May the Best Team Win: Determining the Winner of a Cross Country Race

Stephen Szydlik

Abstract

Finding the winner of an athletic contest or sporting event should be a simple process: the competitor with the most runs, goals, points, or the quickest time should be the winner. In practice, however, choosing a team winner can be a challenge, especially when many teams are involved in a tournament. In this article, we explore some of the specific dilemmas associated with finding the winning team in a cross country running race.

Standard scoring of a cross country race is straightforward. A team typically consists of seven runners, and a team's score is the sum of the placings of its first five runners. A team's sixth and seventh runners do not score points towards their team's total, but their place can serve to increase the team score of their opponents. Teams are ranked by the order of their scores from lowest to highest. While teams often face off head-to-head in dual meets, invitationals of up to 30 teams are also common.

Though simple to implement, this race scoring system can yield surprising and somewhat counterintuitive outcomes. Chief among these are failures of *binary independence*: the relative ranking of two teams in an invitational can depend upon the presence and performance of the other teams in the race. For example, one team (the Acorns) might finish ahead of a second team (the Buckeyes) when a third team (the Chestnuts) has a good day. However, with identical races by the Acorns and Buckeyes, it is possible for the Acorns to finish behind the Buckeyes if the Chestnuts instead have a bad day. Troubling situations such as these are relatively common in cross country running, and are well known to coaches and interested observers. In this article, we study these issues and consider some alternatives to the standard race scoring methods.

Social choice theory, an area of study on the interface of mathematics, political science, and economics, provides a useful lens through which we can examine cross-country race scoring. One particular aspect of this discipline is concerned with studying the science of decision-making in voting and elections, and comparisons with cross-country scoring are striking. It is arguable, for example, that had Ralph Nader not been a candidate for the U.S.

predisdency in the 2000 election, then Al Gore would have beaten George W. Bush. Our electoral system violates binary independence in precisely the way that cross-country scoring does. We examine parallels between cross-country scoring and voting methods with an eye towards Kenneth Arrow's seminal result on the impossibility of finding a completely fair democratic election method. We also draw some important distinctions between the two subjects, particularly in the case of a two-team race.

Warming Up

Collegiate cross country running in the upper Midwest is highly competitive. In the 36 years (through the fall of 2008), that the National Collegiate Athletic Association (NCAA) has sponsored a Division III (non-scholarship) men's team championship, the top team has come from Illinois, Wisconsin, or Michigan 26 times [6]. The University of Wisconsin-Oshkosh had won three championships in the years leading up to the fall of 2001, when this story begins. The Oshkosh team that year had a great mix of talent and experience, and I had hopes that they would contend for the national title. One early-season test for the Titans was at the Jim Drews Invitational, a race bringing together much of the top Midwest Division III talent. The event was hosted by Wisconsin-LaCrosse, a perennial running powerhouse and nemesis of Oshkosh. Also competing was a team from the University of Wisconsin-Madison, a strong Division I school.

The race took place on a wet Saturday in October, with 33 teams and 223 team runners racing over a 5 mile (8 kilometer) course. At the conclusion of the event, the UW-Madison runners had finished in places 1, 2, 3, 8, and 27, those from UW-LaCrosse had finished in positions 4, 12, 15, 24, 35, 49, and 55, and UW-Oshkosh runners had finished in places 10, 11, 13, 28, 30, 43, and 69. Scoring an NCAA cross country meet is straightforward, with three fundamental rules (See [3, p. 122]):

1. A team's score is the sum of the placings of their first five runners.

2. Teams are ranked by the order of their scores from lowest to highest.

3. Although the sixth and seventh runners of a team to finish do not score points toward their team's total, their place, if better than those of any of the first five of an opposing team, serve to increase the team score of the opponents. When it occurs, this situation is called *displacement*.

Using this scoring method, UW-Madison won the invitational with an aggregate score of 41, UW-LaCrosse's total of 90 placed them second, and UW-Oshkosh was a close third with 92 points. The results were disappointing for me, but only because I had hoped Oshkosh would beat UW-LaCrosse. Beating UW-Madison was not a realistic goal, and was in fact irrelevant, as the teams compete in different divisions and would not race each other again. This led to an interesting question: if we eliminated the UW-Madison runners from the race results, would the result be the same? After such a change, UW-LaCrosse runners move up to respective places 1, 8, 11, 20, 30, 44, and 50, while the UW-Oshkosh runners improve to positions 6, 7, 9, 23, 25, 38, and 64, so the teams tie for first place, with 70 points each. Somehow the presence of UW-Madison runners was detrimental to the UW-Oshkosh cause, not just because UW-Madison beat UW-Oshkosh, but also because they helped UW-LaCrosse relative to UW-Oshkosh.

What happens if we ignore all other teams except for LaCrosse and Oshkosh? If we are really trying to determine which of the two teams was stronger on that October day, it makes sense to consider them head-to-head, in a two-team race. A way to picture this reduced race is as a finite sequence of runners:

$$\left\langle \begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ L_1, & O_1, & O_2, & L_2, & O_3, & L_3, & L_4, & O_4, & O_5, & L_5, & O_6, & L_6, & L_7, & O_7 \end{array} \right\rangle$$

Here O_i and L_i represent the ith finishers for Oshkosh and LaCrosse, respectively, and the numbers o_i and ℓ_i above them represent their respective finishing positions. Scoring this head-to-head race yields 27 points for Oshkosh and 28 points for LaCrosse. If Oshkosh and LaCrosse had been the only two teams on the course that day, Oshkosh would have won!

Races like the Jim Drews Invitational occur in high school and collegiate cross country meets across the country many times every fall weekend, and scoring oddities such as those involving UW-Oshkosh and UW-LaCrosse are well known to coaches and interested observers. In this article, we will analyze some of the mathematics of cross country scoring, consider criteria that measure the reasonableness of scoring methods, and explore alternative scoring methods. Readers familiar with the language and theorems of social choice theory will note strong parallels to that area of mathematics, and we will borrow liberally from its terminology. We will also, however, draw some strong distinctions between cross country scoring and vote counting.

Mile 1: Basic Terminology and Some Alternatives

At the start of a large invitational race, runners need to balance carefully two goals: they need to begin quickly enough to establish a favorable position, but not so quickly that they tire and fade at the end of the race. So as the starting gun goes off, we clarify what we mean by a "scoring method," and formalize some of our notations.

Definition 1 *Let $\mathcal{T} = \{T_1, T_2, \ldots, T_n\}$ be a set of n teams, each consisting of k runners. A* race *is an ordered list of the $n \cdot k$ runners. A* scoring method *is a function that accepts a race as an input and produces as output a linear ordering of $\mathcal{T}$. We shall use $T_{i_1} \prec T_{i_2} \prec \cdots \prec T_{i_n}$ to denote the ordering, where team T_{i_1} finishes first, team T_{i_2} finishes second, and so on.*

In this article, we will insist that the ordering be strict. While team ties occur during meets, they are easy to break. For example, the NCAA tiebreaking rule is to compare the tying teams' fifth runners and break the tie in favor of the team whose fifth runner finishes ahead of the other(s). For simplicity, we will also consider only scoring methods that are symmetric with respect to the teams, and anonymous with respect to the runners. By *symmetric*, we mean that the names of the teams are irrelevant, i.e., interchanging the runners of teams A and B in a race will interchange A and B in the team ranking outcome. By *anonymous*, we mean that if any two runners from a team interchange places, the outcome remains the same. Anonymity guarantees that there are $({}_{nk}C_k)({}_{n(k-1)}C_k)\cdots({}_kC_k)$ essentially distinct races.

In scoring a meet, the choice by the NCAA to use five scoring runners and two displacing runners is somewhat arbitrary. The simplest change we can make to the usual scoring method is to change those numbers: let us call a method that uses standard scoring but with m scoring runners and ℓ displacing runners an (m, ℓ)-standard scoring system. For example, the recent world cross country championships in Amman, Jordan in March 2009 used a $(4, 2)$-scoring system (see [4]) to determine the team champion. The choice of k and ℓ can have a great impact on deciding the winning team in a race even if there are only two teams involved (a *dual meet*). In the head-to-head matchup between UW-Oshkosh and UW-LaCrosse, Oshkosh would win using $(5, 2)$, $(6, 1)$ or $(3, 0)$-standard scoring, whereas UW-LaCrosse would win (by virtue of the standard tiebreaker) using $(4, 0)$ scoring. UW-LaCrosse would also win using $(1, 0)$ standard scoring (since LaCrosse's top finisher beat Oshkosh's top finisher) but that is hardly an acceptable method because it depends only on the relative finishing positions of one runner from each team (in this case, the winning team is the team whose first runner finishes first). The notion of single-runner dominance in a race will be important for us later, so let us formalize this idea:

Definition 2 *A scoring method is called* non-team oriented *if there is some i, $1 \leq i \leq k$ such that for every race, the order of the teams is determined solely by the relative positions of the ith finisher for the teams. That is, for all teams A, B in an invitational meet, $A \prec B \Leftrightarrow a_i < b_i$.*

For example, in a race involving five runners per team where we ignore the placings of the first four runners from each team and rank the teams based solely on the ordering of their fifth runner would be a non-team oriented scoring method. We would like to avoid such methods!

The Jim Drews Invitational example points to difficulties with the standard invitational scoring methods. Most obvious is that one team's performance relative to another team can depend on the other teams in the field (as UW-Oshkosh's performance relative to UW-LaCrosse was influenced by the presence of UW-Madison). We call this a failure of the *Binary Independence* condition.

Binary Independence condition: *The relative ranking of teams A and B in the race results should not depend upon any other team.*

More precisely, suppose a race takes place in which team A beats team B in the final ranking. Shortly thereafter a second race takes place with the same teams and the same runners, and in that race some runners change positions, but the relative positions of runners from team A and B don't change at all. Then team A should still beat team B in the rankings.

A scoring method should ideally possess this property because deciding whether team A or team B finishes higher in the rankings should not depend on factors such as injury, illness, stumble, or poor performance by a member of team C. If the relative ranking of A and B is never affected by the positions of other runners in any race, then the scoring method is said to satisfy the Binary Independence condition. We use the word "should" in its statement (and in our statements of other conditions in this article) to emphasize that, while desirable, the property is not possessed by all scoring methods. The Jim Drews example shows that $(5, 2)$ standard scoring violates the Binary Independence condition, since moving UW-Madison runners to the end of the race changes the relative ranking of UW-LaCrosse and UW-Oshkosh to a tie, broken in favor of UW-Oshkosh. The other (m, ℓ) standard scoring methods are similarly flawed.

The Binary Independence condition suggests an alternative to the standard invitational scoring methods. Given a race with two or more teams, why not match each team head-to-head with every other team, and declare the team that wins all of those matchups as the winner of the invitational? This amounts to scoring a series of dual meets, with the winner of the invitational being the team that wins all its head-to-head matchups. Second place could be given to the team that wins all but one of its head-to-heads, third to the team that beats all but the top two teams head-to-head, and so on.

Though attractive at first glance, this scoring mechanism has a fatal problem. Consider the following race between three teams, A, B, and C:

Example 1

$$\left\langle \begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ A_1, & B_1, & C_1, & B_2, & C_2, & A_2, & C_3, & A_3, & B_3, & A_4, & B_4, & C_4, & B_5, & C_5, & A_5 \end{array} \right\rangle$$

If we match each team head-to-head, computation shows that team A beats team B (27-28), team B beats team C (26-29), and team C beats team A (27-28). In this example, we have a nontransitive result: team A beats team B and team B beats team C, but A fails to beat C. There is no team that wins all of its head-to-head matchups. Thus, matching the teams up in this way does not provide us with a scoring method at all, for it does not provide a linear ordering of the teams.

Though matching the teams up head-to-head fails as a scoring method, it suggests another condition for our scoring methods. Though not every invitational has a team that wins all of its head-to-head matchups, if there is such a team, it is reasonable to expect that team to win the invitational. We shall call this condition the *Condorcet Criterion*, and call a team that wins all its head-to-head matchups a *Condorcet team.*

The Condorcet Criterion: *A team that beats all the other teams in an invitational in head-to-head competition should be the winner of the invitational.*

Standard invitational scoring methods do not possess the Condorcet property in general. Creating examples that show violations is a relatively simple matter. Given a (m, ℓ) standard scoring method (with $m \geq 2$), we find a team A that consists primarily of strong runners but also possessing one or two weak runners. The strong runners allow A to win individual head-to-head matchups with other teams, but if we choose a sufficiently large set of teams $\mathcal{T}$ for our invitational, we can ensure that A's bottom runners will finish far enough down in the rankings that A fails to win the invitational. For example, we can show that the standard $(5, 2)$ scoring method violates the Condorcet criterion using an example involving four teams A, B, C, and D.

Example 2

$$\left\langle \begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ A_1, & A_2, & A_3, & B_1, & C_1, & D_1, & C_2, & D_2, & B_2, & D_3, & B_3, & C_3, & C_4, & B_4, \\ 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 \\ D_4, & B_5, & D_5, & B_6, & C_5, & B_7, & C_6, & C_7, & D_6, & D_7, & A_4, & A_5 & A_6 & A_7 \end{array} \right\rangle$$

In this race, runners from team A take the top three places, but the team is fully displaced, that is, the seventh runner from each of the other three teams finishes before A's fourth and fifth runner. In each head-to-head competition involving A then, A takes places 1, 2, 3, 11, and 12 giving it an aggregate score of 29. Each opposing team in the head-to-head with A places runners in positions 4 through 10, giving it a score of 30. Thus A wins each head-to-head matchup and is therefore a Condorcet team. However, when we score the race using the standard method, A has 57 points, B 54 points, and C and D 56. So A not only fails to win the invitational, but actually finishes last. Standard invitational scoring violates the Condorcet criterion, and it can do so in the worst possible way!

Mile 2: Fairness Criteria and Other Scoring Methods

The first mile of a cross country race typically goes quickly. Runners are fresh and the excitement of competition carries them along. In the second mile, runners need to assess their positions, while remembering to be patient as much of the race remains. In our consideration of scoring methods, we'll consider some other slightly more complex scoring methods and ways to measure their reasonableness. But we'll need to be patient.

Let us begin our second mile by considering a slightly different way of comparing two teams, call them P and Q, head-to-head. One way to measure an individual runner's strength is to count how many opponents he is beaten by, ignoring his own team's runners. A team's score could then be determined by aggregating this information among all the team's scoring runners. More formally, given that runner P_i is the ith scoring runner (of m) from team P, let $p_{i,Q}$ be the number of runners from team Q who finish ahead of runner P_i. Then $0 \le p_{i,Q} \le m + \ell$, where ℓ is the number of non-scoring displacing runners. A team score for P in a head-to-head matchup with team Q can be defined as $\sum_{i=1}^{k} p_{i,Q}$, with the lower overall score winning the meet. Let us call this method the (m, ℓ) *Runner Matchup Method*, since each individual runner on a team is compared with the opposing team.

If we apply the $(5, 2)$-Runner Matchup Method in a dual meet between teams A and B from Example 2, we see that $a_{1,B} = 0$, $a_{2,B} = 0$, $a_{3,B} = 0$, $a_{4,B} = 7$, and $a_{5,B} = 7$. Using the Runner Matchup Method, team A scores 14 points. A similar calculation shows that team B scores 15 points.

Is this really a new scoring system? An alert reader will notice that the team tallies in this example (14-15) differ by 15 points from the respective tallies (29-30) obtained using the standard head-to-head scoring. This is no coincidence. Let P and Q be two teams in a dual meet. Given runner P_i from team P ($0 \le i \le m$), let p_i denote P_i's overall finishing place. Since P_i's place in a dual meet with Q is determined by the runners finishing ahead of him (both on team P and on team Q), we have $p_i = i + p_{i,Q}$. Thus the (m, ℓ) standard team score for P in a dual meet with Q is

$$\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} (i + p_{i,Q}) = \frac{m(m+1)}{2} + \sum_{i=1}^{m} p_{i,Q}.$$

Thus $\sum_{i=1}^{m} p_i$ and $\sum_{i=1}^{m} p_{i,Q}$ differ by a term that depends only on m, the number of scoring runners, and not on the placements in any individual race. Thus the (m, ℓ) standard scoring method and the (m, ℓ) Runner Matchup Method are equivalent, yielding identical team rankings and identical differences between team scores in dual meets.

Though the Runner Matchup Methods are no different from the standard scoring methods, they allow us to identify a powerful relationship between a team's aggregate score in an invitational and its head-to-head scores against other teams in the race. To see it, suppose that in an invitational $|\mathcal{T}| = n$ and each team has m scoring runners. As in the dual meet, a runner's finishing place in an invitational is determined by runners finishing ahead of him, both on his and on opposing teams. Thus,

$$p_i = i + \sum_{\substack{Q \in \mathcal{T} \\ Q \neq P}} p_{i,Q}.$$

Then the invitational team score for team P using the (m, ℓ) standard scoring method is

$$\begin{aligned}
\sum_{i=1}^{m} p_i &= \sum_{i=1}^{m} \left(i + \sum_{\substack{Q \in \mathcal{T} \\ Q \neq P}} p_{i,Q} \right) \\
&= \sum_{i=1}^{m} i + \sum_{\substack{Q \in \mathcal{T} \\ Q \neq P}} \left(\sum_{i=1}^{m} p_{i,Q} \right) \\
&= \sum_{i=1}^{m} i + \sum_{\substack{Q \in \mathcal{T} \\ Q \neq P}} \left[(\text{Standard Team Score of } P \text{ vs. } Q) - \sum_{i=1}^{m} i \right] \\
&= (|\mathcal{T}| - 2) \sum_{i=1}^{m} i + \sum_{\substack{Q \in \mathcal{T} \\ Q \neq P}} (\text{Standard Team Score of } P \text{ vs. } Q)\,.
\end{aligned}$$

These show that teams' standard invitational scores can be obtained from the sum of their standard head-to-head scores. We need only subtract a constant based on the number of teams in the invitational. We thus have

THEOREM 1 *A team's score using standard invitational scoring can be determined from its head-to-head matchup scores. If there are n teams and m scoring runners per team (with or without displacement), then a team's invitational score is the sum of its head-to-head scores, minus $\frac{(n-2)m(m+1)}{2}$.*

EXAMPLE 3 Using the data from Example 2, there are $n = 4$ teams and $m = 5$ scoring runners. Team A scores 29 points in each of its head-to-head matchups with teams B, C, and D. The sum of these scores is 87; from this we subtract $\frac{(n-2)(m)(m+1)}{2} = \frac{(4-2)(5)(5+1)}{2} = 30$ to find A's invitational total of 57. The other team's invitational scores may be similarly verified.

Aside from describing a relationship between a team's head-to-head scores and its aggregate score in an invitational, Theorem 1 yields several corollaries, the first of which is immediate. Recall that a Condorcet team in an invitational beats every other team head-to-head.

COROLLARY 1 *Using* (m, ℓ) *standard scoring, the winning team in an invitational is the team whose head-to-head scores have minimal sum, or equivalantly, have minimal average.*

COROLLARY 2 *The* $(m, 0)$ *standard scoring method cannot rank a Condorcet team at the bottom.*

Proof. Without displacement, the total number of points available to both teams in every head-to-head matchup is the same: $\sum_{i=1}^{2m} i = m(2m+1)$. Call this number S. If team P wins all its head-to-head matchups, then P will score fewer than $\frac{S}{2}$ in each of its matchups, and so its overall average will be less than $\frac{S}{2}$ as well. The average score of all n teams across all matchups must be $\frac{S}{2}$, so there must be another team, call it Q, whose average is greater than S. Thus team Q will be ranked below P in the invitational rankings.

Corollary 2 is important because it demonstrates a problem with using displacements. In Example 2, team A was bottom ranked even though it won all of its head-to-head matchups. Corollary 2 tells us that this is a consequence of allowing displacing runners. Though the $(m, 0)$ standard scoring methods fail the Condorcet criterion in general, at least they cannot fail as badly as those (m, ℓ) methods where $\ell > 0$: A Condorcet team cannot be bottom ranked.

Mile 3: More Criteria and Alternative Scoring Methods

Since a Condorcet team A in an invitational beats every other team head-to-head, it is natural to ask if there are any methods that guarantee team A a first place finish. Consider the following scoring procedure: score the teams using a standard $(m, 0)$ scoring method. Once a preliminary ranking of the teams has been established, we drop the lowest ranked team. Since A cannot be ranked last in this method, A will not be eliminated. Now we rescore, with all of the runners from the bottom-ranked team removed. A remains the Condorcet team after this elimination step. We now repeat the process, scoring the invitational as usual and eliminating the team that is now bottom-ranked. We continue until there is only one team remaining, and declare that team to be the winner of the invitational. In this case, we know that the winning team must be A: at every stage, team A remains the Condorcet team, and by Corollary 2 cannot be bottom-ranked at any step in the process. Thus team A can never be eliminated. If we call this method the *Iterated* $(m, 0)$ *Scoring Method*, then we have shown the following result.

COROLLARY 3 *The iterated* $(m, 0)$ *scoring methods satisfy the Condorcet criterion.*

We can extend this to obtain a complete ranking of the teams in the invitational. Once the top-ranked team has been determined, we can eliminate that team and determine the iterated $(m, 0)$ scoring method winner from among the remaining teams, calling that the second place team. This ensures that a team that beats all other teams head-to-head except A is guaranteed to finish second, if such a team exists. Other rankings may be determined similarly.

EXAMPLE 4 Consider the following race:

$$\left\langle \begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ A_1, & A_2, & A_3, & B_1, & B_2, & B_3, & B_4, & B_5, & C_1, & C_2, & C_3, & C_4, & C_5, & A_4, & A_5 \end{array} \right\rangle$$

In head-to-head competition, A beats teams B and C by identical 25-30 scores and is therefore the Condorcet team. The team ranking using $(5, 0)$-standard scoring is $B \prec A \prec C$ by a score of 30-35-55. This shows that we may have a violation of the Condorcet criterion even if displacement is not allowed. However, if we use the iterated $(5, 0)$ scoring method, then team C is dropped at the end of the first round, and teams A and B face off in the second round with A winning. To find the second place team, we allow B and C to face off head-to-head (with A removed). Then B beats C 15-40, so the ranking using the iterated (5,0) scoring method is $A \prec B \prec C$.

We have identified a scoring method that satisfies the Condorcet criterion. The iterated methods are less intuitive than standard scoring (can you imagine coaches trying to predict their teams' places in the middle of a race?), but are significantly more stable. Barring practical concerns, is there any theoretical reason to avoid such a method? Unfortunately, the answer to that question is, "Yes."

Example 5 Consider the following race among three closely matched teams A, B, and C. Midway through the race, the placings of the runners are as follows:

$$\left\langle \begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ C_1, & C_2, & A_1, & A_2, & A_3, & B_1, & B_2, & B_3, & B_4, & B_5, & A_4, & C_3, & C_4, & C_5, & A_5 \end{array} \right\rangle$$

The quick-thinking coach from team A scores the race at the midway point, using the iterated (5,0) scoring method. When she does so, she finds that in the first round of scoring, the scores for A, B, and C are, respectively, 38, 40 and 42. In this preliminary scoring, Team C would be eliminated, and the race would be rescored with only teams A and B. In the second round, A would have runners in positions 1, 2, 3, 9, and 10, for a total score of 25, while B would hold positions 4 through 8 for a total score of 30. So at this point in the race, A is leading the race, but it is close! The coach from team A exhorts her team to push hard over the last half of the course, and one of them, runner A_4, responds. Over the last mile, she surges ahead of all the runners from team B, clinching her team's victory, right? Let's look at the final results.

The final runner placings in the race are

$$\left\langle \begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ C_1, & C_2, & A_1, & A_2, & A_3, & A_4, & B_1, & B_2, & B_3, & B_4, & B_5, & C_3, & C_4, & C_5, & A_5 \end{array} \right\rangle$$

Using the iterated (5,0) scoring method, Round 1 yields scores of 33, 45, and 42 for teams A, B, and C, respectively. Team B is eliminated and we rescore the race. In Round 2, A holds positions 3, 4, 5, 6, and 10, while C holds places 1, 2, 7, 8, and 9. In the final scoring, then, A has 28 points, while C has 27. Amazingly, C wins! Runner A_4's surge through the field of runners perversely hurt her team's cause.

Unfortunately, situations such as this, though relatively rare, can occur with the iterated scoring methods. We say that they violate the *Monotonicity* criterion.

Monotonicity Criterion: *If team P is the winner of a race, and in a second race with the same teams and runners, a runner from team P improves his performance by moving up one or more places and no other runners change position, then team P should remain the winning team.*

Our search for an improved scoring method is beginning to get frustrating! We have seen flaws in the usual system, but each new method possesses some undesirable properties of its own. Before we get too depressed, however, let us make note of some positives for our scoring systems.

First, the (m, ℓ) standard scoring methods satisfy the monotonicity criterion, for improving a winning team's individual placings can only lower that team's score in these point-based systems, and can only raise other teams' scores. Second, all the scoring methods we have described thus far possess a desirable property that we call the *Pareto* condition.

The Pareto Condition: *If a scoring method involves k runners from each of teams A and B, and if $a_i < b_i$ (that is, if runner A_i beats runner B_i) for all i ($1 \leq i \leq k$), then team B should not finish above team A in the rankings.*

The Pareto condition says that if all of my team's runners beat their individual counterparts from your team, then your team should not be ranked above my team by the scoring method.

Since better placings in a race yield lower team scores in the standard scoring methods, the (m, ℓ) standard scoring methods, and by extension, the iterated (m, ℓ) scoring methods satisfy the Pareto condition. Surprisingly, however, there are some seemingly reasonable methods that violate it. Consider the following method, another way to match up the teams head-to-head: create some fixed ordering of the teams in the invitational (perhaps randomly). After choosing a (m, ℓ) standard scoring method, match up the first two teams in the list head-to-head using that method. The winner of that matchup will advance to face the third team in the list in another head-to-head matchup. The winner of that will progress to face the fourth team in the list, and so on. The winner is the team that is left when the list has been exhausted.

Not surprisingly, this method satisfies the Condorcet criterion, since any team capable of winning each of its head-to-head matchups will survive to the end of the competition. On the other hand, this sequential-type scoring method violates the Pareto condition in general.

Example 6 Consider the following race with four teams (A, B, C, and D) with three runners per team:

$$\left\langle \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ C_1, & B_1, & A_1, & B_2, & A_2, & A_3, & D_1, & D_2, & D_3, & C_2, & C_3, & B_3 \end{array} \right\rangle$$

Let us score this meet using $(3,0)$ standard scoring using head-to-head matchups in the sequence $[A,B,C,D]$. We match up teams A and B (ignoring all the other teams) and find that team B wins, 10-11. Team B advances to face team C which beats B by a head-to-head score of 10-11. Team C then advances to face team D, and D wins that matchup by the score of 9-12. So D is the winner of the invitational using this method.

However, runner A_1 beats runner D_1, A_2 beats D_2, and A_3 beats D_3. This example therefore shows a violation of the Pareto condition. In fact, the offense is far worse than a simple Pareto violation. Using this process, Team D wins the invitational even though *all of team A's runners finish before all of team D's runners*. Similar violations of the Pareto condition using general (m,ℓ) sequential scoring of this type can easily be constructed for $m > 3$ and $\ell > 0$.

Mile 4: Some Social Choice Theory

The middle miles of a cross country race are often where the race is won. Runners need to keep their attention on their competition even as they become fatigued. Here, we focus on some of the formal language of voting theory. We will use the notation of [9]: [8] provides an elementary introduction to the subject.

We consider the situation in which a group of people (the voters) wishes to decide among several alternatives (most typically, these alternatives are candidates). Let A be the (finite) set of candidates or *alternatives*, and let V be the (finite) set of voters. We assume that each voter in V has an established preference between any two candidates $a, b \in A$, and that these preferences are transitive. Thus the voter has a linearly ordered *preference list* of candidates, ranking them from top to bottom.

Definition 3 *A* social welfare function *is a function that accepts as input a sequence of individual preference lists of some fixed set A and produces as output a single listing (perhaps with ties) of the set A. This list is called the* social preference list.

Social welfare functions provide the means of counting the ballots in an election, and different social welfare functions yield different election results in general. For example, the well-known *Plurality Method* is the social welfare function that ranks the candidates in order of the number of times they appear at the top of voters' preference lists. The *Borda Count* awards points to the candidates as follows: if there are n candidates, the candidate at the bottom of a list gets zero points, the alternative at the next to the bottom spot gets one point, then next one up gets two points, and so on up to the top candidate on the list who gets $n-1$ points. For each candidate, we add up the total number of points awarded over all the preference lists, and rank the candidates in order of total points, from most to least.

A connection between social welfare functions and cross country scoring methods is immediately apparent. The teams in an invitational can be thought of as the alternatives in an election, with the runners serving as the voters: the number of voters corresponds to the number of runners per team. Finding the winner of an invitational is then analogous to selecting the winner of an election, and cross country scoring methods correspond to social welfare functions. As a simple example, consider a two voter, three candidate election, where voter 1's preferences are candidates B, A, C in that order, while voter 2's preferences are candidates A, B, C. By analogy, we picture this as corresponding to a three-team race with two runners per team: $< B_1, A_1, C_1, A_2, B_2, C_2 >$. We have no need here to make the analogy more precise; rather, we ask the reader only to consider that there is a parallelism between the two domains. In some sense, we can think of the runners in a race as voting with their feet!

While cross country scoring and vote counting share some similarities, they are clearly not the identical. A notable distinction occurs when there are two candidates/teams. May proved (see [2]) that with two candidates, there is only one reasonable method of choosing the winner of an election: majority rule, i.e., the winner of a 2-candidate election should be the candidate who receives more first place votes. However, as we noted in our UW-Oshkosh vs. UW-LaCrosse head-to-head example when we considered different values of m and ℓ in the (m, ℓ) standard scoring method, there are many reasonable ways to score a two-team race. Moreover, even when the number of scoring runners remains fixed, we can see different results.

Example 7 When UW-Oshkosh and UW-LaCrosse are considered head-to-head in the Jim Drews example, Oshkosh runners finish in positions 2, 3, 5, 8, and 9, while LaCrosse runners finish at 1, 4, 6, 7, 10. Rather than finding the team scores using the sum of the placings (in which case Oshkosh wins), we could find the sum of the cube roots of the runners' placings, again giving victory to the team with the smaller score (This effectively places less importance on earlier finishers). LaCrosse wins the head-to-head competition by an approximate score of 8.47 to 8.49.

In this sense, then, there are more cross country scoring systems than there are social welfare functions.

Just as there are criteria measuring the reasonableness of cross country scoring methods, there are means of assessing the value or fairness of different social welfare functions. In fact, our cross country scoring criteria have been drawn directly from analogous criteria in voting theory. For example, the *Pareto Condition* (social choice version) asserts that if every voter prefers candidate x over candidate y on his or her ballot, then y should not win the election. Similarly, a social welfare function is said to satisfy the *Binary Independence Condition* (social choice version) if the following holds: Given our fixed sets A and V, but two different sequences of individual preference lists. Suppose that exactly the same people have candidate x over candidate y in their list. Then we either have x over y in both output rankings of the candidates, or y over x in both output rankings, or y and x tied in both output rankings. That is, the positioning of candidates other than x and y in the individual preference lists is irrelevant to the question of whether x is preferred to y or not in the final output list.

Just as our standard cross country scoring methods' violations of binary independence can have consequences for invitationals, violations of binary independence in the social choice domain can have profound effects on elections. For example, had Ralph Nader not been a U.S. presidential candidate in the 2000 election, then it is likely that Al Gore would have beaten George W. Bush ([7]). The U.S. presidential electoral system violates binary independence in precisely the way that cross-country scoring does.

While fairness criteria and the search for an optimal social welfare function have been the subject of research for well over half a century, Kenneth Arrow proved ([1]) that this exploration is inherently limited.

Arrow's Impossibility Theorem (1950)[1] *If A has at least three elements and the set V of individuals is finite, then the only social welfare function for A and V satisfying the Pareto and Binary Independence conditions is one for which there is a single voter $v \in V$ such that for every choice of individual preference lists by the voters, the social preference list is precisely the same as the individual preference list of v.*

[1]Arrow's Impossibility Theorem earned him the Nobel prize in Economics in 1972.

Arrow's Theorem says, simply, that the only social welfare functions satisfying both the Pareto and Binary Independence conditions are dictatorships (with v serving as the dictator).

Mile 5: Impossibility?

The disheartening implication of Arrow's Theorem is that there is no democratic voting method that satisfies both the Pareto and Binary Independence criteria. When we return to the analogy between cross country scoring and vote counting, we naturally wonder whether Arrow's Theorem yields an analogous result in running. We take up this question in the last mile of our race.

In pursuing alternatives to our standard scoring methods, we have encountered numerous roadblocks. Each of our methods suffered from some notable flaw, whether it be a violation of the Condorcet, Monotonicity, or Pareto criteria. Moreover, we have not yet found a cross country scoring method that does satisfy our condition of Binary Independence, the issue that motivated our initial investigation. Before we show how difficult this condition is to satisfy, let us look at an example.

Example 8 Suppose that teams A and B are the only two teams in a race, and we allow only $k = 2$ runners per team. There are ${}_4C_2 = 6$ possible different races:

1. $\langle A_1, A_2, B_1, B_2 \rangle$,	and symmetrically	4. $\langle B_1, B_2, A_1, A_2 \rangle$
2. $\langle A_1, B_1, A_2, B_2 \rangle$,		5. $\langle B_1, A_1, B_2, A_2 \rangle$
3. $\langle A_1, B_1, B_2, A_2 \rangle$,		6. $\langle B_1, A_1, A_2, B_2 \rangle$.

Without knowing precisely what our scoring method is, we can say a number of things about it. For example, if we insist that the Pareto condition be satisfied, team A necessarily must win Races 1 and 2, while team B wins Races 4 and 5. Only Race 3 and Race 6 offer any possible ambiguity. Suppose that our scoring method allows A to win the third race. Then, by symmetry, B wins Race 6. Recall our definition of *non-team oriented methods* from Definition 2. In this case our scoring method is indeed non-team oriented: in each race, the team winner corresponds to the team whose first runner finishes first. On the other hand, suppose that we decide instead that team A should win race 3 (and, symmetrically, B wins Race 6). This again yields a non-team oriented scoring system, in which the winning team is the team whose *second* runner finishes first, i.e., $A \prec B \Leftrightarrow a_2 < b_2$.

Example 8 shows that when there are two teams and only two runners per team, there are *only* two scoring methods that satisfy the Pareto condition, and both of these are non-team oriented. Though this is disturbing, it is but a single, small example. What if there are more than two teams?

EXAMPLE 9 Suppose that as in Example 8 there are only $k = 2$ runners per team, but now there are three teams in the race. Then there are a total of $({}_6C_2) \cdot ({}_4C_2) = 90$ different races. Consider the race

$$\langle A_1, B_1, B_2, A_2, C_1, C_2 \rangle . \tag{1}$$

If we insist that our scoring method satisfy the Pareto condition, then team C should finish last. However, there is flexibility in the relative positions of teams A and B. Suppose that our scoring method chooses A to be the winning team in the race. If our method satisfies the Binary Independence condition (as well as symmetry), then the method is completely determined. For example, consider the race $\langle B_1, A_1, C_1, C_2, B_2, A_2 \rangle$. The Pareto condition ensures that $B \prec A$, while Binary Independence ensures that $B \prec C$, since ignoring team A gives the race $\langle B_1, C_1, C_2, B_2 \rangle$, which B wins by symmetry. Similarly, $A \prec C$, so $B \prec A \prec C$. The team ranking corresponds precisely to the finishing order of the teams' first runners. This is no coincidence: the team rankings from each of the 90 races is determined solely by the first runner placings. So our choice to make A the winner in (1) gave us a non-team oriented method depending only on a team's first runners. Similarly, if we had chosen B to be the winner in (1), we would have found our method to be non-team oriented, with team rankings dependent only on the teams' second runners.

Examples 8 and 9 highlight the challenge of satisfying both the Pareto and Binary Independence conditions. Unfortunately for seekers of fair scoring methods, Example 9 generalizes to an arbitrary number of scoring runners. This is our main result.

THEOREM 2 **Arrow's Theorem for Cross Country Scoring:** *If there are at least three teams involved in an invitational, then the only scoring methods that satisfy Binary Independence and the Pareto condition are the non-team oriented methods.*

Figure 1: The traditional symbol for cross country running

That is, the only way to score a race consistently that will ensure that the Pareto and Binary Independence conditions are satisfied is to pick an index i ($1 \leq i \leq k$), and rank the teams according to the placings of their ith runners. Thus, the social choice version of Arrow's Theorem yields an analogous version in the cross country running domain, with the collection of ith runners for each team playing the role of the dictator in an election. Our discussion in the previous section highlighted some distinctions between elections and cross country races, so the classes of social welfare functions and cross country scoring methods are different sets. Thus Theorem 2 is not merely a restatement of of Arrow's Theorem in a new domain; it is in fact a slightly different result.

The proof of Theorem 2, like the proof of the social choice version of Arrow's Theorem, is not conceptually difficult, but it is technical, so it is omitted here. The result follows using a modification of Taylor's proof of Arrow's Theorem (social choice version), found in §10.4 of [9].

An interesting coincidence is that the traditional symbol of cross country running is the two letters "CC" cut by an arrow (see Figure 1). Perhaps the symbol designers knew something about scoring methods! And with that final push, we cross the finish line.

Warmdown: Some Concluding Remarks

The cross country version of Arrow's Theorem is disheartening. There is no reasonable scoring method that satisfies both the Pareto condition and Binary Independence. All is not lost, however. Here is one very simple way of finding the winner of an invitational: rank the teams in order of *total finish time* or equivalently, average finish time of the scoring runners. This is the system that is used to rank the teams in the International Association of Ultrarunners 100 Kilometer World Challenge events (see §10.24 of [5]). This system satisfies all of our stated reasonableness criteria. However, there is a tradeoff: according to our definition, it is not a scoring method. Scoring methods accept as input a ranking of runners and do not take into account finishing times.

This distinction is as much philosophical as technical. Since this *Total Time Method* involves only the runners' times and not their placings, Binary Independence is satisfied. UW Oshkosh's relative placing against UW LaCrosse does not rely at all upon the presence of UW-Madison runners. However, this method also removes one exciting aspect of cross country running: the head-to-head individual competition. The total time method involves a race against a clock rather than against opponents. So though it might be a mathematically superior means of scoring a race, it is perhaps less desirable.

After competing strongly all season and climbing to second in the final NCAA Division III cross country poll, the UW-Oshkosh men's cross country team finished sixth at the 2001 NCAA Championships in Rock Island, Illinois. The winner? UW-LaCrosse.

Acknowledgments: The author wishes to thank Joe Gallian for his helpful comments in the preparation of this manuscript and Underwood Dudley and John Beam for their careful editing.

References

[1] Kenneth Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58:328–346, 1950.

[2] Kenneth May. A set of independent, necessary and sufficient conditions for simple majority decision. *Econometrica*, 20:680–684, 1952.

[3] The National Collegiate Athletic Association (NCAA). *2009/2010 NCAA men's and women's track and field and cross country rules*, 2009 (accessed August 27, 2009). Available at `http://www.ncaapublications.com/Uploads/PDF/mw_trackandfield_rulese88cb2a4-3dac-4a6d-b6c2-c0fe5a66ffe7.pdf`.

[4] The International Association of Athletics Federations (IAAF). 37th IAAF world cross country championships. 2009 (accessed August 27, 2009). Available at `http://www.iaaf.org/WXC09/index.html`.

[5] International Association of Ultrarunners. *Technical guidelines for the organisation of a major IAU competition*, 2009 (accessed August 27, 2009). Available at `http://www.iau.org.tw/upload/download/1249000956.doc`.

[6] Kirk Reynolds. Men's history and statistics. 2009 (accessed August 27, 2009). Available at `http://www.physical-education.pomona.edu/womens/crosscountry/stats/XChistorian/ncaaxc.shtml`.

[7] David E. Rosenbaum. The 2004 Campaign: the Independent - Relax, Nader advises alarmed Democrats, but the 2000 math counsels otherwise. *The New York Times*, February 24, 2004.

[8] Peter Tannenbaum. *Excursions in Modern Mathematics*. Pearson Education, Inc., Upper Saddle River, NJ, 6th edition, 2007.

[9] Alan D. Taylor. *Mathematics and Politics: Strategy, Voting, Power and Proof*. Springer-Verlag, New York, 1995.

STEVE SZYDLIK earned an M.A. in 1991 and a PhD. in mathematics in 1997 from the University of Wisconsin, and he has taught at the University of Wisconsin-Oshkosh since 1996. As an undergraduate at Union College, he ran both cross-country and track, and he remains a strong supporter of Division III athletics. His interest in cross country scoring systems arose when he was simultaneously teaching a course on the mathematics of voting and following the UW-Oshkosh cross country season. Mathematics Department, University of Wisconsin-Oshkosh, Oshkosh, WI 54901. szydliks@uwosh.edu.