The problem of determining overall rankings from a panel of voters with varying preferences is not a new one, nor is it an issue just for the ISU. If you have ever tried to decide where to go out to dinner with a group of friends, each with different preferences, you can understand the difficulties which often arise.

The Canadians Should Have Won!?

Maureen T. Carroll
University of Scranton
Elyn K. Rykken
Muhlenberg College
Jody M. Sorensen
Grand Valley State University

The biggest story of the 2002 Winter Olympics was the controversy surrounding the judging of the pairs figure skating event. Going into the final skate of the event, the favored Russian pair of Berezhnaya and Sikharudlidze was in the lead, but not by much. The final skate would determine the medals. The Canadian pair of Sale and Pelletier skated flawlessly, while the Russians faltered. It was a foregone conclusion that the Canadians had won the gold medal until the judges' marks were displayed. The Russians were placed first by five of the nine judges and came away with the gold. Disappointed that the judges did not recognize the flawless performance of the Canadians with the gold medal, Sandra Bezic, an NBC commentator and former Canadian pairs champion, said, "I'm embarrassed for our sport right now." The firestorm of public opinion that followed led to the unusual awarding of a second set of gold medals to the Canadians. This controversy also led the International Skating Union (ISU) to propose changes in the figure skating scoring system.

The problem of determining overall rankings from a panel of voters with varying preferences is not a new one, nor is it an issue just for the ISU. If you have ever tried to decide where to go out to dinner with a group of friends, each with different preferences, you can understand the difficulties which often arise. In this paper we discuss several methods for determining overall rankings from a collection of preferences, using figure skating as our context.

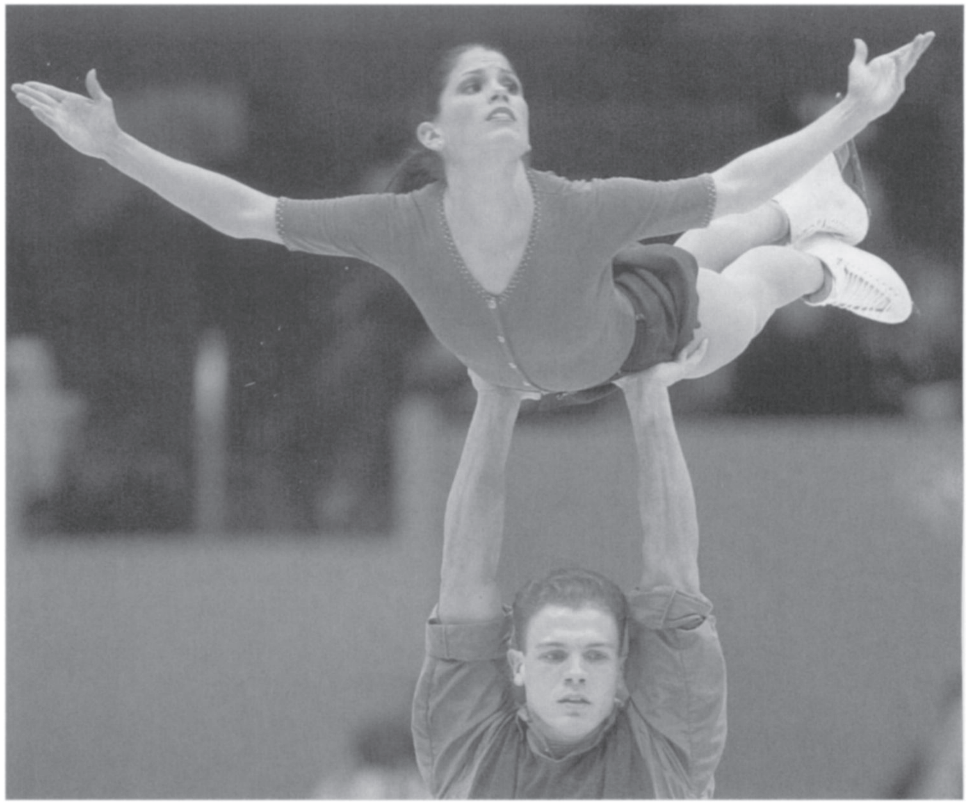
Mean Methods

The simplest possible method of selecting placements for a figure skating competition would be to have a single judge rank all of the skaters. While this method would give unambiguous results, there would certainly be controversy and concern surrounding the possible biases of the judge. For this reason, figure skating events always rely on a panel of judges to determine placements.

Suppose we have a panel of seven judges evaluating the free skate (or long program) for five skaters at a ladies' figure skating event. Traditionally each judge gives a score with one decimal place accuracy between 0 (worst) and 6 (best) for two categories: technical merit and artistic impression. In our examples, we will deal with the sum of these two marks, so our judges will give marks out of 12 points. Current rules do not allow the judges to give the same marks in both categories to two skaters. The sum of a judge's marks can, however, result in a tie. In this case, the skater with the greater artistic mark is ranked higher. We simplify matters here by assuming that a judge cannot give two different skaters the same sum.

The simplest way of determining an overall ranking from a set of judges' marks is to look at averages or means. The *Mean Method* sums the marks of the seven judges for each skater, and then divides this number by seven. We would then award first place to the skater with the highest mean, and so on. This is a straightforward, clear method of awarding rankings—so why isn't it used to judge figure skating? One drawback of the Mean Method is that it can easily result in ties between skaters, so another more complicated rule would have to be in place to break the ties. In addition, this method is extremely susceptible to biased judges—a single judge who gives a skater low marks can destroy that skater's chance of winning. In international competitions such as the Olympics it is often thought that judges are biased in favor of skaters from their own countries.

One way to get around the problem of a biased judge is to use the *Trimmed Mean Method*, in which the high and low marks are discarded for each skater, and the remaining five scores are averaged. This averts the possibility of a single judge's positive or negative bias affecting a skater's final placement. While the Trimmed Mean Method eliminates two outlying scores, it certainly is not perfect. Like the Mean Method, it is prone to ties. In addition, figure skating has a history of "block judging"—a group of judges deciding before an event that they will all vote in favor of one skater and against another. The Cold War, with its competition between East and West, encouraged such behavior. A block of biased judges can easily sway the results if rankings are determined by either means or trimmed means.



Canadian pairs figure skaters Jamie Sale and David Pelletier.

Ordinal Methods

The methods used to prevent the influence of block judging rely on *ordinal* marks. Once each judge has given his or her marks for all of the skaters, the judge's marks are then converted into rankings—first, second, third, and so on, or corresponding ordinals, 1, 2, 3, etc. There are many methods which can then be used to determine the final rankings based on these ordinal marks. We will discuss several of these. Consider the example in Table 1, with seven judges and five skaters.

		Judges						
		1	2	3	4	5	6	7
Skaters	*A*	11.2	11.4	11.6	11.4	11.2	11.4	11.2
	B	11.4	11.2	11.4	11.0	11.4	11.2	11.4
	C	11.6	11.6	11.8	11.6	11.6	11.6	10.6
	D	11.8	11.8	12.0	11.8	10.8	10.8	10.8
	E	11.0	11.1	11.2	11.2	12.0	12.0	12.0

Table 1

The corresponding table of ordinals is given in Table 2.

		Judges						
		1	2	3	4	5	6	7
Skaters	*A*	4	3	3	3	4	3	3
	B	3	4	4	5	3	4	2
	C	2	2	2	2	2	2	5
	D	1	1	1	1	5	5	4
	E	5	5	5	4	1	1	1

Table 2

The scores given in this example reflect a possible incident of block judging—judges 5, 6 and 7 seem to be biased in favor of skater *E* and against skater *D*.

Let's see how our first two methods would rank the skaters in this example. If we use the Mean Method, the means for skaters *A*, *B*, *C*, *D*, and *E* are 11.34, 11.29, 11.49, 11.4 and 11.5, respectively. The skaters finish in the order *E, C, D, A, B*. If we use trimmed means we have 11.32, 11.32, 11.6, 11.4 and 11.5, respectively, and the final order is *C, E, D,* $\{A, B\}$, where brackets represent a tie for fourth place between skaters *A* and *B*. As you can see, the mark of 10.6 given to skater *C* by judge 7 was enough to lower that skater's ranking to second with the Mean Method, but this outlier was eliminated with the Trimmed Mean Method and skater *C* ranked first.

Let's consider some methods which depend on the ordinal scores of the seven judges. Again, the simplest possible way to use these rankings is by averaging them. If we use this method on our example, we get the following average ranks: 3.29, 3.57, 2.43, 2.57, and 3.14. Thus the skaters are ranked *C, D, E, A, B* (remember that these are ordinal marks, so lower marks mean higher rankings). An equivalent method, called the *Borda Method*, was first used by Jean-Charles de Borda in 1781. This method assigns points to each of the five skaters based on each judge's preferences. The last place skater receives zero points, the next to the last skater, one point, and so on. The judge's first place skater receives four points in our example since there are five skaters. We then total the points a skater receives from all of the judges and rank them from highest to lowest. Table 3 shows these Borda points for our example.

		Judges							
		1	2	3	4	5	6	7	Total
	A	1	2	2	2	1	2	2	12
	B	2	1	1	0	2	1	3	10
Skaters	*C*	3	3	3	3	3	3	0	18
	D	4	4	4	4	0	0	1	17
	E	0	0	0	1	4	4	4	13

Table 3

Skaters	5– tuple	*LMR*	*SLM*	Ordinal sum of *LM*	Total sum of ordinals
A	(0, 0, 5, 2, 0)	3	5	15	23
B	(0, 1, 2, 3, 1)	4	6	20	25
C	(0, 6, 0, 0, 1)	2	6	12	17
D	(4, 0, 0, 1, 2)	1	4	4	18
E	(3, 0, 0, 1, 3)	4	4	7	22

Table 4

Notice that the average ordinal and Borda methods are equivalent, meaning that they will always produce the same ranking. For this reason, we will call both of them the *Borda Method*. We could also employ a *Trimmed Borda Method* by eliminating the high and low ordinals for each skater. For our example, the trimmed ordinal scores are 3.2, 3.6, 2, 2.4, and 3.2, producing a final ranking of $C, D, \{E, A\}, B$, where E and A are tied for third place. Note that the four methods considered so far have produced four different final rankings of the skaters.

The Borda and Trimmed Borda Methods, while not equivalent to the Mean and Trimmed Mean, have the same shortcomings. They are prone to ties, and they are susceptible to block judging. The ISU has used two different (and much more complicated) methods, Best of Majority and One-by-One, to correct these problems.

The *Best of Majority* (BOM) method works as follows. For each skater, we create a 5-tuple $(x_1, x_2, x_3, x_4, x_5)$ where x_i is the number of ordinals i that the skater has received. For example, looking at Table 2 we see that the 5-tuple for skater B is (0,1,2,3,1). We then look for the lowest majority rank, or *LMR*, where four or more judges represent a majority since there are seven judges. The lowest majority rank is given by

$$LMR = \min\left\{k \,\middle|\, \sum_{i=1}^{k} x_i \geq 4\right\},$$

that is, the smallest rank for which a majority of judges places a skater at that rank or higher. For skater B the *LMR* is **not** three since only three judges rank this skater three or higher. However, six judges place the skater at rank four or higher, so $LMR(B) = 4$. The size of the lowest majority, *SLM*, is the number of judges which comprise the lowest majority. For skater B we have $SLM(B) = 6$. These two quantities are not always enough information to determine final rankings. There are two sums which are used to break ties. The first is the sum of the ordinals from the judges in the lowest majority, and the other is the sum of all ordinals. Table 4 displays these five pieces of information for all five skaters.

The overall rankings are determined by the *LMR*. For example, skater D ranks first, C second, and A third. In the case of a tie, as with skaters B and E, the skater with the greater *SLM* receives the higher rank. Thus, the overall ranking is D, C, A, B, E. If the *SLM*s were equal, then the skater with the smallest sum of ordinals from her lower majority receives the higher rank. If this sum does not break the tie, then the total sum of ordinals is used. If all of these numbers agree, then the skaters remain tied in the overall rankings.

BOM is complicated, but we can see that it is the first method to overcome the effects of block judging in our example. Skater D, the clear choice of four of the seven judges, ended up in first place, despite the block of judges opposed.

The BOM method is still used to judge US national figure skating competitions. However, in 1998, the ISU changed judging systems from BOM to the *One-by-One* (OBO) system. This switch was made because it was thought that using OBO would eliminate "swaps"—two skaters changing relative placements after a third skater performs. (It can, however, be shown that swaps are possible under OBO.)

The OBO method uses the judges' ordinals to determine the final rankings by comparing each skater to every other skater and giving one point to the winning skater in each pairing. For example, suppose we look at Table 2 and consider the pairing of skater A and skater B. If a judge ranks A above B, then A receives 1 judge-in-favor (*JIF*) point, and vice versa for B. In our example, since judge 1 ranked B third and A fourth, B receives one *JIF* point. When we continue this analysis with the pairing of A and B for all the judges, we see that skater A has 4 *JIF* points and skater B has 3. Then we do this for all of the possible pairings. In our example there are $\binom{5}{2}$ or 10 such pairings. A *JIF* matrix is helpful to display the results of these pairings. Entry (i,j) represents the number of judges who favor the skater in row i over the skater in column j. A column representing the total for each row is included as it will be useful in determining the overall rankings.

	A	*B*	*C*	*D*	*E*	Total
A	–	4	1	3	4	12
B	3	–	1	3	3	10
C	6	6	–	2	4	18
D	4	4	5	–	4	17
E	3	4	3	3	–	13

Since skater A has more *JIF* points than skater B (4 to 3), skater A "wins" over B and receives one *WIN* point. We proceed in this fashion for all possible pairings of skaters to produce the *WIN* matrix below, where each entry (i,j) is 1 if the

skater in row i has more *JIF* points than the skater in row j, and 0 otherwise. The Total column will be used to determine the overall rankings.

	A	*B*	*C*	*D*	*E*	Total
A	–	1	0	0	1	2
B	0	–	0	0	0	0
C	1	1	–	0	1	3
D	1	1	1	–	1	4
E	0	1	0	0	–	1

Final rankings are determined by the number of *WIN*s, with higher *JIF* scores used to break ties. This is unnecessary in our case, and the final rankings are *D, C, A, E, B*. Note that these final rankings, while slightly different than the BOM results, have also succeeded in defeating the block of judges opposed to skater *D*.

Proposed Method

In reaction to the negative publicity surrounding the 2002 Olympics pairs skating event, ISU president Ottavio Cinquanta has proposed a new judging method. The new method will see fourteen judges marking each event, but only nine of these judges' marks will actually be used to determine the final standings. The group of nine will be chosen at random at the conclusion of the skating event, with the same nine judges being used for all skaters. Once the nine judges are determined, the Mean Method will be used to determine the overall ranking. We will call this method *Nine of Fourteen* (NOF).

In order to evaluate this new method, we need to consider some general properties of methods which determine overall rankings from a set of preferences that we might want a judging method to exhibit. We will consider four properties within the context of figure skating: monotonicity, rank majority, the Condorcet winner criterion, and reproducibility.

Monotonicity, also called "incentive compatibility" or "non-perversity," is the property that a skater's final rank cannot be made **worse** by a judge who improves either the skater's raw marks or her ranking. All of the judging methods discussed in this paper, including the proposed new method, satisfy this condition. (For examples of methods that do not exhibit monotonicity, consult Alan Taylor's book *Mathematics and Politics: Strategy, Voting, Power and Proof.*)

The *rank majority* property is one possible interpretation of the idea of majority rule. As such, it is a way of saying that a judging system is not affected by block judging. It is quite limited in that it only applies under a narrow set of circumstances. Specifically, suppose at least half of the judges rank a skater *A* at rank r and at least half of the judges rank skater *B* at rank q. The rank majority property requires that if $r < q$, then overall, skater *A* will be ranked higher than skater *B*. Notice how restrictive this condition is—it requires a majority of the judges to agree on the precise rank of skater *A* and a (possibly different) majority of the judges to agree on the precise rank of skater *B*.

In their article *Rating Skating*, Bassett and Persky show that any system that satisfies both monotonicity and rank majority must be equivalent to BOM. Thus BOM satisfies this property. Notice in Table 2, since a majority of judges rank skater D first and a majority rank skater C second, the rank majority property requires skater D's overall ranking to be above skater C's overall ranking. Thus from this example, it is clear that Mean, Trimmed Mean, Borda, and Trimmed Borda do not satisfy this property. Once the nine judges are chosen, NOF becomes the Mean Method and hence it also does not satisfy this property. To see that OBO does not satisfy this property consider the example given in Table 5, with seven judges and four skaters. Here we give only the ordinals.

The rank majority property requires skater *A* to be ranked above skater *B*, but skater *A* has 1 *WIN* point and skater *B* has 2 *WIN*s, and the final ranking is *D, B, A, C*.

		Judges						
		1	2	3	4	5	6	7
	A	2	2	2	2	4	4	4
	B	1	3	3	3	3	3	3
Skaters	*C*	3	4	4	4	2	2	2
	D	4	1	1	1	1	1	1

Table 5

Another possible interpretation of majority rule is the *Condorcet winner criterion.* A skater *S* is a Condorcet winner if when paired with any other skater, a majority of judges prefer skater *S* to the other skater. (In Table 2 and Table 5, skater *D* is the Condorcet winner.) In any given competition there may well be no Condorcet winner. A judging method that always ranks a Condorcet winner first (when there is one) is said to satisfy the Condorcet winner criterion. If a Condorcet winner exists, then he or she will also be the winner using the OBO method. This is not hard to see. The Condorcet winner has n–1 *WIN* points, where n is the total number of skaters participating in the event. All other skaters will have fewer than n–1 *WIN* points since they lost to the Condorcet winner.

This property only provides information about the winner. One might be tempted to extend this criterion to the more general principal, "If a majority of judges prefers skater *A* to skater *B*, then overall, *A* should be ranked higher than *B*." Unfortunately, this is not possible since such a ranking may not be transitive. In other words, it may be possible for a majority of judges to rank *A* above *B*, *B* above *C*, and *C* above *A*. Consider the following simple example with three judges and three skaters where this situation occurs.

		Judges		
		1	2	3
	A	1	2	3
Skaters	*B*	2	3	1
	C	3	1	2

In such cases, no clear ranking is possible using this general criterion.

In Table 2, using the Condorcet winner criterion, skater *D* must be the winner. Hence it is clear that Mean, Trimmed Mean, Borda, Trimmed Borda, and NOF do not satisfy this property. To see that BOM also does not satisfy this property consider the example given in Table 6, with seven judges and four skaters.

		Judges						
		1	2	3	4	5	6	7
	A	1	1	1	3	3	3	3
	B	2	2	2	2	4	1	4
Skaters	*C*	3	3	3	4	2	2	1
	D	4	4	4	1	1	4	2

Table 6

Here, skater *A* is the Condorcet winner, but BOM places skater *B* in first place. In fact, since any method with the rank majority property would place skater *B* above skater *A*, it is clear from this example that no voting method can satisfy both the rank majority property and the Condorcet winner criterion.

To minimize the impact of block voting, one could argue that the judging method should either satisfy rank majority or the Condorcet winner criterion. It is also important to note that under these conditions if a voting block becomes part of the majority then it would be unfair not to acknowledge their legitimacy to pick the winner. Since NOF does not satisfy either criterion, we consider it less desirable than either BOM or OBO. As it also employs the mean method, NOF is susceptible to the problems inherent therein, namely that it is prone to ties and vulnerable to block voting.

To illustrate the properties of NOF, let's see the rankings that result from Table 1. This example had seven judges, so we will simply use each judges' marks twice to do the calculations. Using *Maple*, we had a computer run through all $\binom{14}{9} =$ 2002 possible ways to choose the nine judges whose marks will count, and determined the rankings in each case. The table below shows the number of times that each of the skaters placed in each position, depending upon which of the nine judges were chosen.

	first	second	third	fourth	fifth
A	0	0	550	1394	58
B	0	0	88	440	1474
C	872	1047	56	0	0
D	190	496	804	60	452
E	940	432	504	108	18

Recall that in this example, our "block" of judges voted for skater *E*, as opposed to the general favorite, skater *D*. The table above shows that about 47% of the time, skater *E* receives first place, while skater *D* (the favorite of the majority of the judges) is ranked first only about 9.5% of the time. We can see that most of the time this block of judges is able to influence the overall rankings.

Finally, we wish to define *reproducibility*. By this we mean that if the competition were held again and the raw marks (and therefore the ranks) were identical, then the final rankings would remain the same. Clearly the only method that does **not** satisfy this property is NOF. To demonstrate the lack of reproducibility of NOF, let's consider the ladies' event from the 2002 Olympic Games, and look at all 2002 of the possible outcomes. Since the current system employs only nine judges, and we needed 14 marks to do this analysis, we chose to duplicate the marks of the same five judges for each skater. (While there are 126 ways to do this, we found similar variance of outcomes regardless of which five we chose to duplicate.) We also limited the scope of this analysis to the top five finishers.

Continued on p. 22

11 A Three-Dice Stack

You need three dice for this test. Toss the first die on the table. On top of it put the tossed second die. The third die goes on top of the other two, turned so its top face is 1.

If you inspect this stack from all sides, you'll note that five faces cannot be seen. Add these faces as follows: check the two touching faces between the top and middle dice. Write down their sum, and put the top die aside. Check the two hidden faces that are touching between the two dice that remain. Add the numbers, write down the sum, and put the top die aside. Subtract both touching face sums from 20, and say this number out loud as you look at the bottom face of the remaining die.

What do you notice?

12 Nine-Card Spell

Remove nine cards from a deck. Shuffle them, and then hold them face down in your left hand. Reverse the third card from the top of the packet. Spell the name of the reversed card as follows. Let's assume it was the queen of hearts. Spell Q-U-E-E-N by dealing five cards to the table, one card for each letter. Place the remaining cards on top of the five just dealt.

Pick up the packet. Spell O-F by dealing two cards to the table. Again, put the remaining cards on top of those just dealt.

Now spell H-E-A-R-T-S. Put the cards in your hand on top of the tabled pile.

Follow this procedure, using the name of the card you have reversed. Note that the number of letters in the name can vary from 10 (for example, the ace of clubs) to 15 (for example, the eight of diamonds).

After spelling the name of the reversed card, how far down is it from the top of the packet? ■

Answers

See answers on page 26.

Continued from p. 9

	first	second	third	fourth	fifth
Hughes	491	1501	10	0	0
Slutskaya	1511	491	0	0	0
Kwan	0	10	1992	0	0
Cohen	0	0	0	2002	0
Suguri	0	0	0	0	2002

As you can see in the table above, the final rankings for the top three skaters depend upon which nine of the fourteen judges are chosen. The determination of the Olympic champion is left to chance. This lack of reproducibility is one of the most troubling aspects of the newly proposed system. For the competitors, this method is especially unfair and capricious. Perhaps the ISU was lured by the idea that introducing an element of randomness into the judging method would increase its fairness. The examples above clearly show that this is not the case. In light of this and the fact that the proposed method would not prevent negative block voting, it appears to be a hasty and ill-planned change.

Thinking back to our motivating example, what is interesting to note is that a majority of judges preferred the Russians to the Canadians. In this case, one should be very skeptical of a system which awards the gold to the Canadians. Instead of changing the judging method, the skating community would be better served if there were changes in the training of judges and in the punishment of those found guilty of block voting. The recent three-year suspensions given to the French judge and her supervisor seem to be a promising sign. However, a recent US proposal requiring a lifetime ban for any judge convicted of ethical violations was voted down by the ISU. Without such sanctions in place, it seems that the judging of figure skating might remain an "embarrassment" for years to come. ■

For Further Reading

For up-to-date information on the proposed judging system, see `www.isu.org`. For more details about the BOM method, read the article "Rating Skating" by Gilbert Bassett and Joseph Persky in the September 1994 *Journal of the American Statistical Association*. For details on judging systems for other sports, see the 1993 book *Mathematics and Sports* by L.E. and A.L. Sadovskii. Finally, for an introduction to collective choice theory, read Alan Taylor's 1995 book, *Mathematics and Politics: Strategy, Voting, Power and Proof*.