Rank injustice?: How the scoring method for cross-country running competitions violates major social choice principles

Article *in* Public Choice · February 2007

DOI: 10.1007/s11127-007-9193-6 · Source: RePEc

1 author:

Thomas H. Hammond
Michigan State University

Public Choice (2007) 133: 359–375
DOI 10.1007/s11127-007-9193-6

Rank injustice?: How the scoring method for cross-country running competitions violates major social choice principles

Thomas H. Hammond

Received: 24 May 2006 / Accepted: 3 June 2007 / Published online: 18 July 2007

Abstract The scoring method used by high schools and colleges in the U.S. to determine which team wins a cross-country meet can violate a major social choice principle, referred to here as Independence from Irrelevant Teams: whether team A is scored as defeating or losing to team B can depend on whether team C's performance is included in the calculations. In addition, if a three-way meet is scored as three dual meets, the scoring method can produce a cycle, thereby violating the principle of Transitivity: team A beats team B, team B beats team C, but team C beats team A. Real-world violations of Independence and Transitivity are reported from a high school cross-country meet held in Michigan in the U.S. in 2003. Several results are presented about the conditions under which these two principles can be violated. An alternative scoring method that will violate neither Independence nor Transitivity is also discussed and evaluated.

Keywords Social choice theory · Sports · Cross-country meets · Transitivity · Independence from irrelevant teams

1 Introduction

It has long been recognized that our methods for making social and political choices may violate desirable principles such as transitivity and the Pareto Principle. These problems have been extensively explored in the social choice literature (see, e.g., Arrow 1963; Sen 1970; Schwartz 1986; Brams and Fishburn 2002), and sprinkled throughout this literature are descriptions of real-world instances in which some method for making social and political choices can be shown to have violated one or more of these principles.

The scoring methods for athletic events have occasionally been used in the social choice literature to illuminate these social choice problems. For example, MacKay (1980, Chaps. 2, 4) shows how the problem of aggregating athletes' performances in the ten events

T.H. Hammond (✉)
Michigan State University, East Lansing, MI 48824, USA
e-mail: thammond@msu.edu

in the Olympic decathlon has some similarities to the problem of aggregating voters' preferences in an election. Saari (2001, pp. 40–41) shows how the scoring methods used in international figure skating competitions have produced some perverse and unexpected results that can be explained by social choice arguments. Saari (2001, p. 55) also suggests that the scoring method for a multi-team ($\geq$3), multi-event track meet can, at least in principle, exhibit a social choice problem: if there are three teams in a track meet, one team that beats another team in a "combined" (three-way) meet may not beat that team when that same track meet is scored as a trio of dual meets.

This paper demonstrates that the scoring method used in the U.S. in high school and college cross-country meets—long-distance footraces that do not take place on a standard oval track—also has some undesirable social choice properties. A team's score in a cross-country meet is based on the sum of the positions in which some k of the team's runners finished the race. For example, in a cross-country meet with two teams (a dual meet), it is usually the case that $k = 5$, so a team's score is based on the sum of the positions in which the team's fastest five runners finished the race. (A cross-country team usually has more than five runners in a meet, but when $k = 5$, only the five fastest runners from each team count in the scoring.) If a team's fastest five runners finished 2nd, 4th, 5th, 6th, and 8th, the team's score would thus be $2+4+5+6+8=25$. Since this is a dual meet, the other team's fastest five runners would have finished 1st, 3rd, 7th, 9th, and 10th, which means that this team's score would be $1+3+7+9+10=30$. In a cross-country meet, the team with the lowest score wins, so the first team here would be counted as defeating the second team by a score of 25 to 30.

On occasion, three or more teams will run a cross-country race on the same course at the same time but the meet will be scored as three or more dual meets. For example, if teams A, B, and C are in the meet, the meet could be scored as a single three-way meet involving A vs. B vs. C; the order-of-finish positions of the fastest five runners from each of the three teams would be counted, for a total of 15 runners. However, such a meet is sometimes scored as three dual meets, involving teams A vs. B, B vs. C, and A vs. C. In each of these dual meets, the order-of-finish positions of the fastest five runners from each of the two teams would be counted, for a total of 10 runners. The scoring would be done as in the dual-meet example already described; it is as if the runners from the third team were not present at all.

The scoring method for cross-country meets is generally used for three somewhat different purposes. First, we usually want to find an *overall winner* of the meet: of all the teams running, which is the "best"? Second, we may also want a *ranking* of how the teams finished: we are often interested in the top-3 finishers, for example, not just the first; it is most desirable, of course, to finish first, but it is often considered noteworthy to have finished second or third (think of the awarding of the silver and bronze medals in the Olympics, in addition to the gold). Third, we sometimes also want to be able to *compare teams by pairs*, as when a multi-team ($\geq$3) meet is scored as several dual meets, so that we may know whether team A is better than, as good as, or worse than team B, no matter how they fared against other teams.

Unfortunately, the scoring method for cross-country meets exhibits two problems in serving these three purposes. The problems involve some *inconsistency* and *ambiguity* in the results that this scoring method can produce. The first problem involves what is, in effect, a violation of the principle known as the Weak Axiom of Revealed Preference (WARP). Plott (1976, p. 550) provides a simple definition of WARP, involving a set of options and a larger set of options containing the first set:

(i) If two options are tied for a "win" in a contest within a small set, then either they tie again over the larger set or neither is a "winner."

(ii) If an option was a winner from a big set, then it should also be a winner in any smaller subset.

For athletic contests, we might call this the principle of *independence from irrelevant teams*, and in a cross-country meet, it would mean that whether team A is scored as defeating, tying, or losing to team B should not be affected by whether team C's results are included in the scoring. The rationale is obvious: if we want to know if team A is better than, just as good as, or worse than team B, there is no reason why this comparison should be influenced by how well, or how badly, team C performed. After all, team C's performance should be irrelevant to a comparison of teams A and B, and this is especially true in a sport like cross-country in which the runners are not supposed to interact physically with each other. However, it turns out that the scoring method for cross-country meets violates this independence principle: as will be demonstrated, whether team A is scored as defeating, tying, or losing to team B can be affected by whether team C's results are included in the calculations. Thus, this scoring method can produce inconsistent results about what the "best" team actually is.[1]

The second problem involves a violation of the principle of *transitivity*. In a three-way cross-country meet, for example, if team A beats team B, and team B beats team C, respect for transitivity means that team A should be expected to beat team C. For a cross-country meet in which the athletes are all running on the same course at the same time, it certainly seems reasonable to want the scoring method to respect transitivity: what value is a scoring method for an athletic contest if it can produce ambiguous results about how to rank the teams in that meet? Nonetheless, it turns out that the cross-country scoring method violates transitivity: even if all the runners from teams A, B, and C are running on the same course at the same time, it will be demonstrated that in the three dual meets, team A can be scored as defeating team B, team B can be scored as defeating team C, but team C can be scored as defeating team A. Because transitivity is violated here, there are ambiguous indications about how to rank the teams.

Of course, intransitivity is often observed *across* athletic contests: in league competition, for example, we often observe that team A will beat team B, team B will beat team C in a subsequent contest, but team C will beat team A in yet another contest. The usual explanations are that a team may be "up" for one game but "down" or "overconfident" for a subsequent game, or some key player may be healthy in one game but injured and unable to play in a subsequent game, or one game is played "at home" but a later game is played "on the road." Indeed, one might argue that it is precisely this possibility of intransitivity across athletic contests, and thus the uncertainty about possible outcomes, that makes them interesting to contestants and observers alike. Hence, we should not be disturbed by intransitivity in a series of athletic competitions; indeed, we should delight in it. But for a *particular* athletic event, such as a multi-team cross-country meet, the case for respecting transitivity seems much stronger: after all, one purpose of the meet is to rank the teams, from best to worst, at least at that particular site on that particular day, and when transitivity is violated, there are ambiguous indications about what the best team is and how the remaining teams should be ranked overall.

---

[1] It is interesting to note that in a brief article on methods for comparing the performances of academic teams, Huntington (1938) demonstrated that "the rank method" failed to satisfy what he referred to as "the postulate of relevancy," which is that "it ought not to be necessary to take into account the performance of any contestants who do not belong to these teams." He then showed "that the rank method fails to satisfy this postulate." He even observed that "The rank method is regularly used in scoring inter-collegiate cross-country runs." His arguments thus accurately foreshadowed some of those made here. Arrow (1963, footnote 3, p. 27) also cited Huntington's paper.

These are not just abstract problems: in this paper I report a case in which both independence from irrelevant teams and transitivity actually were violated in a high school girls cross-country meet. I then characterize the conditions under which violations of independence from irrelevant teams and transitivity can be expected to occur in cross-country meets. I next evaluate an alternative scoring method whose results would violate neither independence nor transitivity. I end the paper by suggesting that understanding scoring methods in sports is not just an academic exercise: both amateur and professional sports have substantial economic importance in modern societies, and developing a social choice theory of sporting events may lead to changes in the rules for these events, and thus to changes in who is counted as winning or losing. This may in turn have significant economic consequences for these teams, their leagues, their owners, and even for the countries that support teams in international competitions.

2 Social choice problems in the scoring method for cross-country meets

To illustrate how independence from irrelevant teams and transitivity can be violated by the scoring method for cross-country meets, consider the following example involving teams A, B, and C. Let m be the number of teams in a meet, so $m = 3$ here. For simplicity, assume that just the orders of finish for the three fastest runners from each team are taken into account in the scoring, so $k = 3$. Since these nine runners all run the race simultaneously, they can be ranked from 1st to 9th in order of finish. Label team A's three runners as A_1, A_2, and A_3, team B's three runners as B_1, B_2, and B_3, and team C's three runners as C_1, C_2, and C_3.

Now assume that the overall order of finish of the nine runners from the three teams is as follows in the three-way meet:

$$\begin{array}{llllllllll} \textit{Order}: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \textit{Runner}: & C_1 & A_1 & A_2 & B_1 & B_2 & B_3 & C_2 & C_3 & A_3. \end{array}$$

In the three-way meet, team A's sum is $2 + 3 + 9 = 14$, team B's sum is $4 + 5 + 6 = 15$, and team C's sum is $1 + 7 + 8 = 16$. Hence, team A is 1st with 14 points, team B is 2nd with 15 points, and team C is 3rd with 16 points; that is, team A beats teams B and C, and team B beats team C.

However, this method of scoring violates independence from irrelevant teams: how two teams fare against each other can be affected by whether a third team is included in the scoring. Consider the positions of the runners for the pairs of teams in the three dual meets. The contest between A and B has:

$$\begin{array}{lllllll} \textit{Order}: & 1 & 2 & 3 & 4 & 5 & 6 \\ \textit{Runner}: & A_1 & A_2 & B_1 & B_2 & B_3 & A_3. \end{array}$$

Team A's score here is $1 + 2 + 6 = 9$, and team B's score is $3 + 4 + 5 = 12$, so in this dual meet team A beats B by 9 to 12 points; independence is not violated here since team A beats team B in the three-way meet as well. Next, we see that team B beats C in the dual meet:

$$\begin{array}{lllllll} \textit{Order}: & 1 & 2 & 3 & 4 & 5 & 6 \\ \textit{Runner}: & C_1 & B_1 & B_2 & B_3 & C_2 & C_3. \end{array}$$

Team B's score here is $2 + 3 + 4 = 9$, and team C's score is $1 + 5 + 6 = 12$, so in this dual meet team B beats team C; again, independence is not violated here since team B beats team

C in the three-way meet as well. However, note that team C beats A in the dual meet:

Order:	1	2	3	4	5	6
Runner:	C_1	A_1	A_2	C_2	C_3	A_3.

Team C's score here is $1 + 4 + 5 = 10$, but team A's score is $2 + 3 + 6 = 11$: whereas team A beats team C in the three-way meet (14 to 16 points, respectively), team C beats team A in the dual meet here; in other words, independence is violated.

Note also that the outcomes from the three dual meets produce a cycle: in the dual meets, team A beats team B (by 9 to 12), team B beats team C (also by 9 to 12), but team C beats team A (by 10 to 11). In other words, transitivity is violated here as well.

3 Intransitive outcomes from a high school cross-country meet in Michigan in 2003

These violations of independence from irrelevant teams and transitivity are not just hypothetical concerns but actually occurred in a cross-country meet held in the state of Michigan on September 30, 2003, involving the girls cross-country teams from East Lansing High School, Okemos High School, and Holt High School in the U.S.[2] Each team had five runners among the fastest 15 finishers (out of 72 total runners). Counting the orders of finish of these fastest 15 runners as if they were in a single three-way meet (i.e., $m = 3$ and $k = 5$), Okemos, Holt, and East Lansing would have placed 1st, 2nd, and 3rd respectively, with scores of 39, 40, and 41; see the bottom of the "Three-way meet" column in Table 1.

However, note that independence from irrelevant teams was violated: in the three-way meet Okemos would have beaten Holt (by a score of 39 to 40) but in their dual meet, Holt beat Okemos (by 27 to 28). Similarly, in the three-way meet Holt would have beaten East Lansing (by 40 to 41) but in their dual meet, East Lansing beat Holt (by 27 to 28). Only the rankings of Okemos and East Lansing were unaffected by Holt: Okemos beat East Lansing both in their the three-way meet (by 39 to 41) and in their dual meet (by 26 to 29).

The dual meet results here also produced a cycle. In a three-way meet, Okemos would have beaten Holt and East Lansing, and Holt would have beaten East Lansing. Yet for the league competition—the teams are in the "Capital Area Activities League"—this meet was scored as three dual meets, and for dual meets Table 1 shows that East Lansing beat Holt by 27 to 28, Holt beat Okemos by 27 to 28, but Okemos beat East Lansing by 26 to 29. Thus, transitivity was violated.

4 Under what conditions can independence and transitivity be violated?

This cross-country meet demonstrates that the principles of independence from irrelevant teams and transitivity can be violated when there are three teams ($m = 3$) and five runners from each team count in the scoring ($k = 5$). We now determine the general conditions for when independence and transitivity can and cannot be violated. As will become apparent, there is an intimate relationship between independence and transitivity for this method of scoring cross-country meets: every order-of-finish that violates independence also violates

[2]The city of East Lansing in Michigan is the home of Michigan State University, Okemos is a community located immediately to the East of East Lansing, and Holt is located just to the South of East Lansing and Michigan State. The meet was held at Kinawa Middle School in Okemos.

Table 1 Results of the three-way and two-way cross-country meets among Okemos High School, Holt High School, and East Lansing High School on 30 September 2003

Runner	High School	Time	Three-way meet	Three dual meets					
			Rank	Ranks for E. Lansing vs. Holt		Ranks for Holt vs. Okemos		Ranks for Okemos vs. E. Lansing	
Darling	Holt	19:27	1		1	1			
De Kroub	Okemos	20:00	2				2	1	
Fessenden	Holt	20:02	3		2	3			
Pugh	Okemos	20:02[a]	4				4	2	
Bialek, D.	E. Lansing	20:05	5	3					3
Bialek, L.	E. Lansing	20:29	6	4					4
Butler	Okemos	20:37	7				5	5	
Roberts	E. Lansing	20:43	8	5					6
Crippen	E. Lansing	20:50	9	6					7
Dell	Holt	20:50[b]	10		7	6			
Douches	Okemos	21:04	11				7	8	
Brewer	Holt	21:10	12		8	8			
Coon	E. Lansing	21:12	13	9					9
Porter	Holt	21:14	14		10	9			
Lovett	Okemos	21:17	15				10	10	
Sum of ranks			Okemos: 39 Holt: 40 E. Lansing: 41	27 E. Lansing	Holt 28	27 Holt	Okemos 28	26 Okemos	E. Lansing 29

[a]Fessenden and Pugh were recorded as having identical times, but Fessenden actually finished slightly ahead of Pugh

[b]Crippen and Dell were recorded as having identical times, but Crippen actually finished slightly ahead of Dell

Table 2 The relationship between the number of teams (m), the number of runners who count in the order-of-finish (k), and violations of independence and transitivity

Number of teams, m	Number of runners per team who count in the order-of-finish, k					
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = \cdots$
$m = 1$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated
$m = 2$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated
$m = 3$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*
$m = 4$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*
$m = 5$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*
$m = \cdots$	Independence & transitivity cannot be violated	Independence & transitivity cannot be violated	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*	*Independence & transitivity can be violated*

transitivity, and vice versa. Consequently, proofs of the following propositions about independence and transitivity use essentially the same procedures.[3]

To organize the analysis consider Table 2; the values of m are shown in the rows and the values of k are shown in the columns. While the table contains cells for just the values of $m = 1, \ldots, 5$ and $k = 1, \ldots, 5$, we want to know whether independence and/or transitivity can be violated for any possible cell in an infinitely extended version of the table. For a starting point, the cross-country meet described in part 3 tells us that independence and transitivity both can be violated when $m = 3$ and $k = 5$.

The first proposition specifies when independence and transitivity cannot be violated:

[3]Respect for WARP (my independence from irrelevant teams) implies respect for transitivity. But it also appears that for this method of scoring, a violation of transitivity implies a violation of WARP, and a violation of WARP implies a violation of transitivity.

Proposition 1 *In a cross-country meet with the scoring method based on the order-of-finish positions of the runners*:

(a) *Independence from irrelevant teams cannot be violated if* $m < 3$.
(b) *Transitivity cannot be violated if* $m < 3$.
(c) *Independence from irrelevant teams cannot be violated if* $k < 3$.
(d) *Transitivity cannot be violated if* $k < 3$.

Proof Part (a) is true by definition: the independence condition refers to the impact of some team C on the outcomes of the race between teams A and B; hence it can be violated only if there are at least three teams. For this reason Table 2 states that the independence condition cannot be violated in the $m = 1$ and $m = 2$ rows.

Part (b) is true by definition as well: no cycle can occur with just one or two teams. Hence, Table 2 states that the transitivity condition cannot be violated for $m = 1$ and $m = 2$.

The proofs for (c) and (d) here are essentially the same and will be presented together. First, consider the case in which $k = 1$. It is obvious that if the fastest runner from team A finishes ahead of the fastest runner from team B, and the fastest runner from team B finishes ahead of the fastest runner from team C, whether team C's runner is included in the calculation cannot affect the result that runner A defeats runner B. Hence, independence cannot be violated when $k = 1$; it is so stated for all cells in the $k = 1$ column in Table 2. Furthermore, no cycle is possible when the order-of-finish positions for just one runner per team ($k = 1$) count in the scoring: if the fastest runner from team A finishes ahead of the fastest runner from team B, and the fastest runner from team B finishes ahead of the fastest runner from team C, then there is no possibility that the fastest runner for team C can finish ahead of the fastest runner for team A; hence, there cannot be a cycle when $k = 1$. For all cells in the $k = 1$ column in Table 2 it is thus stated that the transitivity condition cannot be violated.

Next, consider the case in which $k = 2$. Since we know that independence and transitivity cannot be violated when $m = 1$ or $m = 2$, consider the cell in which $m = 3$ and $k = 2$. But the proof in Appendix 1 demonstrates that neither independence nor transitivity can be violated when $m = 3$ and $k = 2$: adding team C cannot change the relative rankings of the runners from teams A and B. Moreover, since every case with $k = 2$ and $m > 3$ can be subdivided into trios of teams—i.e., team i_1, team i_2, and team i_3—and the proof in Appendix 1 shows that if team i_1 beats team i_2, and team i_2 beats team i_3, then team i_3 cannot beat team i_1. Repeatedly applying this result to all possible trios of teams produces a chain of pair-wise relationships among the teams that protect independence and transitivity from violation. □

The second proposition specifies when independence and transitivity can be violated:

Proposition 2 *In a cross-country meet with the scoring method based on the order-of-finish positions of the runners, independence from irrelevant teams and transitivity can be violated for any combination of* m *and* k, *whenever* $m \geq 3$ *and* $k \geq 3$.

For proof see Appendix 2.

5 Discussion

In a cross-country meet, violation of independence from irrelevant teams means that whether team A is scored as beating or losing to team B can be affected by whether some team C also

Table 3 Results if team times are summed in the cross-country meet among Okemos High School, Holt High School, and East Lansing High School on 30 September 2003

Runner	Holt	Runner	Okemos	Runner	East Lansing
Darling	19:27	De Kroub	20:00	Bialek, D.	20:05
Fessenden	20:02	Pugh	20:02	Bialek, L.	20:29
Dell	20:50	Butler	20:37	Roberts	20:43
Brewer	21:10	Douches	21:04	Crippen	20:50
Porter	21:14	Lovett	21:17	Coon	21:12
Sum of times	102:43	Sum of times	103:00	Sum of times	103:19
Rank in meet	1st		2nd		3rd

is in the race. It might seem odd that this could happen in a sport where physical interactions between the athletes are forbidden, and it would seem even more odd since this violation of independence stems from the *same* performances of the athletes: each girl's performance in each of the three dual meets reported here—her actual times—was precisely the same; that is, she ran the course only once in that meet. Yet whether her team was scored as defeating or losing to another team was affected by how a third team's runners performed, even if the presence or absence of the runners from the third team did not change *any* of the behavior of the runners on the other two teams.

Similarly, transitivity can be violated in the results from a cross-country meet when a three-way meet is scored as three dual meets. Again, this would seem to be at odds with the intuitively sensible notion that we should be able to rank the teams from best to worst when these dual-meet results are all based on the same performances on the same course at the same time. Indeed, the existence of a cycle seems to be especially odd in a competition in which objective measurements of the runners' performances—their order-of-finish positions—are available.

Unlike many social choice situations, however, there is an obvious and workable solution to these problems with the scoring method for cross-country meets. Instead of summing the order-of-finish *positions* of each team's top k runners, summing the *times* of each team's top k runners would avoid both problems. For the cross-country meet reported here, each runner's time is available in Table 1, so the three-way results based on the sum of the times for each team's runners can be computed; these results are summarized in Table 3.

We see in Table 3 that the Holt team has a summed time of 102 minutes and 43 seconds, the Okemos team has a summed time of 103 minutes and no seconds, and the East Lansing team has a summed time of 103 minutes and 19 seconds. Thus, instead of Okemos winning the three-way meet, with Holt finishing second and East Lansing finishing third, as produced by the "sum of the positions" method and shown in Table 1, Holt would now win the three-way meet and Okemos would now finish second, though East Lansing would still finish third.

This "sum of the times" method would not permit cycles: if team A's total time is smaller than team B's total time, and team B's total time is smaller than team C's total time, it cannot be the case that team C's total time will be smaller than team A's total time: the aggregation of cardinal numbers such as "times in a race" renders a cycle impossible. Moreover, this "sum of the times" method cannot violate independence: team C's total time cannot affect whether team A's total time is greater than, equal to, or lesser than, team B's total time. Hence, the "sum of the times" method would avoid *both* of the social choice problems highlighted in this paper, and so should be seen from this perspective as superior to the "sum of the positions" method.[4]

Furthermore, the "sum of the times" scoring method might even seem to have some desirable *incentive* effects for the runners. Under the "sum of the positions" scoring method, if a runner currently in position j for her team during a race ($j \leq m \times k$) thinks that she has no prospect of improving her position by the end of the race, she will have little incentive to run as fast as possible (as long as she does not expect to be passed by someone from another team); within the constraints of maintaining one's position in a race, "loafing" during the race is not penalized by the traditional scoring method. In contrast, under the "sum of the times" scoring method *every* runner (or at least the runners on each team who expect themselves to be among that team's top k runners) would always have an incentive to run as fast as possible since it is her team's aggregate *time* that counts, not its aggregate rank. Hence, runners should be expected to work harder under a "sum of the times" scoring method.

Interestingly, however, the coaches involved in this particular cross-country meet in 2003 all found the costs of this proposed system to be greater than the benefits of eliminating the possibility of violating the independence and transitivity principles in these kinds of dual meets.[5] They cite a number of concerns. For example, coach John Quiring of Okemos High School argued that wise coaches do not want their runners to go "all-out" in most races prior to the district, regional, and statewide meets that determine the statewide champion at the end of the cross-country season. The reason is that runners need to pace themselves throughout the season so as to avoid injury and maintain a healthy mental attitude toward the sport: racing "flat-out" in meet after meet, as induced by the "sum of the times" scoring method, might pose a greater risk of injury and psychological "burnout" for the runners. What was above called "loafing" above is, from this point of view, a desirable phenomenon.

Coach Robert Brown of East Lansing High School agreed with these concerns, suggesting also that the "sum of the times" method "would unduly favor the team with a 'superstar' while penalizing a team that perhaps had four good runners and one very slow one."

Coach David Foy of Holt High School likewise agreed with these criticisms of the "sum of the times" method: in his view, this alternative method would downgrade the aspect of *team* competition in a cross-country meet and put more emphasis on the role of the fastest individual runner or runners on a team. Like coach Quiring, he was also concerned that the new method would force the fastest runner or runners on a team to "perform maximally every race," and he went on to argue that "Top athletes need recovery runs, that is opportunities to go through the motions in a relaxed, yet competitive state. The proposal would expect *all* runners to be stressed during every race." He was also concerned that with the new method "team strategy is lost" during the race itself: "There [would be] little or no way that coaches and runners can make adjustments in [a] race to improve their scoring. In the traditional scoring, counting runners allows racers to know where they stand in the scoring of the event."

Nonetheless, there are some counterarguments to be made in defense of the "sum of the times" method, aside from respect for transitivity and independence. First, as to the concern about putting too much physical and mental stress on the runners over the course

[4]Regarding his competition among academic teams, Huntington (1938, p. 288) made precisely the same observation, noting that what he called the "sum-of-the-grades method," unlike his "rank method," did not violate his "postulate of relevancy."

[5]Unless otherwise noted, the remarks reported in the remainder of this section come from e-mail communications to the author (from coaches Brown and Foy) and personal and telephone communications with the author (from coach Quiring) in October, November, and December in 2003.

of a season, it is not clear how much more stress the "sum of the times" method imposes than the "sum of the positions" method. After all, if a team wants to win a dual meet under the "sum of the positions" method, its five fastest runners will have to finish ahead of a number of the other team's five fastest runners, and accomplishing this goal may well take great effort. In particular, for one runner to overtake another runner (thereby changing their relative positions) in order to win a meet under the "sum of the positions" method may actually take more effort than for the runner to improve her time sufficiently so that her team might win the meet under the "sum of the times" method.

Second, if a coach wants to limit stress on his team's runners, it seems plausible to think that the coach could devise a strategy in which each runner's goal would be to finish ahead of a particular number of runners from the other team or teams; as long as these relative positions are maintained (as occurs under current rules), the team might beat the other team since its "sum of the times" result would necessarily be lower than the other team's "sum of the times" result. Moreover, the success of these individual strategies—"stay ahead of x runners from the other team"—could also be monitored by the coaches and runners during the course of any one meet.

However, another of the coaches' criticisms of the "sum of the times" method—that it de-emphasizes the role of overall *team* performance and puts greater weight on the performance of particularly fast runners (or particularly slow runners)—may be valid. For example, under the "sum of the positions" scoring method, a team with one superstar runner and four mediocre runners would have difficulty defeating a team with five good-though-not-great runners, whereas the team with the superstar might be able to defeat the other team if the "sum of the times" scoring method were used. Similarly, under the "sum of the times" scoring method, a team with four excellent runners and one really mediocre runner might lose to a team with five good runners, whereas the team with the mediocre runner might be able to defeat the other team if the "sum of the positions" method were used. In this regard, it is interesting to note that coach Foy of Holt, whose team would have finished first overall in a three-way meet and would have defeated both Okemos and East Lansing in the dual meets, had the "sum of the times" method been used, thought that a change to this method was not desirable; as he put it, "Even though my team would have benefited from such a change, I feel that an aspect of team is lost."

Finally, it is interesting to note that the coaches all felt that the likelihood of the social choice problems highlighted here—the violations of independence and transitivity—was so low that there was little reason for concern. For example, coach Foy of Holt observed after the September 30 meet that "I've never seen anything like that. In all my years of coaching that was amazing."[6] Coach Brown of East Lansing likewise observed that this meet was "the first time ever that I've heard [of] that happening, too. It's not often that the three teams involved are so close talent-wise. And the chances of this happening again have to be very small .... To change the scoring to avoid this odd occurrence from happening once in a blue moon doesn't make sense to me .... It is an interesting and weird event to occur."[7] And coach Quiring reported that, at a post-season statewide meeting of cross-country coaches in late 2003, no other coaches mentioned to him, in a group discussion of issues related to

[6]This quotation is from Kangas (2003), an article on this high school cross-country meet that appeared in *The Town Courier*, a community newspaper covering the East Lansing and Okemos communities. This article initially alerted me to the existence of a cross-country cycle.

[7]Coach Brown's focus on how evenly matched teams are critical to the existence of cycles is borne out by the examples that have been constructed for this paper (not all of which have been shown here): for each case where independence and transitivity are violated, the differences among the teams' scores are quite small.

rules and scoring methods, that they had ever heard of any other cyclical outcomes before. (A number of these coaches had become aware of the curious results in the Holt–Okemos–East Lansing meet.) Hence, all three coaches involved in the Okemos/Holt/East-Lansing meet felt that while these violations of independence and transitivity may be regrettable, a change to the "sum of the times" scoring method would solve problems that appeared to occur only very rarely.

It is not clear how much foundation there is to these arguments about the improbability of a cycle. For a preliminary estimate, 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 $m = 3$ and $k = 3$. There are 290 of these orders of finish, and it turns out that for only 15 of these 290 orders of finish (5.2%) does team C beat team A, thereby generating a cycle and a violation of independence.[8]

However, some caution may be warranted in asserting that these problems are so rare that they are insignificant. First, something that is likely to occur in only 5.2% of all cross-country meets is still occurring once in every 20 meets, and given the many dozens of meets in any state each year, the resulting number of independence and transitivity violations might be rather substantial. Moreover, as a theoretical matter, while this $m = 3$ and $k = 3$ case suggests that violations of independence and transitivity might occur only some 5.2% of the time, it is not clear what happens to their likelihood when $m \geq 3$ and $k = 5$. The greater degrees-of-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). And as an empirical matter, coach Quiring noted that these three-way ($m = 3$, $k = 5$) dual meets had been held for only the past five or so years; holding separate ($m = 2$, $k = 5$) dual meets was the norm beforehand. Hence it might have been too soon to draw firm conclusions about the rarity of cycles in these meets.

Furthermore, it may well be that the "sum of the positions" scores in the conference, regional, and statewide meets (when $m > 3$, $k = 5$) actually conceal more frequent violations of the independence and transitivity principles than these coaches realize. After all, even though one team may be scored, under the "sum of the positions" method, as defeating a second team in the overall meet with three or more teams, retabulating the results as a large number of dual meets might nonetheless show that the second team would have defeated the first *if precisely the same times had been run in a two-way meet between the two teams*. So it may be that these principles are, in fact, violated substantially more often than the participants realize: since the dual-meet re-tabulations that would reveal such problems are apparently rarely conducted in district, regional, and statewide meets (where $k = 5$ and $m =$ several dozen teams), the true incidence of the violation of these social choice principles is largely unknown.

6 Conclusion

The general enterprise of social choice theory is to list desirable sets of principles for social decision-making, and then to explore whether these principles are compatible with each other. And if there is one overriding lesson from social choice theory, it is that desirable

[8]The example in Sect. 2 is one of these 15 cycle-inducing orders of finish. Dropping the subscripts for the individual runners from each team, the 15 cycle-inducing orders of finish are: CAABBBCCA, CAABBCBCA, ABCCABBCA, CABABBCCA, CABABCBCA, ABBCCCAAB, ABBCCACAB, ABCBCCAAB, ABCBCACAB, CABBCAABC, BCCAAABBC, BCAABCCAB, BCACABABC, and BCCAABABC.

sets of principles are often internally incompatible. In this paper I have demonstrated that the traditional method for comparing the performances of cross-country teams—the "sum of the positions" rule—violates the principles of independence from irrelevant teams and transitivity.

While I have shown that there exists a simple alternative—the "sum of the times" rule—that does, in fact, respect both independence and transitivity, the history of social choice theory also suggests that we are likely to find additional desirable principles by which scoring methods for cross-country meets might be evaluated. For example, the coaches involved in this noteworthy 2003 cross-country meet advanced what might be taken as important additional principles, such as "don't subject your runners to unnecessary wear and tear prior to the end-of-season championship meets" and "the method of scoring should not be excessively sensitive to 'superstar' runners nor to especially mediocre runners." Respecting these other principles (or at least concerns) seemed paramount to the coaches, especially since their perceptions were that violations of the independence and transitivity principles are extremely rare. (Whether such violations are actually this rare is something that requires more theoretical and empirical analysis.)

Of course, one might argue that this analysis of scoring methods for cross-country meets, while perhaps intellectually interesting, is of little significance to public choice scholars. My response would be to point out that athletic competitions around the world have great social and economic significance. Billions of dollars are spent to host the Olympic Games, for example, and billions of dollars are spent on advertising for sporting events around the world. And who wins and who loses these athletic contests has significant economic consequences for the teams themselves, for the athletes and coaches on the teams, and for the persons owning the teams (for professional sports) or the institutions having jurisdiction over the teams (for college sports). If the scoring methods can affect who the winners and losers are, then these methods are themselves of significant social and economic importance. In fact, even the procedure by which the members of the International Olympic Committee (IOC) select a host city for the Olympics Games—a decision that triggers the expenditure of billions of dollars in the "winning" country—might itself be analyzed from a social choice viewpoint. In this procedure, each IOC member has one vote, and balloting occurs over several rounds, with the city with the fewest votes in each round being dropped from the next round and a new vote being held on the remaining candidate cities; the winner is the first city to gain a majority of the votes of IOC members. (In any round, a country's IOC members cannot vote if their country has a host city still in contention.)

To be sure, cross-country meets are of negligible social and economic consequence. But the scoring methods for cross-country meets clearly do raise some interesting conceptual issues that may be relevant to scoring methods for many other sports. So given the social and economic importance of sports in modern societies, perhaps it is time to develop a "social choice theory of athletic contests" so that ambiguity and inconsistency, and the resulting controversies, can be eliminated from these economically important sporting events, at least to the extent that such a social choice theory tells us that this is possible.

Acknowledgements I thank Steven Brams for his support and his suggestions about initial versions of the propositions. Brian Silver and Nathan Monroe provided helpful remarks on early versions of this paper. I also thank the three coaches whose runners were involved in this cross-country meet—John Quiring of Okemos High School, Robert Brown of East Lansing High School, and David Foy of Holt High School—for their useful responses and arguments. An anonymous referee pointed out errors in my original submission and suggested insightful ways of discussing this subject. And an editor suggested the "Rank injustice" phrase that I have adopted for the primary title.

Appendix 1 Proof that independence and transitivity cannot be violated when $k = 2$ and $m = 3$

Because the calculation of scores for teams A, B, and C does not depend on which *particular* runner for a team finished in what rank, it is not necessary in the proof to distinguish between the two runners who count for each team. Hence, the sequence AABB means that team A's runners finished 1st and 2nd and team B's runners finished 3rd and 4th.

Begin by noting that in a dual meet between teams A and B, there are only six possible orders-of-finish: AABB, ABAB, ABBA, BBAA, BABA, and BAAB. But team A can beat team B in only two of these six ways: the order-of-finish must be either AABB or ABAB. With AABB, team A gets 3 points while team B gets 7 points; with ABAB, team A gets 4 points while team B gets 6 points. The only other possible orders of finish to consider would be ABBA and BAAB, both of which yield a tie (each team gets 5 points), and the BBAA and BABA sequences, in which team B beats team A. Since the proof depends on constructing sequences like "team A beats team B and team B beats team C," we are interested only in those sequences by which team A can beat team B (AABB and ABAB) and team B can beat team C (BBCC and BCBC).

The proof involving transitivity consists of showing that for both of the ways that team A can beat team B in a dual meet, given both of the ways that team B can beat team C in a dual meet, there is no possible ordering of all three team's runners that would allow team C to beat team A in a dual meet. The proof involving independence consists of showing that, given both of the ways that team A can beat team B in a dual meet and both of the ways that team B can beat team C in a dual meet (which collectively imply that team A always beats team C as well), there is no possible ordering of all three team's runners in which team A does not finish 1st, team B does not finish 2nd, and team C does not finish 3rd. Tentative "trial" placements of the team C runners are indicated by the superscripted ^{C}.

First, consider the AABB ordering for the race between teams A and B; as noted, team A beats team B here. Since there are two possible ways that team B can beat team C, their orders-of-finish must either be BBCC or BCBC. First assume a BBCC ordering. Given the initial AABB ordering and this BBCC ordering, there is just one way that the C team runners could have finished: the final order-of-finish would have to be AABB^{CC}; however, team A beats team C here by 3 to 7. So with the AABB and BBCC orderings, it cannot be the case that team C beats team A. So instead of the BBCC ordering, assume a BCBC ordering. Given the initial AABB ordering and this BCBC ordering, there is just one way that the C team runners could have finished: the final order-of-finish would have to be AAB^{C}B^{C}; however, team A beats team C here by a 3 to 7 score. So with the AABB and BCBC orderings, it cannot be the case that team C beats team A. In other words, given the AABB ordering for teams A and B, and the BBCC or BCBC orderings for teams B and C, there does not exist any final order-of-finish for team C that will produce a cycle. Furthermore, for both the AABB^{CC} and AAB^{C}B^{C} orders-of-finish, we know that in the dual meets team A beats team B, team B beats team C, and team A beats team C. But in both of these three-way meets, the orders-of-finish for the teams is that team A is 1st, team B is 2nd, and team C is 3rd. Hence, there does not exist any final order-of-finish for team C that will produce a violation of independence.

Second, then, consider the ABAB ordering for the race between teams A and B; as initially noted, team A beats team B here. Now recall that there are two ways that team B can beat team C: their orders-of-finish can be either BBCC or BCBC. Begin with the BBCC ordering. Given the initial ABAB ordering and this BBCC ordering, there is just one way that the C team runners could have finished: the order-of-finish would have to be ABAB^{CC};

however, team A beats team C here by a 3 to 7 score. So with the ABAB and BBCC orderings, it cannot be the case that team C beats team A. So instead of the BBCC ordering, assume a BCBC ordering. Given the initial ABAB ordering and this BCBC ordering, there are two ways that the C team runners could have finished: the order-of-finish could be either $\mathrm{AB^CAB^C}$ (but team A beats team C here by a 4 to 6 score) or $\mathrm{ABA^CB^C}$ (but team A beats team C by a 3 to 7 score). So with the ABAB and BCBC orderings, it cannot be the case that team C beats team A. In other words, given the ABAB ordering for teams A and B, and the BCBC or BBCC orderings for teams B and C, there does not exist any order-of-finish for team C that will produce a cycle. Furthermore, for these same three orders-of-finish—$\mathrm{ABAB^{CC}}$, $\mathrm{AB^CAB^C}$, and $\mathrm{ABA^CB^C}$—we know that in the dual meets team A beats team B, team B beats team C, and team A beats team C. But in all three of these three-way meets, team A is 1st, team B is 2nd, and team C is 3rd. Hence, there does not exist any final order-of-finish for team C that produces a violation of independence.

Since there are no other orderings by which team A can beat B, and by which team B can beat C, we conclude that there is no way that team C can beat team A. Hence, transitivity cannot be violated when $m = 3$ and $k = 2$, nor can independence from irrelevant teams be violated.

Appendix 2 Proof that violations of independence and transitivity can occur whenever $m \geq 3$ and $k \geq 3$

The proof demonstrates that when $m \geq 3$ and $k \geq 3$, every possible combination of m and k contains an order-of-finish for the runners that violates independence and transitivity.

The proof has two stages. First, we consider the case in which $m = 3$ and $k = 3$. It will be demonstrated that when $m = 3$, every possible value of $k \geq 3$ contains an order-of-finish for which independence and transitivity are violated. Second, we then build on the results from this first stage to demonstrate that for every value of $k \geq 3$, every possible value of m contains an order-of-finish for which independence and transitivity are violated.

For the first stage, involving the case in which $m = 3$ and $k = 3$, select an $m = 3$ and $k = 3$ order-of-finish that violates independence and transitivity, such as the CAABBBCCA order-of-finish discussed in part 2 of the main text. With this order-of-finish, team A beats team B by 9 to 12, team B beats team C by 9 to 12, yet team C beats team A by 10 to 11; and while team C beats team A in their dual meet, in the three-way meet team A (14 points) beats team C (16 points).

Next, take this initial $k = 3$ order-of-finish—CAABBBCCA—and append three runners in the BCA order, yielding the CAABBBCCA $\oplus$ BCA order-of-finish (where the symbol "$\oplus$" indicates the merging of the preceding and following orders-of-finish into a combined order-of-finish that respects the relative initial positions of the runners who were merged). This procedure generates the $m = 3$ and $k = 4$ case (each of three teams has four runners). The resulting order-of-finish violates both independence and transitivity: team A beats team B by 17 to 19, team B beats team C by 16 to 20, and team C beats team A by 17 to 19. In the three-way meet, team B is 1st with 25 points, team A is 2nd with 26 points, and team C is 3rd with 27 points; since team A beats team B in their dual meet, independence is thereby violated.

Given that $m = 3$, we have just shown for both the $k = 3$ case (an odd number) and the $k = 4$ case (an even number) that independence and transitivity are both violated.

Now take 6 additional runners (2 more for each of the three teams) and assume they have the ABCCBA order-of-finish; this ordering can be seen as a "null" ordering in the sense that if these 6 additional players are appended, in the ABCCBA order, to some previous order-of-finish, then these additional runners will not affect the outcomes (such as independence and transitivity both being violated) from this previous order-of-finish. The reason is simply that these three teams are all tied in this ABCCBA addition: for the team A vs. team B dual meet, team A gets $1+4=5$ more points while team B gets $2+3=5$ more points; the same holds for the dual meets between team B and team C and between team C and team A; in addition, the three teams are all tied in their three-way meet (each has 7 points). Thus, if we start with the CAABBBCCA order-of-finish and append this ABCCBA ordering, we have an CAABBBCCA $\oplus$ ABCCBA ordering, generating a case in which $m=3$ and $k=5$. But with this new ordering independence and transitivity are again violated (since the ABCCBA addition has no impact on the basic relationships among the scores of the teams in the initial ordering): in this new ordering, for example, we see that team A beats team B by 26 to 29, team B beats team C by 26 to 29, yet team C beats team A by 27 to 28, while in the three-way meet, team A is 1st with 39 points, team B is 2nd with 40 points, and team C is 3rd with 41 points; since team C beats team A in their dual meet, independence is thereby violated. In fact, by further repeating this process ($k=5+2=7$, $k=7+2=9$, and so on), we can generate every case when k is an odd number. Since the previous results are never changed when the ABCCBA null ordering is appended to a previous ordering, we have thereby demonstrated that independence and transitivity can be violated whenever k is an odd number ($k \geq 3$).

Next, return to the $k=4$ case, with the CAABBBCCA $\oplus$ BCA order-of-finish. To generate the $k=6$ case, simply append the ABCCBA null ordering to this $k=4$ order-of-finish, yielding the CAABBBCCA $\oplus$ BCA $\oplus$ ABCCBA order-of-finish. The resulting order-of-finish violates both independence and transitivity: team A beats team B by 38 to 40, team B beats team C by 37 to 41, and team C beats team A by 38 to 40 (independence is violated), while in the three-way meet, team B is 1st with 56 points, team A is 2nd with 57 points, and team C is 3rd with 58 points; since team A beats team B in their dual meet, independence is also violated. (Note that these general results—who beats whom in dual and three-way meets, and the violations of independence and transitivity—replicate the results for the $k=4$ case.) In fact, by further repeating this process ($k=4+2=6$, $k=6+2=8$, and so on), we can generate every case when k is an even number. Since the previous results are never changed when the ABCCBA null ordering is appended to a previous ordering, we have thereby demonstrated that independence and transitivity can be violated whenever k is even ($k \geq 4$).

Thus, it has been demonstrated that whenever $m=3$, independence and transitivity can be violated for every value of $k \geq 3$.

For the second stage of the proof, it must be shown that independence and transitivity can be violated for any value of $m > 3$, given every possible value of $k \geq 3$. But demonstrating this is trivial: simply take the order-of-finish that violates independence and transitivity for each value of $k \geq 3$, given $m=3$, and then append a fourth team at the end of the previous order-of-finish; this will turn the $m=3$ into $m=4$ and do so without undoing the violations of independence and transitivity that existed with the previous order-of-finish for that value of k. For example, take the order-of-finish CAABBBCCA that violates independence and transitivity when $m=3$ and $k=3$ and then append DDD at the end, thereby generating the CCAABBBCCA $\oplus$ DDD ordering. While we now have four teams (for the $m=4$, $k=3$ case), the addition of team D has no impact on the relationships among the scores for the first three teams: independence and transitivity are still violated by the relationships among the scores from the dual meets and the three-way meet among teams A, B, and C. Now take this $m=4$, $k=3$ case and append EEE at the end, thereby generating the

CCAABBBCCA $\oplus$ DDD $\oplus$ EEE ordering. What results is five teams (for the $m = 5$, $k = 3$ case), but independence and transitivity are still violated by the relationships among the scores from the dual meets and the three-way meet among teams A, B, and C. This can be repeated for each of the remaining k values for the $m = 3$ case, with the same results. In this way, it is shown that for each of the $k \geq 3$ orders-of-finish that violates independence and transitivity, when $m = 3$, there exists a corresponding value of m ($m > 3$) containing an order-of-finish that also violates independence and transitivity.

This completes the proof of the proposition that for every possible combination of m ($m \geq 3$) and k ($k \geq 3$), there exists an order-of-finish that violates both independence and transitivity.

References

Arrow, K. J. (1963). *Social choice and individual values*. New Haven: Yale University Press.

Brams, S. J., & Fishburn, P. C. (2002). Voting procedures. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (pp. 173–236). Amsterdam: North-Holland/Elsevier Science.

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

Kangas, W. (2003). Close cross country meet separates division opponents. *The Town Courier*, 5 October 2003, 23.

MacKay, A. F. (1980). *Arrow's theorem: the paradox of social choice: A case study in the philosophy of economics*. New Haven: Yale University Press.

Plott, C. R. (1976). Axiomatic social choice theory: an overview and interpretation. *American Journal of Political Science*, *20*, 511–596.

Saari, D. G. (2001). *Decisions and elections: explaining the unexpected*. New York: Cambridge University Press.

Schwartz, T. (1986). *The logic of collective choice*. New York: Columbia University Press.

Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.