Andrew Duong
20772131

Bias in the 2008 Beijing Olympics
(Gymnastics)

The Olympic Games are designed to promote peace and universal moral principles all over the world. In order to promote these ideals, it is necessary for the Games to display a high level of integrity in upholding these ideals. The performance of many athletes of the Games can be measured objectively, such as the finish time of a sprinter or the accuracy of an archer. However, many sporting events in the Olympics involve varying degrees of subjectivity, which in some cases may determine the ultimate results of the competitions. Among other Olympic sports, gymnastics is heavily influenced by the subjective assessment of athletes by the judges, which naturally raises the question of whether nationalism influences the outcomes of these events.

In the 2008 Games, a new scoring system was introduced in Gymnastics which removed the upper limit of an athlete's score from a 'perfect 10'. The new system instead introduced an open-ended scoring system. Two judges mark the technical aspect of a routine, which is essentially limitless. The technical aspect rates the degree of difficulty of a routine and encourages competitors to perform increasingly difficult skills to earn extra points. The bulk of the subjectivity in gymnastics is introduced through another six judges who mark the execution of a routine starting from 10 and deducting points for errors in the execution of a routine. Of the six execution scores, the highest and lowest scores are removed in an attempt to eliminate some of the bias, and the average of the remaining four scores is taken as the final execution score. The final score of a routine is the sum of the technical and execution scores less any penalties.

Total Score = Technical Score + Execution Score – Penalties

Further attempts to remove bias from the scoring involves selecting judges from countries different from one another, but does not necessarily avoid the judges' nationalities coinciding with the athletes' nationality during the heats. However, such a conflict of interest is prohibited in the finals of the competition because any bias will significantly influence the final medal results. Additionally, if any particular judge scores an athlete by more than an acceptable margin, a head judge has the discretion to advise the judges to conference and request the deviating judge to change his/her score.

Despite these safeguards, it is important to remove bias from all the judges as it may potentially influence the final score after the highest and lowest scores are removed. In addition, in the event of

a tie under the normal scoring system, the four median execution scores of the six judges is expanded to include the score of all six execution judges. This rare occurrence was witness during the finals of the women uneven bar event in the 2008 Games between Nastia Liukin from USA and Kexin He from China, and highlights the importance of removing bias from all the judges. An explanation of the scorecard of this incident can be found in Appendix 1.

The analysis of bias in the 2008 Olympic scoring in gymnastics is based on data from five men events: the floor, pommel horse, rings, parallel bars and horizontal bar; and three women events: the floor, uneven bars and beam. The scores from each judge were taken from the rotations during the heats since there were cases where the nationality of the judge and athlete coincided. The data includes 623 scores from the selected events against 25 judges, in 62 of these cases the nationality of the judge was the same as that of the gymnast. Only the execution scores are analysed for bias, as the technical score is much more objective.

On a simple level, if there was no bias, judges would be expected to award gymnasts from their own nation the highest score approximately one sixth of the time. Since there are 62 observations, this equates to approximately 10 times. However, the data shows that this occurred 30 times which is approximately half of all the observations. Similarly by the same argument, the lowest score is expected have been awarded to gymnasts from their own nation approximately 10 times, but occurred only 3 times. These results certainly provide an early indication that bias may exist among some of the judges at the 2008 games.

The model used to provide a numerical value for the bias displayed by each judge against each gymnast was built based on ideas from a model used by Eric Zitzewitz in his paper titled "Nationalism in Winter Sports and Its Lessons for Organizational Decision Making". The model identifies that each judge is unique and variances in the scores are not only a result of bias, but are due to the varying degrees of leniency of each judge. Hence each execution score awarded by a judge can be broken down as follows:

Judge's Execution Score = Objective score + Judge's Leniency + Bias

The objective score of an athlete is difficult to measure, if this was possible then there would be no need for the subjective assessment by judges. The model uses the average score awarded by the other 5 judges not in question as a proxy for the objective score of a gymnast's performance. This calculation assumes that the average leniency and bias of the other judges is zero.

The leniency of a judge may vary slightly from performance to performance, so for the purposes of this model, each judge's leniency was measured as the average difference of the judge's official score minus the objective score measured previously among all athletes whose nationality is not in question. Another important assumption when making this calculation is that the average bias among all athletes whose nationality is not in question is zero.

Each judge's bias for each performance is quite simply the residual term once the other components of the equation have been calculated. Judges who did not judge athletes from their own nation displayed an average bias very close to zero in comparison to the judges where there was a conflict of interest. As a consequence of this result, relative confidence can be gained from using this model to measure bias at an individual level.

TABLE 1.

The Measured Bias of Judges from the Men Gymnastic Events at the Beijing 2008 Olympic Games.

Country	N	Bias	P-Value
BEL	1	0.38	N/A
COL	1	0.36	N/A
KOR	5	0.25	0.037
ROU	5	0.24	0.000
VEN	1	0.22	N/A
UKR	1	0.18	N/A
JPN	5	0.18	0.009
USA	5	0.18	0.001
CAN	5	0.13	0.057
ESP	5	0.13	0.000
GER	5	0.10	0.000
SUI	1	0.08	N/A
GBR	2	0.02	0.401
LUX	1	-0.06	N/A
EGY	1	-0.30	N/A

Table 2.
The Measured Bias of Judges from the Women Gymnastic Events at the Beijing 2008 Olympic Games.

Country	N	Bias	P-Value
KOR	1	0.92	N/A
ESP	2	0.49	0.000
BUL	1	0.43	N/A
BLR	1	0.42	N/A
GER	5	0.37	0.009
GBR	5	0.3	0.001
NED	1	0.25	N/A
COL	1	-0.12	N/A
POL	1	-0.26	N/A

After implementing the model, the results show there is a statistically significant level of bias among many of the judges. Some results were based on only one observation of a judge judging their own nation, which made it difficult to assess the significance of the measured bias. Despite the lack of data from some judges, all the judges except for Luxemburg and Egypt in the men events and Colombia and Poland in the women events displayed positive bias, as it can be seen from Table 1 and Table 2. The judges who displayed negative bias were based on only one observation, which would suggest that the observed negative bias is not statistically significant.

To test the significance of the results for judges with more than 1 observation, the null hypothesis assumes that the bias is zero. Among the judges who judged their own nation 5 times, the p-value was always under 0.05, except in the case of the Canadian judge who obtained a p-value of 0.057, which can still be considered fairly significant.

The Korean judge in Table 1 displayed a bias of 0.25. An interpretation of this result suggests that on average the Korean judge scored Korean gymnasts 0.25 points higher than he would for the exact execution of a performance by a gymnast from a country that the judge felt indifferently towards. In a competition where the final scores of athletes are separated by as little as one hundredth of a point, such a display of bias raises questions about the integrity of the judges and injustice towards athletes of other countries, in particular to those who are not represented on the judging panel.

Many interesting links and ideas can be drawn from the result generated by the model. One of these is an analysis of the correlation between the biases of a judge from a particular nation with that of

the nation's level of corruption. For the purposes of this analysis, corruption is measured by the Corruption Perception Index[1] (CPI), compiled by Transparency International, and rates the "degree to which corruption is perceived to exist among public officials and politicians" (Corruption Perception Index, 2008) on a scale of zero to ten, where ten indicates little to no corruption perceived and zero indicates extreme levels of corruption.

Due to the lack of significance and reliability of the calculated bias from judges with only one instance of judging their own country, these judges were not included in the analysis of the correlation (Appendix 2). Among the other judges, the correlation was found to be -0.37 which indicates a moderate relationship between the two variables, which can be seen in Appendix 3. This finding is similar to what Zitzewitz found in his analysis of judging in winter sports, where the correlation between the measured bias and the CPI was much stronger in ski jumping (-0.59) and similar for figure skating (-0.38) . Although correlation does not imply causation, it does suggest that some of the bias may be explained by the corruption levels of the judges' nations. However, the CPI of judging nations was not a factor used to discriminate judges, as the CPI of judges in the recent games ranged widely from two to nine.

It is possible to obtain a bias-free score for each athlete by using the model. The bias-free score is essentially the objective score plus the leniency of each judge. Leniency is not removed because it will always be evident in a subjective assessment of any performance and cannot be controlled. So analysing the bias-free score provides the most relevant measure of performance that can be realistically achieved.

As would be expected, the bias-free score exhibits that the rankings of the gymnasts after each heat rotation are, in many cases, different from the official ranking. The majority of gymnasts moved up or down 1 or 2 positions from what they were officially ranked. However, in an extreme case, a gymnast was found to have advanced by 10 positions in the absence of bias, while another gymnast would have resultantly been 14 positions lower in the absence of bias. Such cases of large movements certainly highlights the dramatic affect bias can have on the position of a gymnast, and thus their chances of obtaining a spot in the finals.

In addition to analysing the nationalistic bias displayed by the judges, it is possible to use the model to analyse any bias from judges towards other nationalities. The model is implemented in much the same way as when analysing each judge's nationalistic bias, except that the nationality of the competing nation modified when in question. The analysis was performed for all judges on nations

---

[1] Corruption Perception Index for 2007 was used.

who were represented by 5 gymnasts in each event, which provided a sufficient amount of data. The

result of this analysis is displayed in Table 3.

Table 3.
Cross Comparison of Bias of each Judge by Athlete's Nation.

		Judging Nation											
		BEL	ROU	NZL	KOR	BUL	GRE	VEN	EGY	SLO	TPE	ISR	SUI	GER
Athlete Nation	BLR	-0.144	0.096	-0.120	0.072	0.000	0.076	-0.092	-0.044	-0.020	0.124	0.020	0.050	-0.100
	CAN	-0.020	-0.116	0.100	0.004	0.124	0.092	0.044	-0.220	0.068	-0.052	-0.016	-0.016	0.008
	CHN	0.052	0.052	-0.044	-0.020	0.076	-0.140	-0.110	0.070	-0.110	0.070	0.015	-0.015	0.015
	ESP	0.020	0.044	0.044	-0.124	0.068	0.216	-0.216	-0.336	-0.072	0.192	-0.016	-0.136	-0.016
	FRA	-0.004	0.188	-0.124	0.020	0.020	0.164	0.140	-0.220	-0.028	-0.028	-0.136	0.032	0.056
	GER	0.080	0.080	0.032	-0.064	-0.016	0.000	-0.024	-0.120	0.048	0.048	-0.044	0.052	0.100
	ITA	-0.068	0.220	0.100	-0.284	0.052	0.108	-0.252	-0.156	0.132	-0.060	-0.130	0.200	0.050
	JPN	0.004	0.052	0.004	-0.116	0.052	-0.040	-0.016	0.080	-0.088	0.152	0.025	-0.065	0.085
	KOR	-0.108	-0.036	-0.060	0.252	0.012	-0.092	-0.092	0.076	-0.068	0.028	0.056	0.104	0.032
	ROU	0.148	0.244	-0.236	-0.164	0.052	0.168	-0.144	0.072	-0.048	-0.120	0.128	-0.064	0.056
	RUS	-0.132	0.084	-0.060	0.060	0.036	0.060	-0.204	-0.012	0.036	-0.060	0.128	0.008	-0.136
	USA	-0.064	0.032	0.056	-0.088	0.104	0.024	0.072	0.024	0.072	-0.144	0.112	-0.128	0.016

		Judging Nation											
		ESP	CAN	NED	GBR	USA	ALG	UKR	ARG	COL	JPN	AUT	LUX
Athlete Nation	BLR	0.080	-0.040	-0.145	-0.025	0.095	0.005	0.035	-0.016	-0.016	0.080	-0.016	-0.112
	CAN	-0.040	0.128	-0.104	-0.080	-0.080	0.040	0.040	-0.120	0.090	-0.090	0.060	-0.060
	CHN	-0.105	0.135	-0.120	0.024	0.024	0.000	-0.072	0.252	0.060	-0.036	-0.156	-0.060
	ESP	0.128	0.032	0.096	-0.096	0.000	0.000	0.072	0.016	0.136	0.064	-0.152	-0.104
	FRA	-0.040	-0.040	-0.076	0.020	-0.004	-0.004	-0.028	0.028	-0.020	0.172	-0.092	0.172
	GER	-0.068	0.004	0.080	0.056	-0.112	-0.040	-0.040	0.016	-0.008	0.064	-0.104	0.016
	ITA	0.020	-0.190	0.068	-0.004	-0.004	-0.052	0.068	0.096	0.240	0.072	-0.072	-0.144
	JPN	-0.005	0.025	-0.132	0.156	-0.060	0.084	-0.108	0.008	-0.016	0.176	0.008	-0.112
	KOR	-0.064	-0.040	-0.024	0.048	0.000	0.072	-0.192	-0.064	-0.040	0.128	-0.040	-0.016
	ROU	-0.064	0.056	0.000	-0.072	0.024	0.144	-0.024	0.020	0.308	-0.220	0.020	-0.004
	RUS	-0.016	0.032	-0.052	-0.076	-0.172	0.044	0.140	0.044	0.092	-0.076	-0.148	0.116
	USA	-0.128	0.160	0.008	0.008	0.176	-0.088	-0.112	0.012	0.084	-0.036	-0.036	0.012

Notes: Blue bold numbers highlight positive bias greater than 0.2. Red bold numbers highlight negative bias more than -0.2. Black bold highlight nations with tensions/disputes.
- China considers Taiwan as a renegade province.
- China claims the Japanese-administered Senkaku-shoto (Senkaku Islands/Diaoyu Tai).
- Algeria fought a war of independence against France.

A possible source of positive bias may come from countries with strong ties, such as when the

judging and competing nations are a part of the same trading bloc or are regional neighbours. On

the other hand, negative bias may come from countries where there are ongoing conflicts or disputes over various issues, such as land rights. However, looking at Table 3 no clear pattern of significant bias emerges within these groups.

The implications of this finding suggest that the motivation for judges to introduce bias into the scores of other nations other than their own is not as strong as the motivation to positively score the judge's own nation. Nationalism seems to be the strongest driver of bias, as judges do not seem to be influenced by tensions or relations when judging gymnasts from different countries. Despite this, significant bias still exists in some cases which have been highlighted in Table 3. There are many potential explanations for these biases, which ultimately come down to factors which affect each individual judge. For example, the Korean judge's large negative bias towards Japan, Spain, Italy and Romania may be an attempt to lower the Korean competitors' scores to give Korean gymnasts a better chance of ranking higher. When comparing the bias towards Italian and Korean gymnasts given by the Korean judge, the effect of the bias gave Korean gymnasts more than a 0.5 point advantage over the Italian gymnasts. This creates a major disadvantage to the Italian gymnasts relative to the Korean gymnasts.

The analysis raises serious question as to the integrity of the scoring system used in gymnastics. The main problem is that there is no alternative that will provide a better objective measure of the performance of a routine than by relying on the subjective opinion of trained judges. An obvious suggestion would be to select judges from nationalities different from those of the competitors to reduce the main source of bias, but this makes it difficult to find quality judges which tend to come from countries who are strong performers in the sport and thus also competing in the event. Choosing judges from nations with a high CPI is another suggestion, but would discriminate and goes against some of the moral principles promoted by the Olympic Games.

The sport of Olympic gymnastics will always involve a certain element of bias among the judges, which can at times be highly controversial because of its potential influence on the results of the competition. The only possible method to create a fairer competition is to create incentives for judges to be less motivated to introduce nationalistic bias into their scores. However, this is an extremely difficult task because the Olympics Games is considered as the pinnacle of many sporting events, including gymnastics, which brings out nationalistic pride and a natural instinct to support one's own nation, even while on the judging panel.

Appendix.

Appendix 1.
Score cards for the Women Uneven Bar Finals at the 2008 Beijing Olympics.

Normal Bias removing system:

Rank	Country	Name	A Score	B1 AUS	B2 POL	B3 NZL	B4 RSA	B5 BRA	B6 BUL	Total Score
1	CHN	Kexin HE	7.7	9.3	9.1	9.1	8.9	9	8.9	16.725
2	USA	LIUKIN Nastia	7.7	9	9.3	9	8.8	9	9.1	16.725

Note: For He, the scores of judges B2, B3, B5 and B6 were included. For Liukin, the scores of judges B1, B3, B5 and B6 were included. This resulted in exactly the same Total Score.

Tie resolving system:

Rank	Country	Name	A Score	B1 AUS	B2 POL	B3 NZL	B4 RSA	B5 BRA	B6 BUL	Total Score
1	CHN	Kexin HE	7.7	9.3	9.1	9.1	8.9	9	8.9	16.750
2	USA	LIUKIN Nastia	7.7	9	9.3	9	8.8	9	9.1	16.733

Note: The average of all judges, B1 to B6 were all included for the calculation of the execution scores. He scored higher overall to win the gold medal.
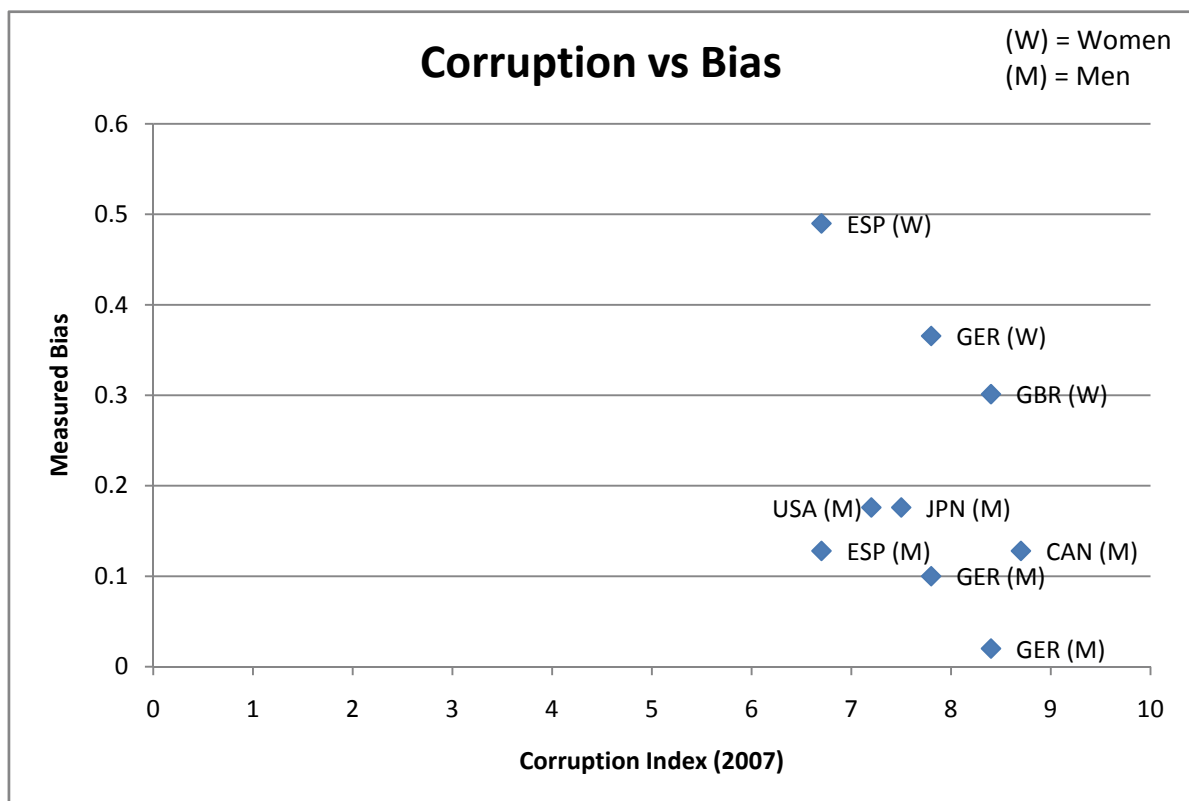
Appendix 2.

Table of Data used for the Correlation between CPI and Measured Bias.

Country	N	Bias	Corruption Perception Index (2007)
ESP	2	0.49	6.7
GER	5	0.37	7.8
GBR	5	0.30	8.4
JPN	5	0.18	7.5
USA	5	0.18	7.2
CAN	5	0.13	8.7
ESP	5	0.13	6.7
GER	5	0.10	7.8
GBR	2	0.02	8.4

Appendix 3.

Scatter plot of Corruption Index (2007) against measured bias in Men and Women events.

References.

Corruption Perception Index. (2008, November 29). Retrieved December 6, 2008, from Wikipedia: http://en.wikipedia.org/wiki/Corruption_Perceptions_Index

Ellenberg, J. (2008, August 12). *Down With the Perfect 10!* Retrieved December 6, 2008, from Slate: http://www.slate.com/id/2197336/pagenum/all/

Faucette, J. (2008, August 18). *2008 Olympic Women's Gymnastics Scoring*. Retrieved December 6, 2008, from suite101.com: http://gymnastics.suite101.com/article.cfm/2008_olympic_womens_gymnastics_scoring

Internat ional Olympic Committee. (n.d.). *OLYMPIC PROGRAMME COMMISSION.* Retrieved December 6, 2008, from http://multimedia.olympic.org/pdf/en_report_953.pdf

NBC. (2008). *Results and Schedules*. Retrieved December 6, 2008, from 2008 Beijing Summer Olympics - Gymnastics: http://www.nbcolympics.com/gymnastics/resultsandschedules/index.html

Popović, R. (2003). International Bias Detected in Judging Rhythmic Gymnastics Competition at Sydney-2000 Olympic Games. *Physical Education and Sport* , 1-13.

Scoring System. (2008). Retrieved December 6, 2008, from NBCOlympics.com: http://www.nbcolympics.com/gymnastics/insidethissport/scoring/newsid=116807.html

Theatres of conflicts around the world. (n.d.). Retrieved December 6, 2008, from Did you know?: http://www.didyouknow.cd/story/conflicts.htm

Whissell, R., Lyons, S., Wilkinson, D., & Whissell, C. (1993). National Bias in Judgements of Olympic-Level Skating. *Perceptual and Motor Skills* , 355-358.

Zitzewitz, E. (2006). Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making. *Journal of Economics & Management Strategy* , 67-99.