore+=counter;

        }
        else if(str.charAt(i)==y)
        {
            yScore+=counter;
        }
    }
    return (xScore<yScore);
}
```

Fig. B.3. Test for win when ranks among other team(s) are considered.

has 0 wins, there is no cycle; otherwise there must be a cycle, as this is the necessary and sufficient condition for a three group cycle.

Checking the outcome where four teams are involved is slightly more complicated. First, the program independently compares each of the six possible pairwise scores (2^6 possibilities) and accumulates each win and loss for each group. Then, the number of wins and losses is checked for each group. If at least three have at least one win and one loss, there is either a three or four group cycle, fulfilling the necessary and sufficient condition. This last portion is summarized in Fig. A.3.

Appendix B. Java code supporting violation of independence probability calculations

Similar to testing the number of transitivity violations, we developed a Java application to compute the number of possible independence violations. Each possible generic outcome from the tree algorithm was tested for a violation of independence. A violation of independence occurs when the outcome of two groups is dependent on the relative placement of one or more additional groups. We can test this by first independently ranking the two groups by removing the other teams. This outcome is then compared by ranking the two groups while including the other group(s). If the result changes across comparisons, a violation of independence occurred. Fig. B.1 lists the code for this basic test. Respectively, Figs. B.2 and B.3 list code when a given pair of groups is ranked without (with) other groups.

References

Arrow, K.J., 1951. Social Choice and Individual Values, Vol. 12. Yale University Press.

Boudreau, J.W., Sanders, S.D., 2013. Competitive balance and ranking cycles. Mimeo.

Gordon, S., Truchon, M., 2008. Social choice, optimal inference and figure skating. Soc. Choice Welf. 30 (2), 265–284.

Hammond, Thomas., 2007. Rank injustice? How the scoring method for cross-country running competitions violates major social choice principles. Public Choice 133 (3–4), 359–375.

Huntington, E.V., 1938. A paradox in the scoring of competing teams. Science 88 (2282), 287–288.

MacKay, A.F., 1980. Arrow's Theorem, The Paradox of Social Choice: A Case Study in the Philosophy of Economics. Yale University Press, New Haven.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60.

Mixon Jr., F.G., King, E.W., 2012. Social choice theory in 10,000 meters: examining independence and transitivity in the NCAA cross-country championships. Amer. Econ. 57 (1), 32.

Saari, D.G., 2001. Decisions and Elections: Explaining the Unexpected. Cambridge University Press.

Szydlik, S., 2010. May the best team win: determining the winner of a cross country race. In: Galian, J.A. (Ed.), Mathematics and Sports. In: Dolciani Mathematical Expositions, vol. 43. Mathematical Association of America, Washington, DC, pp. 295–314.

Truchon, M., 1998. Figure skating and the theory of social choice. Cahier, 9814.

Truchon, M., 2004. Aggregation of rankings in figure skating. Cahier, 0402.

Truchon, M., 2008. Borda and the maximum likelihood approach to vote aggregation. Math. Social Sci. 55 (1), 96–102.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1 (6), 80–83.

Wilcoxon, F., Katti, S.K., Wilcox, R.A., 1970. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. In: Selected Tables in Mathematical Statistics, Vol. 1. pp. 171–259.