



2003

# On the Selection of Judges in International Figure Skating

Michael B. Abramowicz

*George Washington University Law School, [abramowicz@law.gwu.edu](mailto:abramowicz@law.gwu.edu)*

Follow this and additional works at: [http://scholarship.law.gwu.edu/faculty\\_publications](http://scholarship.law.gwu.edu/faculty_publications)

 Part of the [Law Commons](#)

---

## Recommended Citation

Michael Abramowicz, On the Selection of Judges in International Figure Skating, 6 GREEN BAG 2D 339 (2003)

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in GW Law Faculty Publications & Other Works by an authorized administrator of Scholarly Commons. For more information, please contact [spagel@law.gwu.edu](mailto:spagel@law.gwu.edu).

## On the Selection of Judges in International Figure Skating

Michael Abramowicz

A 5-4 DECISION with international ramifications. Television commentators who had confidently predicted a victory for the ultimately defeated duo. Opponents of the decision who argued that the result was illegitimate and the judges biased. First *Bush v. Gore*,<sup>1</sup> then *Salé/Pelletier v. Berezhnaya/Sikharulidze*.<sup>2</sup> In democracy, contests are

supposed to be resolved by the vote, not the blade; in figure skating, it's the other way around. And yet both the 2000 presidential election and the 2002 pairs figure skating competition ended with accusations of partisanship. Even in the supposedly objective realms of sport and vote counting, personal preferences can sometimes matter.

---

*Michael Abramowicz is an Associate Professor of Law at George Mason University and a Visiting Associate Professor of Law at George Washington University. He thanks Debbie Mellis for her research assistance. Any remaining technical errors in this essay are balanced by its artistic presentation.*

<sup>1</sup> 531 U.S. 98 (2000).

<sup>2</sup> The controversy began at the 2002 Salt Lake Olympic Games when Canadian pairs skaters Jamie Salé and David Pelletier lost to Russians Elena Berezhnaya and Anton Sikharulidze, despite claims by many that the Canadians had skated better. See, e.g., Amy Shipley, *Russian Pair Garners Gold; Controversial Judging Keeps Canadians 2nd*, WASH. POST, Feb. 12, 2002, at D1. But see Ivan Lebedev, *The Better Pair Won the Gold; Controversy Aside, Russians Earned It*, WASH. POST, Feb. 14, 2002, at D10 (presenting the Russian perspective that the Russians deserved the gold). The controversy intensified when French judge Marie Reine Le Gougne tearfully claimed that there had been an improper attempt to influence her, although she later claimed that it was supporters of the Canadians, not the Russians, who had attempted to do so. See, e.g., Amy Shipley, *French Judge Now Blames ISU Officials*, WASH. POST, Feb. 22, 2002, at D11. Eventually, the Canadian skaters were awarded a duplicate set of gold medals, even though six of nine judges gave the Russians higher marks for artistic performance in the long program. See, e.g., Amy Shipley, *In Skating, a Golden Glow; Canadian Pair Finally Receives First-Place Medals*, WASH. POST, Feb. 18, 2002, at A1. Le Gougne, meanwhile, received a three-year suspension from judging. See *Suspended French Judge Drops Appeal*, WASH. POST, July 26, 2002, at D2.

The method by which judges are selected can matter too. In a forthcoming essay in the *California Law Review*, Stephen Choi and Mitu Gulati propose, apparently seriously, that Supreme Court Justices be selected from a tournament of courts of appeals judges.<sup>3</sup> Among the relevant criteria would be opinion publication rates, citations received, and reversal rates.<sup>4</sup> Even if these proxies are weak, Choi and Gulati suggest, at least the system would eliminate the taint of political motivation.<sup>5</sup> The method for selecting Supreme Court Justices is unlikely to be changed,<sup>6</sup> let alone to a system modeled on sports. The International Skating Union's system is almost sure to be changed, however, and a tournament model might well work there.

The system that the ISU long used seems worse than advice-and-consent procedures in any event. The national skating federations select judges, with the countries represented chosen at random from among the nationalities of the skaters participating at the World Championships. This creates two problems. First, federations have an incentive to pick judges who are most likely to favor their interests, whether because those judges happen to like a particular style of skating, or because the judges seem particularly eager to bring home glory. Supreme Court Justice selection may be based on similar considerations,<sup>7</sup> but at least it is tempered in times of divided government by the requirement of confirmation. Second,

skating judges are beholden to their national federations, and their incentives are to be partisan rather than objective, at least if they hope to judge again.

This essay offers an alternative approach to selection of skating judges, describing a tournament-like system that could be used to pick judges objectively. The proposal itself is simple: The judges themselves should be judged. Using another set of judges to rate the actual judges would be cumbersome, and it would merely recreate the problem of bias at a different level of abstraction. Rather, judges could be rated based on the correlation of their scores with those of other judges. Judges with high correlation scores would earn the right to judge ever more important levels of competition, while judges with low scores would suffer demotion. The judges at the most prestigious events – the World Championships and the Olympics – would thus be those who have consistently judged like other judges.

This system might seem to reward the pursuit of averageness. It takes considerable skill, however, to be consistently average, and idiosyncrasy in skating judges is a vice. Part I will evaluate the reforms that the ISU has considered to address the problems of judging in skating, and it will explain why these reforms are unlikely to be successful. Part II will provide a game theoretic justification for my proposal. I will conclude by suggesting extensions.

3 Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. (forthcoming Jan. 2004), available at <http://papers.ssrn.com/abstractid=394700>. The winner of the tournament, Choi and Gulati write, would receive "presumptive promotion to the Court." *Id.* (manuscript at 5). A goal of the proposal, they indicate, is to stimulate analysis of objective factors in selecting Supreme Court Justices.

4 *Id.* (manuscript at 5).

5 *Id.* (manuscript at 6).

6 Choi and Gulati acknowledge this. *Id.* (manuscript at 30).

7 See, e.g., Gerald N. Rosenberg, *The Supreme Court and the Attitudinal Model*, L. & CTS. NEWSL., Spring 1994, at 6 (reviewing JEFFREY A. SEGAL & HAROLD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL* (1993)) (arguing that the appearance of Justices' voting consistently with their presumed political opinions is best explained by the theory that the President picks Justices whose interpretive philosophy will tend to produce results favored by the President).

## I. THE ISU'S PROPOSALS & REFORMS

The ISU rejected a proposal in 2001 that would have centralized judge selection in the ISU itself.<sup>8</sup> After the pairs controversy, the ISU adopted a new package of solutions, using more judges but having some scores randomly disregarded, and changing from the traditional 6.0 scale to a more complex approach taking into account the degree of difficulty.<sup>9</sup> I will consider these approaches in turn, but first it is worth assessing why we should even be concerned about vote trading by figure skating judges. Economists, after all, generally believe that trades will produce allocative efficiency, and political scientists note that logrolling may be useful where a minority on a particular issue has strong preferences on that issue.<sup>10</sup> Perhaps vote trading in figure skating makes everyone better off.

A technical response might focus on agency costs or externalities. The agency cost argument is that even if judges maximize their joint utility by engaging in trading, they may not be acting on behalf of the broader figure skating community. The externality argument is that logrolling itself imposes a cost on figure

skating fans, who will like the product more if they believe that there is integrity in the judging. After all, figure skating presumably would be less interesting if judges or competitors were allowed to bid on the right to a gold medal.

To be sure, some corruption may be better for the sport and its television ratings than none. Tonya Harding's supporters helped promote figure skating by clubbing Nancy Kerrigan,<sup>11</sup> although that form of cheating has little to do with the judging. More relevantly, the recent pairs controversy increased television ratings.<sup>12</sup> Even if corruption has a positive short-term impact, however, it conceivably could have a corrosive long-term effect.<sup>13</sup> Moreover, corruption in figure skating may attract attention to that sport only at the expense of the image of the Olympics as a whole.<sup>14</sup> In any event, I will consider reforms under the assumption that a process with integrity is indeed the goal.

### A. Centralized Judge Selection

The theory behind centralizing judge selection is that centralization would prevent individual countries from choosing individuals to serve their own interests.<sup>15</sup> The

8 See Selena Roberts, *Judge's Early Tears Indicated Controversy to Come*, N.Y. TIMES, Feb. 17, 2002, § 8, at 1 (discussing the consideration of the earlier proposal).

9 See, e.g., *Pairs Scandal Yields Reform; Point System Among Changes Proposed by Skating Union*, AP, Feb. 19, 2002, available at 2002 WL 6448241; Amy Shipley, *A Small Change for Big Effect: ISU's New Judging System May Discourage Cheaters*, WASH. POST, June 9, 2002, at D6.

10 See, e.g., DENNIS C. MUELLER, PUBLIC CHOICE II at 82-87 (1989).

11 See Richard Sandomir, *No Knee-Whacking, but It's a Ratings Spike*, N.Y. TIMES, Feb. 15, 2002, at 7.

12 *Id.*

13 At least, sports leagues respond as if corruption would have an adverse future effect on the league. See ELIOT ASINOF, EIGHT MEN OUT: THE BLACK SOX AND THE 1919 WORLD SERIES (1988) (discussing the 1919 scandal and its aftermath).

14 Cf. Michael Wilbon, *Go Figure: Skating Lands on Thin Ice*, WASH. POST, Feb. 14, 2002, at D1 (arguing, in the wake of the recent controversy, that "any event in which the outcome is determined exclusively by judging ain't a sport").

15 ISU President Ottavio Cinquanta argued against the proposal on the ground that "even if it is the I.S.U. picking the judges, these judges are coming from nations." Roberts, *supra* note 8. This explanation is unpersuasive, because national federations are likely to pick the most nationalistic judges. Moreover, he disregards that judges may feel they must follow instructions from their federations, which is precisely the alleged source of the current controversy.

The method by which judges are selected can matter too. In a forthcoming essay in the *California Law Review*, Stephen Choi and Mitu Gulati propose, apparently seriously, that Supreme Court Justices be selected from a tournament of courts of appeals judges.<sup>3</sup> Among the relevant criteria would be opinion publication rates, citations received, and reversal rates.<sup>4</sup> Even if these proxies are weak, Choi and Gulati suggest, at least the system would eliminate the taint of political motivation.<sup>5</sup> The method for selecting Supreme Court Justices is unlikely to be changed,<sup>6</sup> let alone to a system modeled on sports. The International Skating Union's system is almost sure to be changed, however, and a tournament model might well work there.

The system that the ISU long used seems worse than advice-and-consent procedures in any event. The national skating federations select judges, with the countries represented chosen at random from among the nationalities of the skaters participating at the World Championships. This creates two problems. First, federations have an incentive to pick judges who are most likely to favor their interests, whether because those judges happen to like a particular style of skating, or because the judges seem particularly eager to bring home glory. Supreme Court Justice selection may be based on similar considerations,<sup>7</sup> but at least it is tempered in times of divided government by the requirement of confirmation. Second,

skating judges are beholden to their national federations, and their incentives are to be partisan rather than objective, at least if they hope to judge again.

This essay offers an alternative approach to selection of skating judges, describing a tournament-like system that could be used to pick judges objectively. The proposal itself is simple: The judges themselves should be judged. Using another set of judges to rate the actual judges would be cumbersome, and it would merely recreate the problem of bias at a different level of abstraction. Rather, judges could be rated based on the correlation of their scores with those of other judges. Judges with high correlation scores would earn the right to judge ever more important levels of competition, while judges with low scores would suffer demotion. The judges at the most prestigious events – the World Championships and the Olympics – would thus be those who have consistently judged like other judges.

This system might seem to reward the pursuit of averageness. It takes considerable skill, however, to be consistently average, and idiosyncrasy in skating judges is a vice. Part I will evaluate the reforms that the ISU has considered to address the problems of judging in skating, and it will explain why these reforms are unlikely to be successful. Part II will provide a game theoretic justification for my proposal. I will conclude by suggesting extensions.

3 Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. (forthcoming Jan. 2004), available at <http://papers.ssrn.com/abstractid=394700>. The winner of the tournament, Choi and Gulati write, would receive "presumptive promotion to the Court." *Id.* (manuscript at 5). A goal of the proposal, they indicate, is to stimulate analysis of objective factors in selecting Supreme Court Justices.

4 *Id.* (manuscript at 5).

5 *Id.* (manuscript at 6).

6 Choi and Gulati acknowledge this. *Id.* (manuscript at 30).

7 See, e.g., Gerald N. Rosenberg, *The Supreme Court and the Attitudinal Model*, L. & CTS. NEWSL., Spring 1994, at 6 (reviewing JEFFREY A. SEGAL & HAROLD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL* (1993)) (arguing that the appearance of Justices' voting consistently with their presumed political opinions is best explained by the theory that the President picks Justices whose interpretive philosophy will tend to produce results favored by the President).

## I. THE ISU'S PROPOSALS & REFORMS

The ISU rejected a proposal in 2001 that would have centralized judge selection in the ISU itself.<sup>8</sup> After the pairs controversy, the ISU adopted a new package of solutions, using more judges but having some scores randomly disregarded, and changing from the traditional 6.0 scale to a more complex approach taking into account the degree of difficulty.<sup>9</sup> I will consider these approaches in turn, but first it is worth assessing why we should even be concerned about vote trading by figure skating judges. Economists, after all, generally believe that trades will produce allocative efficiency, and political scientists note that logrolling may be useful where a minority on a particular issue has strong preferences on that issue.<sup>10</sup> Perhaps vote trading in figure skating makes everyone better off.

A technical response might focus on agency costs or externalities. The agency cost argument is that even if judges maximize their joint utility by engaging in trading, they may not be acting on behalf of the broader figure skating community. The externality argument is that logrolling itself imposes a cost on figure

skating fans, who will like the product more if they believe that there is integrity in the judging. After all, figure skating presumably would be less interesting if judges or competitors were allowed to bid on the right to a gold medal.

To be sure, some corruption may be better for the sport and its television ratings than none. Tonya Harding's supporters helped promote figure skating by clubbing Nancy Kerrigan,<sup>11</sup> although that form of cheating has little to do with the judging. More relevantly, the recent pairs controversy increased television ratings.<sup>12</sup> Even if corruption has a positive short-term impact, however, it conceivably could have a corrosive long-term effect.<sup>13</sup> Moreover, corruption in figure skating may attract attention to that sport only at the expense of the image of the Olympics as a whole.<sup>14</sup> In any event, I will consider reforms under the assumption that a process with integrity is indeed the goal.

### A. Centralized Judge Selection

The theory behind centralizing judge selection is that centralization would prevent individual countries from choosing individuals to serve their own interests.<sup>15</sup> The

8 See Selena Roberts, *Judge's Early Tears Indicated Controversy to Come*, N.Y. TIMES, Feb. 17, 2002, § 8, at 1 (discussing the consideration of the earlier proposal).

9 See, e.g., *Pairs Scandal Yields Reform; Point System Among Changes Proposed by Skating Union*, AP, Feb. 19, 2002, available at 2002 WL 6448241; Amy Shipley, *A Small Change for Big Effect: ISU's New Judging System May Discourage Cheaters*, WASH. POST, June 9, 2002, at D6.

10 See, e.g., DENNIS C. MUELLER, PUBLIC CHOICE II at 82-87 (1989).

11 See Richard Sandomir, *No Knee-Whacking, but It's a Ratings Spike*, N.Y. TIMES, Feb. 15, 2002, at 7.

12 *Id.*

13 At least, sports leagues respond as if corruption would have an adverse future effect on the league. See ELIOT ASINOF, EIGHT MEN OUT: THE BLACK SOX AND THE 1919 WORLD SERIES (1988) (discussing the 1919 scandal and its aftermath).

14 Cf. Michael Wilbon, *Go Figure: Skating Lands on Thin Ice*, WASH. POST, Feb. 14, 2002, at D1 (arguing, in the wake of the recent controversy, that "any event in which the outcome is determined exclusively by judging ain't a sport").

15 ISU President Ottavio Cinquanta argued against the proposal on the ground that "even if it is the I.S.U. picking the judges, these judges are coming from nations." Roberts, *supra* note 8. This explanation is unpersuasive, because national federations are likely to pick the most nationalistic judges. Moreover, he disregards that judges may feel they must follow instructions from their federations, which is precisely the alleged source of the current controversy.

problem with this approach is that it merely changes whom the national federations would seek to influence. Suppose, for example, that the ISU board were to choose judges. Then, if the Canadian federation wanted to promote judges who would favor its skaters, it would have an incentive to place individuals sympathetic to it on the ISU board, in the hope that those individuals in turn would pick sympathetic judges, or at least judges who would want to cater to the interests of the board members. (Plus, the ISU board could always give Canadian skaters a duplicate set of gold medals if the judges did not cooperate.)

Centralized selection could be even worse than selection by national federations. The present method at least produces geographical balance. Russia, for example, can pick at most just one judge of the nine. As the recent scandal shows, national federations may get around this problem by engaging in logrolling. Such cheating, however, at least requires some effort and entails some risk of exposure. There would be much less of both if Russia managed to install a sympathetic group of judge selectors. Even if that group were constrained to pick a geographically balanced slate, it might pick that slate in a way favorable to Russia, for example by picking judges who are likely to care so much about later appointments that they favor the Russian skaters. Centralization, in short,

introduces the risk that all judges might be corrupted by a single source. That might be worse than a regime in which each judge is corrupted by a different source.

## B. The ISU's New Scoring System

Immediately after the pairs scandal, ISU President Ottavio Cinquanta promised a reform of the judging process, and he delivered.<sup>16</sup> The reform, however, involves not how judges are selected, but how the judging process itself would function. The first element is the use of fourteen judges rather than nine, with a computer randomly selecting the scores of nine of the judges. I will break this aspect into two: the increase in the number of judges and the random selection of scores. I will then consider the second element of the reform, which replaces the 6.0 scoring scale with a more complex one including a degree of difficulty component.<sup>17</sup>

### 1. Increase in Number of Judges

Perhaps the primary motivation behind increasing the number of judges is to ensure that there would still be a reasonable number even after random selection of a subset of them. But one might argue that increasing the number of judges is a valuable idea in its own right. Such an argument might be based loosely on the Condorcet Jury Theorem,<sup>18</sup> which as its name implies seeks to provide a justification for using juries instead of

16 See Heather A. Dinich, *Officials Present New Scoring System*, WASH. POST, Mar. 26, 2003, at D10.

17 An alternative proposal floated by U.S. skating officials would have left the system intact, but used the median instead of the mean score. See Amy Shipley, *U.S. Skating Unveils Plan for Judging*, WASH. POST, April 26, 2002, at D1. This system would minimize the extent to which a single judge could affect the score but could not prevent bias from influencing the results. Salé-Pelletier would have had a median score of 5.85 and Berezhnaya-Sikharulidze would have had a median score of 5.80 (assuming each judge's technical merit and artistic presentation scores were averaged) in the 2002 contest.

18 For discussions, see Bernard Grofman, *Judgmental Competence of Individuals & Groups in a Dichotomous Choice Situation: Is a Majority of Heads Better Than One?*, 6 J. MATH. SOC. 47, 48-52 (1978); Saul Levmore, *Conjunction & Aggregation*, 91 MICH. L. REV. 723, 734-39 (2001); and Nicholas R. Miller, *Information, Electorates, & Democracy: Some Extensions & Interpretations of the Condorcet Jury Theorem*, in INFORMATION POOLING & GROUP DECISION MAKING 173, 175-77 (Bernard Grofman & Guillermo Owen eds., 1986).

individuals to resolve cases.<sup>19</sup> The central idea underlying the theorem is that as long as any particular decisionmaker is more likely than not to obtain the right answer to a problem, then a majority vote among a group of such decisionmakers is more likely to produce a correct answer, with the likelihood of a correct answer increasing with increasing group size. As Saul Levmore has pointed out, this is a familiar phenomenon to viewers of *Who Wants to Be a Millionaire*, where the audience usually provides the right answer even to relatively hard questions.<sup>20</sup> And so too in figure skating, if any given judge is likely to make a proper comparison between any two skaters, considering the decisions of a number of judges is likely to improve accuracy.

The problem with applying this argument to figure skating is that the problem it addresses is one of accuracy rather than one of corruption. Suppose that there is some percentage of judges who can be bought and that this percentage is not affected by the number of judges on a panel. Having four corruptible judges on a hypothetical 18-judge panel would seem no better than having two such judges on a nine-judge panel, as both corrupt groups seem to have an approximately equal chance of swaying a close outcome. Perhaps corruption is more likely to make a difference in cases when the judges who are not corrupt make a strange decision, and so improving accuracy by increasing the validity of the noncorrupt judges' assessments may decrease the number of cases in which corruption affects the results. The effect, however, does not seem likely to make a substantial difference.

A counterargument might be that the percentage of judges who are corruptible would decrease with larger panels, because the cost of payoffs might go up. It might be just as expensive (whether in dollars or in vote trades) to corrupt a single judge on a 14-judge panel as it would be to corrupt a judge on a nine-judge panel, because each judge may be willing to take the risks associated with bribes only for a certain minimum amount. This is possible, but larger panels might well reduce the size of the necessary bribe. Larger panels bring less scrutiny to each individual judge, and so corruption might be easier. Which of these effects is larger is an empirical question, but there is little reason to think that any improvement would be large.

## 2. Random Selection of a Subset of Judges

To assess the rationale for randomly selecting judges from a larger group, let us consider Cinquanta's own explanation:

Suppose that I want to ask the judge to help my skater. OK, I go there, I say, 'Here's a beautiful month in Honolulu for you and your girlfriend.' To whom do I say this? First, I don't know if this judge I'm sending to Honolulu is one voting. Secondly, this judge can go to Honolulu with a nice girl, I pay for his expenses, and then he doesn't vote for me because nobody knows who has voted.<sup>21</sup>

The first of these reasons ignores that the success of any corruption will be probabilistic in any event, for even if one knows who will be voting, one does not know whether the judge will have a decisive vote and thus cannot ensure a victory by bribing a small number of judges. The number of judges that one will have to bribe to *ensure* a favorable result will be

19 I say "seeks" because the stylized jury of the Condorcet Jury Theorem is one that engages in no deliberation and resolves decisions by majority vote. For a defense of unanimity rules for jury decisionmaking, see Richard A. Primus, *When Democracy Is Not Self-Government: Toward a Defense of the Unanimity Rule for Criminal Juries*, 18 CARDOZO L. REV. 1417 (1997).

20 See Levmore, *supra* note 18, at 734 n.22.

21 Paul Oberjeruge, *Move Made to Settle the Score*, L.A. DAILY NEWS, Feb. 19, 2002, at S5.



greater when some judges' scores are disregarded, but in practice perfect assurance is likely to be impossible anyway, because even bribed judges are unlikely to vote for, to take an extreme example, a skater who falls repeatedly and fails to complete a program.

The second of Cinquanta's reasons may initially seem more compelling, that it would allow judges who have agreed to cheat the freedom to follow pangs of conscience. Of course, anonymity could be achieved without randomly eliminating some scores, but if the scores are simply scrambled a judge might face the danger that even the most favorable rating to the party the judge wishes to show allegiance might not be favorable enough. While anonymity may facilitate reneging on under-the-table deals, it will also provide less public accountability for the judges' votes.<sup>22</sup> Thus, a judge who does enter a seemingly implausible result faces less of a danger of public scrutiny. A judge might still cheat because achieving the desired result may make the briber marginally more likely to believe that the judge has kept the promise and no one will be able to attack the judge's decision directly. Once again, the proposal thus produces competing effects, and it is hard to evaluate whether honesty would be increased or reduced.

Even if anonymity has benefits, it comes at a high price. First, fewer judges will be scoring skaters, so to the extent accuracy in judging is a concern, the Condorcet Jury Theorem argument discussed above cuts in reverse. Second, random selection emphasizes the arbitrariness of judging. It may be no less

accurate to have nine judges of fourteen evaluating skaters than to have nine of nine, but in close cases, it will be readily apparent that different winners likely would have been selected if different random selections had been made. The randomness of computer selection among judges may depreciate the gold medal more than a random process for picking judges before the competition, because computer-driven arbitrariness is likely to be more cognitively salient than any arbitrariness in picking judges. The problem will be especially severe if the other judges' scores are eventually released, since we then may end up sure that a given result was just a result of a quirky selection. Meanwhile, arbitrariness will seem especially great if different judges are used for different skaters, as some competitors might just have the bad luck of drawing those with relatively high standards. These are, to be sure, arguments about appearances only. Holding constant the number of judges whose scores will be used, it does not substantively matter how many other judges have voted and had their scores ignored. In figure skating, though, there is little at stake in the accuracy of judging besides appearances, and random selection of judges hurts appearances without clear benefit in improving integrity.

### 3. Scoring System Changes

Perhaps the most sensible change in policy is the adoption of a more complicated scoring system, in which each technical element has a value, with more difficult elements having greater potential value than less difficult ones. The technical scores would be added up, along

<sup>22</sup> George Rossano offers a similar criticism: "If I am concerned that officials in a basketball game are making terrible calls and may be on the take from team owners, the solution is to let the officials randomly work half of each game and let them make all their decisions during that time in secrecy?" George Rossano, *Do the Math to Improve Judging*, N.Y. TIMES, May 19, 2002, § 8 (Week in Review), at 11; see also Vicki Michaelis, *New Judging for Skating Still Leaves Room for Political Bias*, USA TODAY, June 10, 2002, at C7 ("Though federation officials won't be sure their judges' votes will count, they still could make deals and pressure their judges to vote a certain way. The new system just means they would be playing longer odds whether the pressure has any effect.").

with evaluations for one or more categories akin to the present evaluation of artistic merit. Commentators have noted that such a system is more objective,<sup>23</sup> though the real difference is that the subjective decisions about the relative value of different aspects of a performance are made in advance and publicly. This makes it more difficult for a judge to justify a seemingly odd ranking by pointing to one aspect of a performance and ascribing undue weight to it.<sup>24</sup> Of course, judges still might be biased in evaluating each of the individual categories, especially those that are not susceptible of purely objective analysis, but it is quite possible that the reputational costs of an equivalent unjustified change in the ultimate score would be greater with more specific criteria.

Elimination of the traditional scoring system is not without sentimental costs, however. As a newspaper commentator noted, “[G]one, too, would be the wondrous score of 6.0, figure skating’s mark of perfection.”<sup>25</sup> Perhaps one reason people watch sports is because of the possibility of witnessing something that is objectively extraordinary – a perfect game in baseball, a perfect score in figure skating – and being able to tell their friends and grandchildren that they were there when it happened, as if their presence means that they deserve some of the credit for the achievement. Moreover, the 6.0 system is

accessible, allowing even the casual viewer of skating to form a tentative opinion about a performance before the official scores are displayed. The new system eliminates this transparency, and while that may well be an underlying goal of the reform, it is not a noble one.

## II. AN ALTERNATIVE APPROACH

My alternative approach for selecting judges is quite simple. Judges at any competition would be evaluated by how closely their scores correlated with those of other judges. Though there are a number of ways of implementing such a system, a straightforward one would be simply to calculate the correlation of each judge’s scores with the average of all other judges’ at a particular competition. Judges who performed well would be promoted; those who performed poorly, demoted. At minor competitions, judges would have an incentive to perform well in order to increase their chances of judging more significant competitions, and at the Olympics, high quality performance would be necessary for continued high-level judging or perhaps for promotion to important positions within the ISU.<sup>26</sup>

Wouldn’t this system be circular? Each judge interested in promotion, after all, would judge on the basis of expectations about what the other judges would do. A judge who cares

23 See, e.g., Allison Sherry, *Skate-Judging Overhaul Urged*, DENVER POST, Feb. 19, 2002, at A1.

24 The 6.0 system is particularly problematic for two reasons: First, it does not allow for as many gradations as might be desirable, since scores are given in tenth of a point increments. Second, when the total scores for two skaters or teams are equivalent, judges simply break the tie. Thus, a judge who favored the Russian skaters could give them a higher mark than the Canadians by one tenth on artistic presentation, a lower mark by one tenth on technical merit, and then declare that the Russians were better overall. Thus, the ultimate decision often may depend on an assessment for which no scores need be given.

25 Abby Haight, *Skating Official Proposes Reforms*, PORTLAND OREGONIAN, Feb. 19, 2002, at A1.

26 The allegedly corrupt French judge in the recent pairs competition was said to be influenced by her desire for an ISU post. See, e.g., Christie Blatchford, *French Judge Now Blames Canadians*, NAT’L POST, Feb. 19, 2002, available at 2002 WL 11861084 (“Skating insiders charge that Le Gougne allegedly succumbed to the pressure because of her own ambitions to win a spot on the ISU’s technical committee in June and need for support from the influential Russians.”).

only about retaining power would give a skater who had performed poorly high marks if the judge expected other judges to do so. But there is no reason to think that the other judges would be more likely to rate bad skaters above good ones. With everyone trying to guess what everyone else thinks, the best strategy would be to grade skaters on the merits.

This logic has roots in game theory. In *The Strategy of Conflict*, Thomas C. Schelling conducted a poll in which he asked people what they would do if they had to meet up with someone the next day in New York but could not contact the person in advance.<sup>27</sup> Large numbers answered that they would go to the information booth at Grand Central Terminal at noon.<sup>28</sup> This is not because of any particular affinity for that tranquil spot. In game theory lingo, the information booth is a “focal point” for people trying to find each other. A visitor trying to figure out where the other visitor will go, knowing that the other visitor will be reasoning in exactly the same way, will seek out a focal point rather than pick a random spot in Manhattan.

Judges in figure skating seeking to maximize the correlation of their ratings with those of other judges face essentially the same problem as the New York visitor. The situation might seem to be different because the judges are competing with one another rather than cooperating, but that will not change the incentives, because there are more than two judges. In effect, each judge is seeking to cooperate with a hypothetical judge who represents the average of all the other judges. The focal point is almost certain to be the scores (or the relationship between the scores, if correlation only is measured) that actually correspond to any particular figure skater’s performance. While one might be able to think of other focal points – prime numbers

or general reputation or projected future success – there is no particular reason to think that other judges are likely to find any one of these particularly focal. Moreover, there is a very strong reason to think that actual performance will be a focal point, because that of course is what judges are supposed to be considering. Thus, a unique and appropriate focal point is even more likely to emerge than in Schelling’s hypothetical involving New York, where one visitor might well end up at the top of the Empire State Building.

Let me consider several arguments against this approach. First, judges might have an incentive to disregard small mistakes, because other judges might not notice them. If all judges could be counted on to make decisions based on their true assessments of performance, a skater would be hurt even by a small mistake that only one judge caught, because that judge’s penalty would affect the average of all the scores. This is an important caveat, but at least judges will have an incentive to ignore a mistake only if they believe that the majority of other judges will not notice it. Mistakes that are sufficiently prominent that they are generally noticed will be sufficiently prominent to affect the focal point. In any event, if a mistake is so subtle that most expert judges do not see it, the corresponding deductions probably would not be so big as to make a difference in most cases.

Second, a judge who expects others to be corrupt will take that into account in his own scoring. For example, if one judge tells another that she will favor the Russians, the second judge has an incentive to move his score in that direction. Such mimicking, however, will not make a difference in the mimicking judge’s ordinal rankings (except perhaps to break a tie), because as long as only a minority of judges are viewed as

27 See THOMAS C. SCHELLING, *THE STRATEGY OF CONFLICT* 54-57 (1980).

28 *Id.* at 56.



corrupt, the incentive will be to move only a bit in the direction of corruption. Of course, a corrupt judge's score could make a difference, but individual judges will at least have less incentive to be corrupt because they are not beholden to their national federations, and the system should screen out those who have been corrupt in lower-level competitions. No system is likely to be wholly successful in eliminating the effect of corruption that occurs, but the aim of the correlation approach is to reduce the likelihood that judges will want to act corruptly.

Third, the geographical balance of the judging panel might affect the focal point. For example, if four of seven judges are Russians, then judges might have an incentive to favor a Russian style of skating or even a Russian skater. If judges are selected objectively and fairly, however, there is no reason to think that the nationalities of the other judges should matter, and thus nationalities are unlikely to affect the focal point. But there are easy solutions in any case. Judges might be selected in part based on diversity criteria, to prevent any one country or region from seeming to have too great a presence on the judging staff. In addition, judges might be evaluated in their deviation from all other judges not of the same nationality as the current skater. Thus, if there were a Russian skater and a Russian judge, that judge would vote, but a Canadian judge's score would not depend on the Russian judge's. An easy way to implement this would be, instead of using correlation coefficients, to calculate the average absolute deviation from a judge's scores with that of the other relevant judges.


While I have focused on figure skating, the approach that I have described has other potential applications. Suppose a school wishes to reward teachers who devise effective exams and grade them carefully. An easy way to do this, at least if there is a blind grading system, would be to measure the correlation of any particular teacher's exams with those of others. The system is not perfect – perhaps performance in one class is less likely to be correlated with grades than performance in another – but it may succeed, especially over years, in identifying those who are particularly careful or careless. In any event, the analysis here suggests that rewarding teachers for being like their colleagues is more likely to produce a race to the top than one to the bottom.

More adventurously, the approach might be used as a way of providing expert witnesses incentives to be truthful.<sup>29</sup> The incentives of expert witnesses are roughly as bad as those facing figure skating judges. If expert witnesses developed reports independently and then were compensated in part based on whether they agreed with other experts (perhaps the other side's expert and an unknown court-appointed expert), their incentives might improve. Instead of insisting that removing financial incentives is necessary to obtain unbiased testimony, this proposal would harness financial incentives to induce such testimony.

The approach, however, is not easily transferable to judging in courts. It would, though, help solve two potential problems

29 Robert Cooter and Winand Emons have recently offered a model suggesting that a strict liability perjury rule would improve witnesses' incentives to be honest. See Robert Cooter & Winand Emons, Truth-Revealing Mechanisms for Courts (2002) (unpublished manuscript, on file with author). They apply their model on the assumption that a court eventually finds out whether a witness was being truthful. *Id.* at 10. They note, however, that the analysis could be extended to cases in which the court has only probabilistic information about truthfulness. Cooter and Emons, however, do not explain where the court obtains such information.

with Choi and Gulati's proposal: first, that the proxies determining who will become a Supreme Court Justice might be manipulated,<sup>30</sup> and second, that the system cannot easily measure whether judges reach sensible decisions.<sup>31</sup> The challenge in applying the proposal to courts would be to create a system in which different judges decided cases simultaneously without contacting one another. Perhaps a random sample of cases could be assigned to multiple judges or panels, but

ratings would be a reliable measure of judicial performance only if judges could be kept from knowing which cases were the ones that would determine their ratings. In the United States, the complications associated with creating such a system are probably not worth it. But a modified version of the proposal, such as a system in which judicial compensation depends on results, might well be worth considering in any judicial systems as rife with corruption as the ISU. 

---

<sup>30</sup> Lawrence Solum has suggested that "there will be endless opportunities for gaming the citation count measure." Lawrence Solum, *A Tournament for Judges. Mad? Brilliant? Clever?*, Legal Theory Blog, Apr. 17, 2003), available at [http://lsolum.blogspot.com/2003\\_04\\_01\\_lsolum\\_archive.html](http://lsolum.blogspot.com/2003_04_01_lsolum_archive.html) (last visited Apr. 18, 2003).

<sup>31</sup> Choi and Gulati do suggest considering reversal rates. See Choi & Gulati, *supra* note 3 (manuscript at 11). A problem with this is that it would tend to entrench the views of the existing members of the Supreme Court. Choi and Gulati also suggest considering reversal rates by courts sitting en banc. See *id.* (manuscript at 6). A circuit court as a whole, however, may be quite liberal or quite conservative, so this would reward judges not for acting moderately, but for mimicking the political biases of their circuits. A possible solution would be a system in which three-judge panels were reviewed en banc by a random selection of judges from other circuits. See Michael Abramowicz, *En Banc Revisited*, 100 COLUM. L. REV. 1600 (2000).