

# Chance, On and Off the Ice

(Published in *CHANCE* Vol. 20, No. 2, 2007)

John W. Emerson

---

In figure skating, past controversy over judging has led to new rules that invoke a random selection of judges. That randomness can, and in fact has, affected the outcome of competitions.

---

The medals awarded in the Pairs competition of the 2006 World Figure Skating Championships did not reflect the consensus of the judges. Why? Under the International Skating Union (ISU) rules, chance (not *CHANCE*) played a role in scoring the competition. The scores of three judges were discarded at random in the Short Program, and another set of three judges' scores were eliminated in the Free Skate.

Unfortunately for Russian bronze medalists Maria Petrova and Alexei Tikhonov, the randomness introduced by the ISU rules (not simply their performances on the ice) made the difference between silver and bronze. In fact, the competition was close enough that different panels of judges could easily have awarded them gold. I can't give Petrova and Tikhonov the silver medal I think they deserve, or the gold they might have won with a bit of luck. However, I can say that the consensus of 12 judges would have awarded the silver medal to Petrova and Tikhonov at the 2006 World Figure Skating Championships.

## **Olympic Winter Games, February 2006**

During NBC's Prime Time broadcast of the 2006 Olympic competition, commentator Bob Costas referred to the new figure skating scoring system as "designed to increase fairness"—fallout from the judging scandal in Salt Lake City, Utah. Two-time gold medal winner and veteran analyst Dick Button offered his support of the new system. The viewer was comforted; the integrity of the Olympic Games was intact.

On some level, the new scoring system increases fairness, but it has introduced the unsettling possibility of luck influencing the medal standings. In a close competition, the outcome will likely be determined by the random selection of panels of nine judges. For more than 100 years, panels of judges used the 6.0 standard of scores. Judging was not anonymous, and accusations of favoritism were common. The starting order often influenced the scores, with earlier skaters receiving lower scores to "leave room" for possible superior performances later in the session.

In place since the 2004 World Championships, the new system awards points for technical elements and five program components: skating skills, transition/linking

footwork, performance/execution, choreography/composition, and interpretation. The scores for the technical elements depend on a base value for the level of difficulty of the elements. The 12 judges add or deduct points from this base value, thereby acknowledging the “grade of execution” of the elements. Program component scores reflect the overall presentation of the program and quality of the figure skating.

Judging is now anonymous. Nine of 12 judges are selected at random for the Short Program; the random selection is repeated for the Free Skate. Scores for each executed element or program component are calculated using a trimmed mean, as in the old system, by dropping the maximum and minimum of the nine scores.

Random elimination of three judges produces 220 possible nine-judge panels for each program segment, each with slightly different scores. However, only one of these panels actually determines the competition scores for a program segment. Katherine Godfrey provided one of the earliest assessments of the new scoring system in 2003, prior to its official adoption:

The ISU has claimed that random selection of judges will make it impossible to create blocs, because the dealmakers would not know whom to approach. However, picking a random number of judges from a panel won't eliminate the ability to set up blocs. They may have to be bigger than before to ensure results, but they'll have the advantage of being undetectable once they're formed. ...Selecting random judges from a completely honest panel creates a fairness issue that does not exist when you use all the judges' marks.

In the absence of the proverbial train wreck, the problem with the scoring system remained abstract. The ISU acknowledged that the random selection of panels “may affect to a small degree” the overall score, but told ABC World News Tonight that “it stands by the system.” At the Olympics, the ISU was lucky.

### **World Figure Skating Championships, March 2006**

At the 2006 World Figure Skating Championships, some of the skaters were lucky; others were not. In the Pairs competition, the random selection of judges resulted in an unlikely event, given the performances on the ice: the Chinese team of Zhang Dan and Zhang Hao were placed ahead of the Russians, Petrova and Tikhonov.

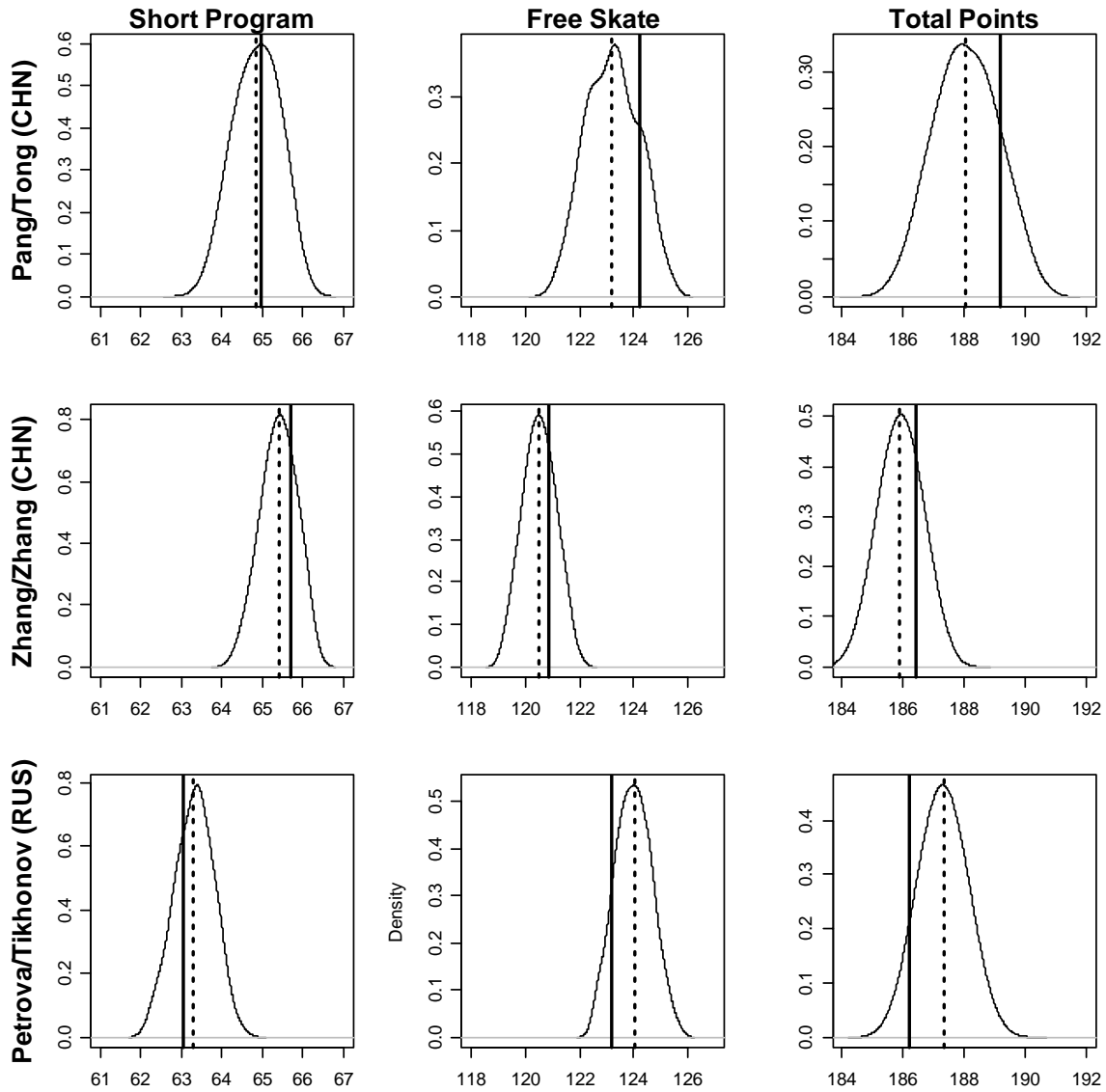
Although I planned to analyze the results immediately following the competition, I didn't complete the analysis until eight months later. Perhaps I shouldn't have been surprised by the results, but I asked my graduate students to validate them. Even Google found no evidence of outraged fans or an analysis of the competition results. The results can be described in terms of what might have occurred (with other panels of judges selected by the computer), or with respect to the consensus of the entire panel of 12 judges. No matter how you look at it, Petrova and Tikhonov were victims of a scoring system based on chance.

2006 World Figure Skating Championships	Actual Competition Results		Consensus Results (uses entire panel of 12 judges)	
	Team (Country)	Points	Medal	Points
Pang/Tong (China)	189.20	Gold	188.05	Gold
Zhang/Zhang (China)	186.42	Silver	185.90	Bronze
Petrova/Tikhonov (Russia)	186.22	Bronze	187.32	Silver

**Table 1.** The actual results relied on the random selection of scores from nine of twelve judges in the Short Program, and a different random selection of nine judges' scores in the Free Skate. The consensus results include the scores of all twelve judges for both programs, and would have awarded bronze medalists Petrova and Tikhonov the silver medal.

Table 1 compares the actual competition scores to the consensus of the entire panel of judges. In the competition, two panels of nine judges were selected at random, and two sets of three judges' scores were unused. This random selection could have produced any one of  $220 \times 220 = 48,400$  possible combinations of panels. Of these, a mere 2,713 (5.61%) placed Zhang and Zhang ahead of Petrova and Tikhonov, while 13,385 (27.65%) awarded Petrova and Tikhonov the gold medal, ahead of actual winners Pang Qing and Tong Jian (also from China).

The kernel density estimates in Figure 1 show the distribution of total points, as well as scores in the Short Program and Free Skate, induced by the random selection of panels. For each team, each of the 220 possible panels of nine judges produces slightly different scores in each program segment. Thus, the distribution of total points reflects 48,400 possible combinations of panels. The dotted line denotes the consensus score of the entire panel of 12 judges, while the solid line shows the scores observed through the random selection of panels in the competition. Both Chinese teams were lucky in the random selection of panels, receiving higher scores in the competition than would have been awarded by the consensus of the 12 judges.



**Figure 1.** Pang and Tong (China) won gold, Zhang and Zhang (China) won silver, and Petrova and Tikhonov (Russia) won bronze; the actual scores are marked by the solid lines. The dotted lines indicate the consensus scores of the entire panel of twelve judges, which would have awarded the Russians the silver, relegating Zhang and Zhang to the third spot on the podium.

## **Does the New Scoring System Reduce the Impact of Biased Judges?**

It would be nice to believe all judges try to score the competition fairly. However, the scoring system was changed partly in response to alleged corrupt judging. Can the new scoring system reduce the impact of corrupt judges on the results of the competition? After all, the scores of a single judge would only count with probability 0.75 in any segment of the competition. And, if a score might not count, maybe a corrupt judge wouldn't even try to influence the results? Perhaps illicit purchasing or exchanging of points will quietly vanish, because nobody could guarantee a positive (or negative) tweak in the scores without involving at least four judges?

I can only speculate on the psychological impact of the random selection of judges on potential score-fixing. I don't happen to believe it provides much of a deterrent, and the new anonymity of scores won't help, either. I can only speculate that anonymity of scores may relieve judges of nationalistic pressures.

I can, however, compare two scoring systems in the presence of a biased judge. If 12 judges score the competition, I recommend using all 12 sets of scores, along with the current system of using trimmed means. This will guarantee that the biased judge influences the results, unless the judge's scores are dependably trimmed as the maximum (or minimum) scores among the 12. I will compare this method to the official scoring system of the ISU.

I selected one of the judges from the 2006 World Championships and decided to see how this judge could help the unfortunate Petrova and Tikhonov. I added 0.25 points to each of the program component scores for this judge and one "grade of execution" level to each of the elected element scores (as long as the resulting score did not exceed the maximum possible score). I calculated the effect of this "judging bias" on the consensus score of the 12 judges and examined the impact on the distribution of the 48,400 possible scores under the ISU rules.

Obviously, the scores of both Chinese teams remained the same. The consensus of the 12 judges would have risen to 187.85 for the Russians (up from 187.32), enough for the silver medal, but not enough for the gold. An examination of the 48,400 combinations of panels produced the following results: 43.29% of the panels would have awarded the Russians the gold (up from 27.65% without the biased judge) and 97.64% would have placed Petrova and Tikhonov ahead of Zhang and Zhang (up from 94.39% in the actual competition).

The impact of the biased judge on the consensus score was  $187.85 - 187.32 = 0.53$ , and 59.24% of the panels would have provided increases larger than 0.53. Although these results would vary slightly if a different judge was used as the "biased judge," the implication is clear: The ISU rules can either magnify or mute the influence of a biased judge, relative to the consensus of the entire panel of judges. We are left hoping that the new scoring system provides a psychological deterrent against corrupt judging, because it certainly isn't a mathematical deterrent.

## **Clear Consequences**

The figure skating community continues to debate refinements to the scoring system. I hope the ISU will follow the lead of U.S. Figure Skating, which uses panels of nine judges without any random selections. Of course, members of the ISU may believe there are advantages to the random selection of judges. But there are clear consequences to the skaters in close competitions, as demonstrated at the 2006 World Figure Skating Championships.

Before publishing this article, I tried unsuccessfully to contact Petrova and Tikhonov; I wanted to congratulate them on the quality of their final competition, which was judged by 12 professionals to be worth a silver medal. Petrova and Tikhonov retired following the 2006 World Figure Skating Championships; they ran out of chances to win gold. Under the ISU rules, they had a great chance to win silver because they skated well enough for a good chance at gold. Instead, they went home with bronze—dumb luck off the ice.