

Scoring reversals: a major league dilemma

Jean-Pierre Benoit*

Department of Economics, and School of Law, New York University, New York, NY 10003, USA

Received June 28, 1991 / Accepted December 24, 1991

Abstract. Recent work shows that, *in theory* scoring methods are very sensitive to the scoring points used. I analyze the results of Most Valuable Player voting in baseball as a test of the practical relevance of this work. I find that Most Valuable Player rankings are indeed quite dependent on the particular scoring points in use.

Perhaps the most common method for ranking a set of alternatives is the plurality method – each individual votes for a single choice and the alternatives are ranked by the number of votes received. Borda (1781) argued that this method is extreme in that it considers only how often an alternative is ranked at the top by some individual, paying no heed to how well the alternative fares when it is not placed first. He proposed that each individual assign points to the alternatives – one to her least preferred alternative, two to her second least preferred, three to the next one, and so forth – and that the alternatives then be ranked by the total number of points received. This method is now known as the Borda Count.

Although Borda suggested that points with constant differences be used, one may question this particular configuration. With four alternatives, why not assign points of 1, 2, 3, and 5, for instance? There is a certain arbitrariness inherent in a Borda Count¹. However, this is not an arbitrariness introduced by Borda. After all, the common plurality method “arbitrarily” assigns weights of one to the top choice, and zero to all others.

Both the Borda count and plurality voting are special instances of *scoring methods*. With n alternatives to be ranked, a scoring method has each individual

* The author thanks Bill Deane of the National Baseball Hall of Fame for supplying voting data, and Gary Cohen and Andy Schotter for discussions. Technical support from the G. V. Storr Center for Applied Economics is gratefully acknowledged.

¹ Recently, some authors have found intriguing advantages to the Borda Count. See, for instance, Saari (1990).

assign points (from top to bottom) of w_1, w_2, \dots, w_n to the alternatives, where $w_1 \geq w_2 \geq \dots \geq w_n$, and $w_1 > w_n$. The points are totaled and the alternatives are ranked from most points received to least.

Given a set of individuals voting (sincerely) according to their preferences, we can expect plurality voting and the Borda Count to result in different rankings of the alternatives. Indeed, there would be little point in introducing a Borda count if this were not the case. The basic appeal of a Borda Count, however, is that it gives weight to preferences beyond each individual's top choice, and not that it uses the specific point assignment suggested by Borda. Accordingly, one might hope that "minor" changes in the points used would have little effect on the final rankings. Such a hope is in vain. Fishburn (1981) showed that given any two different scoring methods², a set of individuals and preferences can be found such that the rankings yielded by the two methods are exactly the opposite of each other. Saari (1984) extended this result as follows: given a set of $n \geq k + 1$ alternatives, k arbitrary rankings of these alternatives, and k distinct scoring methods, a set of individual preference ordering can be found such that all k of the rankings are generated using the different scoring methods³ (with the same individual preferences). This is true no matter how similar the scoring methods appear to be.

Though this finding indicates the extreme arbitrariness of scoring methods *in principal*, it does little to tell us if we are likely to encounter ranking reversals, i.e. changes in the rankings of alternatives, in practice for "reasonable" alterations in scoring methods. This type of question is notoriously difficult to answer. There is no clear way to place an a priori probability distribution over individual preferences⁴.

While I will not be able to definitively determine if ranking reversals are a serious practical problem, I hope to provide some insight by examining a situation for which considerable voting information is available – the selection of the Most Valuable Player awards in Major League Baseball.

Major League Baseball is the most important professional baseball organization in the United States of America. Currently, its primary component is 26 teams divided into two leagues, known as the National League and the American League. These teams enjoy immense popularity, many drawing over two million, and some over three million, paying spectators a year. In addition, the teams have lucrative television contracts.

At the end of each baseball season two players, one in the National League and one in the American League are honored by being named Most Valuable Player (MVP) in their respective leagues. Although the genesis of the Most Valuable Player award can be traced back to 1875, the "official" MVP trophy as we now know it has been given annually by the Baseball Writers Association of America (BBWAA) beginning in 1930. Since 1938 this honor has been determined as follows. Three writers in each city with a major league baseball team are polled (in 1961 the number was reduced to two). They are asked to list their top ten players. Each time a player appears first on a sportswriter's list, that

² The scoring methods must be truly different. Two methods which use weight vectors w and v , such that $w = sv + t(1, 1, \dots, 1)$ for some positive integers s and t , are equivalent and will always produce identical rankings.

³ Saari uses the term "positional voting method" for what is called a scoring method here.

⁴ Compare efforts to determine how often the Condorcet voting paradox is likely to arise.

player receives 14 points. For a second place listing a player receives 9 points. Third place is worth 8 points, fourth place is worth 7 points, and so forth down to tenth place, which is worth 1 point.

From 1938 to 1960 there were eight National League and eight American League teams, so that the maximum number of points a player could receive was 336. From 1961 to 1968 in the American League and from 1962 to 1968 in the National League the number of teams was increased to ten. In 1969 the number of teams in each league was increased to 12, and in 1977 the number of teams in the American League was again increased, this time to 14. Currently, a unanimous first place selection gives a National League player 336 points and an American League player 392 points⁵.

I examine MVP results for the years 1943 through 1989. Of these years, I am missing vote information for eight elections⁶, leaving eighty-six cases to analyze. Determining whether or not a change in scoring methods can cause a ranking reversal is a trivial task. Define r_{ik} as the number of people who rank player i at position k or higher. Player i can be ranked above player j by some scoring method if and only if $r_{ik} > r_{jk}$ for some k . Calculating the players' positions under a variety of scoring methods requires only the patience to type numbers into a spreadsheet program. All manipulations are carried out under the assumption that the sportswriters would not submit different rankings as the scoring methods changed.

The selection of a most valuable player involves many objective criteria, such as players' batting averages, number of home runs hit, number of runs batted in, and so forth.... For this reason we might expect to find considerable agreement in the rankings of the various sportswriters. Indeed, there are eight instances in which one player received all the first place votes, and 40 instances in which the first place votes were distributed among three players or less. This occurred although the number of players in each league ranged from two hundred to three hundred and fifty for the time period in question. It is doubtful that there are many voting situations with such widespread agreement among the voters. Since reversals tend to occur when there is a large variance in voter rankings, the presence of a significant number of reversals here would indicate that their possibility does indeed constitute a practical problem.

I first consider changes in the scoring system which could have resulted in the naming of a different MVP. In 24 out of 86 cases, a change in the scoring points could have had an effect on the player named most valuable. In one of these cases a two-way tie for first could have been broken. In two cases, the original winner would have been renamed but a tie for first could have been created. Nineteen times the second place finisher could have been named MVP instead of the actual winner; in five cases the third place finisher could also have been the winner. Somewhat oddly, in one instance the third place, but not the second place, finisher could have been named MVP and once the fourth place, but not the second or third place, finisher could have ended up first. Finally, in four cases the third place finisher, in one case the fourth place finisher, and in one

⁵ See Deane (1988) for a more complete history of the MVP, as well as Cy Young and Rookie of the Year awards. Contained therein is a complete listing of points received by players, but insufficient information for determining reversals. These reversals were determined from additional voting data supplied by Deane.

⁶ National League: 1946, 1948, 1949, 1950, 1951
American League: 1945, 1953, 1954

Table 1. Years subject to first place reversals. The numbers indicate the original finishing place of a player who could have finished first with a different scoring method. Numbers in brackets are for players who could have finished above the original winner, but not first

Year	American	National
1943	2	
1944	2	2
1946	2	
1947	2, 3 (5)	2 (3)
1951	2, 3 (3-way tie)	
1952		3
1955	2, 3	2 (4)
1957		2, 3
1958	4	
1960	2, 3	
1961	2	
1962		2
1963		2 (3)
1964		2
1966		2
1968		2
1969		2 (2-way tie)
1979		tie broken (3)
1981	2	
1983		2
1984	2 (3)	

case the fifth place finisher could have placed ahead of the original winner with a change in the scoring method, but would have been unable to overtake the second place finisher. I will have more to say about these peculiarities later.

My findings are displayed in Table 1, where the numbers in the columns indicate the ranking of the players who could have finished first. Numbers in brackets indicate players who could finish above the original first place winner, but only with a scoring system which would still have placed them behind the number 2 player. Note that no more than three players were ever in contention for the MVP title.

These changes are all for scoring methods that preserve a strictly decreasing sequence of points for player rankings that receive a positive point score⁷. In the present context a scoring system that, for instance, assigned the same score to a first and second place sportswriter ranking does not seem reasonable. Still, some of the alternate scoring systems needed to obtain reversals are extreme and are also not likely to be considered reasonable⁸. It is worth noting how many reversals are generated by a few simple scoring methods.

The scores $w_1 = 10$, $w_2 = 9 \dots$, $w_{10} = 1$, produce MVP winners different than the actual ones in eight elections. The scores $w_1 = 5$, $w_2 = 3$, $w_3 = 1$, which are the scores used in selecting the Cy Young pitcher, produce four differences. A simple plurality alters 9 of the elections. Many of these changes occur in the same elections; in fifteen different elections at least one of the above scoring methods would have had an effect on the winner. Of course, the number of times some "plausible" difference in scoring method would have mattered is greater still.

⁷ Also, $w_i = 0$ for $i > 10$ since the voting data only indicates the sportswriters' top ten players.

⁸ For instance, the 1958 American League reversal requires a scoring method with $w_9 > 1/2 w_1$

Table 2. 1957 National League MVP finishing using different scoring systems. Numbers in brackets indicate original finishes. A name in bold face indicates a player whose position has changed (including a tie being broken)

Players					
1st	Aaron	Schoendienst (3)	Musial (2)	Musial (2)	Musial (2)
2nd	Musial	Musial (2)	Aaron (1)	Aaron (1)	Aaron (1)
3rd	Schoendienst	Aaron (1)	Schoendienst (3)	Schoendienst (3)	Schoendienst (3)
4th	Mays	Mays (4)	Mays (4)	Mays (4)	Mays (4)
5th	Spahn	Spahn (5)	Spahn (5)	Spahn (5)	Spahn (5)
6th	Banks	Hodges (7)	Banks (6)	Banks (6)	Sanford (10)
7th	Hodges	Banks (6)	Hodges (7)	Hodges (7)	Banks (6)
8th	Mathews	Mathews (8)	Mathews (8)	Mathews (8)	Hodges (7)
9th	Robinson	Sanford (10)	Sanford (10)	Robinson (9)	Bouchee (12)
10th	Sanford	Robinson (9)	Robinson (9)	Sanford (10)	Hoak (11)
11th	Hoak	Hoak (11)	Hoak (11)	Hoak (11)	Mathews (8)
12th (tie)	Blasingame	Bouchee (12)	Blasingame (12)	Blasingame (12)	Robinson (9)
12th (tie)	Bouchee	Blasingame (12)	Bouchee (12)	Bouchee (12)	
14th	Buhl	Buhl (14)	Buhl (14)	Buhl (14)	
15th (tie)	Ennis	Ennis (15)	Ennis (15)	Ennis (15)	
15th (tie)	Groat	Groat (15)	Dark (17)	Groat (15)	
17th	Dark	Dark (17)	Groat (15)	Dark (17)	
18th	Snider	Snider (18)	Snider (18)	Snider (18)	
19th (tie)	Thomas	Drysdale (19)	Thomas (19)	Thomas (19)	
19th (tie)	Drysdale	Thomas (19)	Drysdale (19)	Drysdale (19)	
21st (tie)	McMillan	McMillan (21)	Drott (21)	McMillan (21)	
21st (tie)	Drott	Drott (21)	McMillan (21)	Drott (21)	
23rd	Hamner	Hamner (23)	Hamner (23)	Hamner (23)	
24th	Burdette	Burdette (24)	Burdette (24)	Burdette (24)	
25th (tie)	Logan	Logan (25)	Logan (25)	Logan (25)	
25th (tie)	Anderson	Anderson (25)	Anderson (25)	Anderson (25)	
Points					
w_1	140	155	155	100	100
w_2	90	105	105	90	70
w_3	80	65	80	80	50
w_4	70	50	60	70	30
w_5	60	45	50	60	10
w_6	50	42	40	50	0
w_7	40	39	30	40	0
w_8	30	35	20	30	0
w_9	20	20	10	20	0
w_{10}	10	10	5	10	0

Although being named the Most Valuable Player carries considerably more prestige than any lower place finish, these other positions are also important. Many players contracts carry provisions for bonuses which depend on where they finish in the MVP balloting. In addition, their future salaries are undoubtedly affected by where they find themselves in the voting. For this reason, as well as for pure social choice theory interest, it is important to look beyond reversals in the top finish. In every single election lower place finishers were subject to ranking reversals.

Let us examine this ranking reversal phenomenon in greater detail. Table 2 lists the finishing order for all players receiving at least one point in the 1957 National League MVP balloting using five different scoring methods (see Tables 3

Table 3. BBWAA most valuable player vote tally – 1957 National League

Name	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
1. Aaron, Braves	9	2	6	5	2	239
2. Musial, Cards	5	9	9	1	230
3. Schoendienst, Cards	8	7	3	1	.	1	1	2	.	.	221
4. Mays, Giants	1	5	4	7	4	2	174
5. Spahn, Braves	1	.	2	5	6	4	1	2	.	.	131
6. Banks, Cubs	.	.	.	2	2	2	2	.	6	4	60
7. Hodges, Dodgers	.	.	.	1	1	1	2	6	3	4	54
8. Mathews, Braves	2	1	4	3	1	1	45
9. Robinson, Redlegs	2	2	3	1	2	1	42
10. Sanford, Phils	.	1	.	1	.	2	1	2	1	1	39
11. Hoak, Redlegs	3	2	.	1	.	.	31
12. Blasingame, Cards	1	3	1	.	.	1	26
Bouchee, Phils	.	.	.	1	1	.	1	1	2	2	26
14. Buhl, Braves	2	.	1	1	.	15
15. Ennis, Cards	2	1	1	.	13
Groat, Pirates	1	.	3	3	13
17. Dark, Cards	1	.	1	2	.	13
18. Snider, Dodgers	2	.	.	2	10
19. Thomas, Pirates	1	.	1	.	.	8
Drysdale, Dodgers	2	.	.	.	8
21. McMillan, Redlegs	1	1	1	6
Drott, Cubs	1	.	1	.	6
23. Hamner, Phils	1	.	.	3
24. Burdette, Braves	2	2
25. Logan, Braves	1	1
Anderson, Phils	1	1

and 4 for the breakdown of the voting and point totals). The first column gives the official finishing order using (ten times) the scoring points actually in use by the Baseball Writer's Association. In the next four columns finishes are recomputed using different points.

The points used in the second column produce a crucial change – the order of the top three finishers is exactly reversed. Other movements are of no more than one position. The next column has less dramatic reversals at the top but uses “nicer looking” numbers. Note that in both these cases, the new point systems are in the spirit of the original points used. In particular, they preserve the feature that the point differential between a first and second place finish is greater than the differential between any other two adjacent finishes. The fourth column uses a simple constant differential for the top ten finishes. The fifth column has the sportswriters rank five rather than ten candidates, a quite plausible alternative.

In the fifth column, there are significant changes; in particular, the tenth place finisher jumps to sixth place, and the twelfth place finisher, Bouchee, moves into ninth place. However, although Bouchee has moved ahead of ninth place finisher Robinson, no change in the scoring method could have enabled him to overtake tenth place finisher Sanford⁹. This illustrates an important feature of scoring methods. The rankings depend not only on how well a candidate does “overall”, but also on just how the votes are distributed. We saw a similar peculiarity in Table 1, where in some cases third, fourth and fifth place finishers could overtake

⁹ This can easily be seen from voting Table 3.

Table 4. 1957 National League – Total scores received by players using different scoring methods

Aaron	2390	2335	2485	2030	1510
Musial	2300	2355	2500	2100	1610
Schoendienst	2210	2371	2385	1890	1470
Mays	1740	1554	1700	1700	900
Spahn	1310	1082	1145	1270	410
Banks	600	512	440	600	80
Hodges	540	525	380	540	40
Mathews	450	423	335	450	20
Robinson	420	376	315	420	20
Sanford	390	378	330	390	100
Hoak	310	254	250	310	30
Blasingame	260	220	205	260	10
Bouchee	260	229	190	260	40
Buhl	150	139	110	150	0
Ennis	130	133	90	130	0
Groat	130	129	75	130	0
Dark	120	117	80	120	0
Snider	100	98	70	100	0
Thomas	80	77	60	80	0
Drysdale	80	78	60	80	0
McMillan	60	65	35	60	0
Drott	60	59	40	60	0
Hamner	30	35	20	30	0
Burdette	20	20	10	20	0
Logan	10	10	5	10	0
Anderson	10	10	5	10	0
Points					
w_1	140	155	155	100	100
w_2	90	105	105	90	70
w_3	80	65	80	80	50
w_4	70	50	60	70	30
w_5	60	45	50	60	10
w_6	50	42	40	50	0
w_7	40	39	30	40	0
w_8	30	35	20	30	0
w_9	20	20	10	20	0
w_{10}	10	10	5	10	0

the MVP winner, but could not overtake the second place finisher, while other times the number two finisher could never win although lower place finishers could.

Reversals tend to occur when the voting is close, so that a large point differential is a good indication that no reversals are possible, and a slim margin of victory is a good indication that a reversal may be possible – a good indication, but far from a certainty.

Thus, in Dale Murphy's easy 1983 victory he received 318 points to Andre Dawson's 213 points and Mike Schmidt's 191 points. A look at the vote breakdown in Table 5 confirms the impression that Murphy was a solid winner. Indeed, his selection was nearly unanimous. Nonetheless, a different scoring method (albeit a significantly different one) could have resulted in the naming of Dawson as MVP instead.

Compare Mickey Mantle's 1957 MVP selection with a 24 point differential (233 points to Ted Williams' 209 points, and Sievers' 205). Although Mantle's

Table 5. The vote distribution for the top three finishers in selected years

1983 National League											
Player	1	2	3	4	5	6	7	8	9	10	Total
1. Murphy, D.	21	2	.	.	1	318
2. Dawson, A.	1	16	6	1	213
3. Schmidt, M.	1	5	9	7	1	1	191
1957 American League											
Player	1	2	3	4	5	6	7	8	9	10	Total
1. Mantle, M.	6	9	5	4	233
2. Williams, T.	5	5	7	5	1	1	209
3. Sievers, R.	4	6	6	1	6	.	1	.	.	.	205
1951 American League											
Player	1	2	3	4	5	6	7	8	9	10	Total
1. Berra, Y.	6	5	.	4	1	2	1	1	2	.	184
2. Garver, N.	6	2	3	2	1	.	1	1	.	4	157
3. Reynolds, A.	6	2	2	.	.	.	1	1	.	.	125

margin of victory was less than Murphy's, and the vote breakdown (Table 5) appears to make him a less convincing winner, no change in the scoring system could deprive him of his victory.

Perhaps even more dramatic is Yogi Berra's 1951 crowning (Table 5). Though a pure plurality would have resulted in a three way tie, no other scoring method could have deprived him of sole possession of the MVP title, despite the fact that the voting was quite close indeed.

Thus, the possibility, or lack thereof, of a first place reversal is not a good gauge of the decisiveness of a player's victory. Rather, reversals may simply reflect the particulars of the vote allocation.

As has been noted, for the period 1943–1989 there are 24 instances when a change in the scoring system would have resulted in a different MVP. Dividing this time period into two intervals (with 42 elections in one, 44 elections in the other¹⁰), reveals a striking fact. Of the 24 first place reversals, 18 occur from 1943–1967, whereas only 6 take place between 1968 and 1989.

This difference may be due to the increased media attention given to the selection of the Most Valuable Players in more recent years. The widespread press discussion of who should be selected as the National and American League MVPs in the months preceding the end of the baseball season results in somewhat of a consensus being reached by the end of the season, reducing the possibility of a first place reversal¹¹. This explanation is supported by the fact that of the forty times that three players or less have received all the first place votes, twenty six occur since 1968. Furthermore, from 1968 on all but three MVP winners have received at least five more first place votes than anyone else. Prior to 1968, seventeen MVPs had first place vote differentials of less than five.

¹⁰ Recall that we are missing data for eight elections.

¹¹ This explanation was suggested to me by Gary Cohen, radio announcer for the New York Mets.

Conclusion

From Condorcet's discovery of voting cycles through Arrow's impossibility theorem to more recent work on scoring methods, Social Choice theorists have shown that voting procedures may display anomalies. They have had somewhat less success, however, in determining how much of a problem these theoretical difficulties pose in practice. I have taken a modest step in that direction.

I have found that MVP voting displays significant sensitivity to the particular scoring points which happen to be in use. Furthermore, this sensitivity itself is dependent on the particular way in which votes happen to be distributed. Certainly, both these features of scoring methods could have been demonstrated more starkly with specially constructed examples. My interest, however, is in the fact that these theoretical possibilities have been realized.

Although not every instance of a ranking reversal occurs with a "reasonable" change in scoring methods, the findings suggest that the phenomenon of reversals is of practical concern. In any case, reversals may be of some concern to Mickey Mantle. As impressive as his three MVP titles were, he could just as easily have been a five-time winner without having scored a single extra run. All he needed was a change in the scoring system¹²!

References

1. De Borda JC (1781) Memoire sur les elections au scrutin, histoire de l'Academie Royale des Sciences, Paris
2. Deane B (1988) Award voting, Society for American Baseball Research
3. Fishburn P (1981) Inverted orders for monotone scoring rules. *Disc Appl Math* 3: 27-36
4. Saari DG (1984) The ultimate of chaos resulting from weighted voting systems. *Adv Appl Math* 5: 286-308
5. Saari DG (1990) The borda dictionary. *Soc Choice Welfare* 7: 279-317

¹² Mantle was a unanimous winner in 1956 and an overwhelming selection in 1962. As previously noted, he won less convincingly in 1957. In all three of his victories his title was immune to scoring changes. In 1960 and 1961 he finished a close second to Roger Maris; both finishes could have been reversed with different scoring points.