Booklet position effects, and two new statistics to gauge voter understanding of the need to rank candidates in preferential elections

Franz Ombler franz@franzo.com

1 Introduction

In 2004 the Single Transferable Vote (STV) method replaced plurality for the election of members of New Zealand's District Health Boards (DHBs) [1]. While being unable to assess ballot position effects due to unrecorded random ordering of candidates' names on each ballot paper this article demonstrates effects that may be explained by the order of candidates' names in an accompanying booklet of the candidates' profiles. Such effects undermine the intended benefits from randomly ordering candidates' names on ballot papers, but prove useful in questioning voter understanding of the need to rank candidates. Two new statistics are proposed to better gauge voter understanding of a preferential voting method: the percentage of plurality style informal ballots and a rank indifferent percentage.

2 The elections

Two elections are considered: the Canterbury DHB election and the Otago DHB election. In both cases seven candidates were to be elected. Ballot papers were sent to voters by post. The ballots for the DHBs were printed with candidates' names randomly ordered such that each ballot paper might be unique. An accompanying booklet with candidates' profiles listed the candidates alphabetically [2]. It seems likely that few candidates for the elections were previously known to voters and the election would seem relatively non-partisan. Voters were allowed to rank order any number of candidates and a ballot was deemed informal if there was no 'unique first preference' indicated on the ballot [3].

2.1 Canterbury

The Canterbury DHB election was run alongside other territorial elections including those for the Christchurch City Council mayor, ward councillors and Canterbury Regional Council. These other elections continued to use plurality, so the voter had to contend with two methods in their ballot papers. There were 29 candidates. Of 117,852 non-blank ballots, 8,986 (7.6%) were deemed 'informal' and removed from the count. Of these, 7,579 (84.0% of informal votes, or 6.4% of to-tal votes) marked all of the candidates for whom they voted as a first preference (either with a tick, or by writing '1'), presumably unaware of the need to rank candidates and thus voting as if it were a plurality election.

2.2 Otago

The Otago DHB election was run alongside territorial elections like those for Canterbury, but all elections were conducted using STV. There were 26 candidates. Of 65,389 non-blank ballots, 3,016 (4.6%) were deemed 'informal' and removed from the count. Of these, 1,315 (43.6% of informal votes, or 2.0% of total votes) marked candidates as if it were a plurality election.

As can be seen from the second-last row of Table 3.1, Canterbury DHB voters were over three times more likely to waste their vote by treating the election as a plurality election (6.4% versus 2.0%). This is probably because the Otago DHB election voters were more familiar with STV due to its use for all the elections on the Otago ballot papers. To better gauge voter understanding of preferential elections the percentage of plurality style informal ballots could be reported alongside the more usually reported total number of informal ballots.

	Canterbury	Otago
Number of seats	7	7
Candidates	29	26
Non-blank ballots	117,852	65,389
Formal ballots	108,866	62,373
Informal ballots	8,986 (7.6%)	3,016 (4.6%)
Informal ballots with multiple first preferences only (plurality-style)	7,579 (6.4%)	1,315 (2.0%)
Rank indifferent (see below)	5.1%	2.9%

Table 3.1: The Canterbury and Otago DHB elections

3 Ballot position effects

The voter burden of ordering the candidates is higher when the candidates are unfamiliar to voters, when there are so many candidates (29 for Canterbury, 26 for Otago), and where the district magnitude is high (seven) [4]. Furthermore, due to the lack of familiarity with candidates, position effects are probably greater [4], and these effects have greater consequences when voters are required to rank order candidates [5]. These effects may also be expected to be amplified by voters' lack of experience in rank ordering candidates, especially when they have to contend with multiple methods on their ballot papers as in the Canterbury election.

Candidates' names were randomly ordered during ballot paper printing, presumably to prevent ballot position effects, that is, where the positions of the candidates' names on the ballot affect voters' selection or ranking of the candidates. Randomising candidate name order should certainly have reduced the effect of 'donkey votes': ballots in which the voter ranked all the candidates in the order in which they appeared on the ballot. However, the number of donkey votes cannot be assessed due to the absence of information as to the order in which the candidates were listed on each ballot sheet. For the same reason, other ballot position effects cannot be assessed either.

4 'Booklet position effects'

Due to voters' lack of familiarity with the candidates many voters would have relied heavily on the booklet of candidates' profiles to draft their selections and rankings. The booklet listed the candidates alphabetically. We might call ensuing effects 'booklet position effects', which will dilute the intended benefits from randomly ordering the candidates' names on ballot papers; indeed it is interesting to consider (although not demonstrated here) whether booklet position effects may be greater than ballot position effects in elections in which voters are less familiar with the candidates. Certainly, the cost-effectiveness of randomising ballot paper candidate name order is questionable if the order of candidates' profiles in an accompanying booklet is not also randomised.

Assigning the candidates numbers according to their positions in the booklet (alphabetically) helps compare the rankings of candidates on each ballot with the order in which they appear in the booklet. The real ballot '2 10 14 17 19 24 26', where this voter has ranked candidate number 2 first (that is, they wrote the number one beside the candidate who appeared second in the booklet), candidate number 10 second and so on, may be described as perfectly ordered as it lists the candidates in the same order in which they appeared in the booklet. Similarly, '9 6 14 19 21 24 27' seems near perfectly ordered.

Spearman's rank correlation coefficient (r_s) may be used to assess the correlation of two rankings. We can apply this to each ballot, finding the r_s of the rankings of candidates in the ballot and the same ballot with candidates re-ordered alphabetically. For example, the r_s of the ballot '2 10 14 17 19 24 26' with its ordered self (the same ballot) is exactly 1.0, showing a perfect positive correlation; while r_s for '9 6 14 19 21 24 27' and its ordered self ('6 9 14 19 21 24 27') is 0.96.

The average r_s of each formal ballot's ranking of candidates with its ordered self is only 0.06 for Canterbury and 0.03 for Otago, showing such weak positive correlations that one might be tempted to infer an absence of booklet position effects. This is likely to draw criticism that it proves nothing due to 'failure to randomly assign groups of voters to different name orders' [4]. Indeed it would be consistent with this bare analysis to claim that position effects were present to a large degree and that if the booklets had been printed randomly that we would have seen a lower average r_s . This might be true to some extent but we are unable to assess it properly due to the absence of information about the order of names on each ballot; however, even without this information, booklet position effects can be demonstrated.

If we assess the frequency of the various values of r_s for the ballots, we find inordinately high numbers of perfectly ordered and near perfectly ordered ballots. Figure 3.1 (the data for which is presented in Table 3.2) shows such an analysis of the 51,730 ballots that listed exactly seven candidates in the Canterbury

election. In light grey is the exact distribution of r_s for N=7, as would be approximated by randomly ordering these same ballots. Clearly there is a heavy tail on the right for the real ballots. Focussing on the rightmost bar, these 1,286 ballots (2.49%) are listed perfectly in order, but the expected number of ballots to be found in order for these 51,730 voters is only ten (0.02%) if preferences are randomly distributed.

Analyses of ballots listing other numbers of candidates (but more than 1) also find a notably higher than expected number of perfectly ordered ballots, 2,962 more than expected in total (see Table 3.3).

The Otago DHB election shows a similar but less prominent pattern (Figure 3.2). Given the similarity of the elections in other respects, this difference might be best explained by the use of STV in all of the elections on the Otago ballot papers and therefore greater voter awareness and understanding of the method.

Booklet position effects are apparent, but there are other potential explanations. It is conceivable that some voters are strongly biased towards candidates whose names start with letters nearer the beginning of the alphabet and admittedly booklet position effects cannot be distinguished from alphabetic effects in this election [4]. It is also possible that a group of candidates may actually be preferred in alphabetical order, perhaps by a small group of voters, perhaps following how-to-vote cards with candidates ordered alphabetically. However, as discussed above, the Canterbury voters would have been less aware of STV, they were more than three times more likely than Otago voters to vote as if the election were being run as a plurality election, and the charts show a greater percentage of perfectly ordered ballots for the Canterbury election. I contend that the charts' heavy tails primarily demonstrate ignorance of, or indifference towards, the ranking of candidates.

5 A measure of voter indifference to ranking

Where booklet or ballot position order can be assessed it may be worthwhile reporting a 'rank indifferent' statistic alongside the percentage of informal votes usually reported in elections. However, it isn't easy to say how many voters are rank indifferent.

Considering the Canterbury DHB election, it certainly seems reasonable to assert that most of the 1,286 voters who listed seven candidates in perfect order were rank indifferent: all but the ten expected, perhaps (refer Table 3.3). It would also seem true of the remaining 151 who listed more than seven candidates in perfect order, as the probability of this occurring is so low. It is less compelling to argue that 38 of the 2,526 voters who listed only two candidates in perfect order should also count, as the probability of this occurring by chance is so much greater. The appropriateness of this measure would then depend on some aspects of the election: if the number of candidates is low or if there are few candidates with popular support, sincere preferences are far more likely to happen to accord with ballot or booklet position and this may result in an inordinate number of perfectly ordered or near perfectly ordered ballots.

One way to avoid this problem is to count the higher than expected number of ordered ballots only when the probability of this occurring is extremely low, below 1% perhaps, which would only assess ballots listing five or more candidates. The Canterbury DHB election would then have a statistic of 2%. However, this seems conservative given the significantly more than expected number of near perfectly ordered ballots shown in the second-to-rightmost bar in Figure 3.1. Therefore one might also consider those ballots with an r_s , such that, say, less than 1% of ballots are to be expected to be found with this r_s or higher. The appropriate choice of r_s will then depend on the number of candidates in the ballot.

Taking this approach encapsulates the above in which we ignored ballots with less than five candidates, as with fewer than five candidates, there are fewer possible values of r_s and the probability of finding ordered ballots is greater than 1%. For example, where a ballot ranks only two candidates, there are only two possible arrangements resulting in an r_s (with its ordered self) of either 1 or -1, and with a probability of 50% either way. With three candidates there are only four possible values of $r_s: -1, -0.5, 0.5$ and 1, and the expected number of ballots having an r_s of 1 is one in six (16.7%) [6]. For four candidates, the expected number of ballots with an r_s of 1 is 4%. It is not until we reach five candidates that the expected number of ballots with an r_s of 1 drops below 1%. For six candidates, the expected number of ballots with an $r_s \ge 0.94$ (an r_s of either 0.94 or 1) is less than 1%, so we now count near perfectly ordered ballots as well as perfectly ordered ballots.

The appropriate values to use for r_s are thus the critical values to be found tabulated in textbooks. The expected number of ballots can be calculated from the probability of an r_s greater than or equal to the critical value: this might be assumed to be 1%, but it varies due to the discrete nature of r_s . Thus we also need to look up the probability of this value of r_s and calculate the number of ballots that may be expected to have this r_s if the ballots were randomly ordered. Critical values for the number of candidates in the ballot from 5 through 50 and the probabilities of finding these values are listed in Table 3.4.

Thus one can step through each ballot that ranks five or more candidates, correlating the ballot with its ordered self, and counting those that are 'highly ordered', that is, those with an r_s greater than the critical value for its number of candidates. One can then subtract the expected number of highly ordered ballots, which can be simply calculated by counting the number of ballots with each number of candidates and multiplying this by the probabilities listed in Table 3.4. Dividing this difference by the total number of formal ballots provides an accessible statistic. This statistic may be interpreted as the percentage of voters that were almost certainly rank indifferent. For the Canterbury DHB election this is 3.8% and for the Otago DHB election it is 1.9%.

However, the probability of a voter being rank indifferent can be expected to be unrelated to the length of the ballot even though we cannot identify rank indifference in shorter ballots with confidence. This seems reasonable when one considers that there is no reason to believe that voters who ranked fewer candidates might have had any greater understanding of STV than those who listed five or more candidates. Therefore, we should really divide the difference by the number of formal ballots that listed five or more candidates. For the Canterbury DHB election the rank indifferent statistic is then 5.1% and for the Otago DHB election it is 2.9% (see Table 3.5 for working).

6 Conclusions

Booklet position effects should be considered when assessing the cost-effectiveness of randomising the order of candidates' names on the ballot paper, especially if voters are unfamiliar with the candidates or if the need to rank candidates might be poorly understood.

Two new statistics may be reported to better gauge voter understanding of preferential voting: first, the percentage of plurality-style informal ballots, that is, ballots in which the voter marked all of the candidates (for whom they voted) with a tick or a '1'; and second, for elections where voters might be expected to rank order five or more candidates, the percentage of voters that were almost certainly rank indifferent. However, in interpreting the rank indifferent percentage one should be wary of other potential causes of perfectly ordered or near perfectly ordered ballots such as how-to-vote cards.

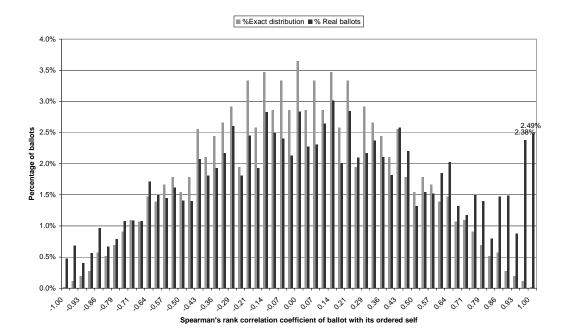


Figure 3.1: Canterbury DHB: frequency of ballots for Spearman rank-order correlation coefficients of voters' ballots with their ballots ordered alphabetically, for ballots listing seven candidates.

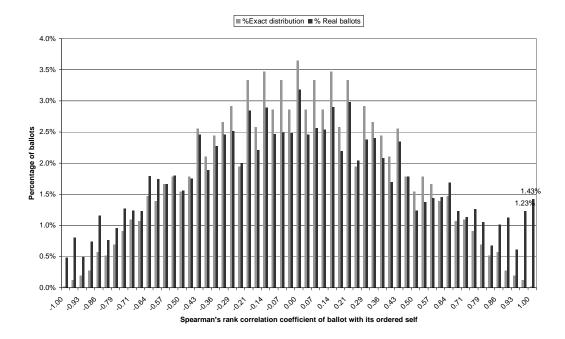


Figure 3.2: Otago DHB: frequency of ballots for Spearman rank-order correlation coefficients of voters' ballots with their ballots ordered alphabetically, for ballots listing seven candidates.

	Canterbury	/ DHB da	ata			Otago D	Otago DHB Data	
rman's rank prrelation pefficient	%Exact distribution	Real ballots	% Real ballots	Exact distribution	Spearman's rank correlation coefficient	%Exact distribution	Real ballot	
-1.00	0.02%	249	0.48%	10	-1.00	0.02%	12	
-0.96		357	0.69%		-0.96	0.12%	20	
-0.93		211	0.41%	103	-0.93	0.20%	12	
-0.89		292	0.56%		-0.89	0.28%	18	
-0.86	0.58%	501	0.97%	298	-0.86	0.58%	29	
-0.82	0.52%	346	0.67%	267	-0.82	0.52%	194	
-0.79	0.69%	406	0.78%	359	-0.79	0.69%	244	
-0.75	0.91%	559	1.08%	472	-0.75	0.91%	322	
-0.71	1.09%	564	1.09%		-0.71	1.09%	314	
-0.68		557	1.08%		-0.68	1.07%	313	
-0.64		885	1.71%		-0.64	1.47%	455	
-0.61	1.39%	778	1.50%		-0.61	1.39%	443	
-0.57	1.67%	749	1.45%		-0.57	1.67%	423	
-0.54		836	1.62%		-0.54	1.79%	457	
-0.50		729	1.41%		-0.50	1.55%	397	
-0.46		726	1.40%		-0.46	1.79%	445	
-0.43		1072	2.07%	1324	-0.43	2.56%	625	
-0.39		935	1.81%		-0.39	2.10%	479	
-0.36		1000	1.93%		-0.36	2.44%	578	
-0.32		1123 1348	2.17%		-0.32	2.66%	625 638	
-0.29		936	1.81%		-0.29	1.94%	507	
-0.23	3.33%	1268	2.45%	1724	-0.25	3.33%	723	
-0.21		999	1.93%		-0.21	2.58%	562	
-0.14		1463	2.83%		-0.14	3.47%	734	
-0.11	2.86%	1293	2.50%		-0.11	2.86%	627	
-0.07	3.33%	1245	2.41%		-0.07	3.33%	632	
-0.04		1103	2.13%		-0.04	2.86%	631	
0.00		1465	2.83%		0.00	3.65%	808	
0.04		1179	2.28%		0.04	2.86%	624	
0.07	3.33%	1196	2.31%	1724	0.07	3.33%	652	
0.11	2.86%	1367	2.64%		0.11	2.86%	645	
0.14		1559	3.01%		0.14	3.47%	737	
0.18		1041	2.01%		0.18	2.58%	557	
0.21	3.33%	1473	2.85%		0.21	3.33%	757	
0.25		1086	2.10%		0.25	1.94%	518	
0.29		1124	2.17%		0.29	2.92%	604	
0.32		1227	2.37%		0.32	2.66%	609	
0.36		1090 939	2.11%		0.36	2.44%	528 431	
0.39		1333	2.58%		0.39	2.10%	595	
0.43		1141	2.38%		0.43	1.79%	453	
0.40		681	1.32%		0.40	1.75%	315	
0.54		799	1.54%		0.54	1.79%	350	
0.57	1.67%	788	1.52%		0.57	1.67%	365	
0.61	1.39%	957	1.85%		0.61	1.39%	369	
0.64		1048	2.03%		0.64	1.47%	429	
0.68		682	1.32%		0.68	1.07%	313	
0.71	1.09%	606	1.17%		0.71	1.09%	288	
0.75	0.91%	779	1.51%	472	0.75	0.91%	321	
0.79		724	1.40%		0.79	0.69%	269	
0.82		412	0.80%	267	0.82	0.52%	172	
0.86		761	1.47%	298	0.86	0.58%	257	
0.89		771	1.49%		0.89	0.28%	287	
0.93		453	0.88%		0.93	0.20%	156	
0.96		1233	2.38%		0.96	0.12%	313	
1 00	0.02%	1286	2.49%		1.00	0.02%	362	

Table 3.2: Data for Figures 3.1 and 3.2: the numbers of ballots for each possible value of r_s and the exact distribution (as would be approximated by randomly ordered ballots) for ballots ranking seven candidates [7]

Exact

71

373

741

655

675

534 650

lots distribution

Candidates in ballot (<i>n</i>)	Ballots (b)	Perfectly ordered (p)	Probability of being in order (1/ <i>n</i> !)	Expected ballots in order (<i>b/n</i> !)	% found in order (p/b)	Number of times more than expected (<i>p/(b/n</i> !))
1	5691	5691	1.000000	5691	100.00%	1.00
2	4977	2526	0.500000	2489	50.75%	1.02
3	8483	1766	0.166667	1414	20.82%	1.25
4	8030	817	0.041667	335	10.17%	2.44
5	8639	514	0.008333	72	5.95%	7.14
6	5857	229	0.001389	8	3.91%	28.15
7	51730	1286	0.000198	10	2.49%	125.29
8	3331	55	0.000025	8.3E-02	1.65%	665.75
9	2224	39	2.8E-06	6.1E-03	1.75%	6363.45
10	2721	27	2.8E-07	7.5E-04	0.99%	36007.94
11	1107	12	2.5E-08	2.8E-05	1.08%	4.33E+05
12	1170	3	2.1E-09	2.4E-06	0.26%	1.23E+06
13	503	6	1.6E-10	8.1E-08	1.19%	7.43E+07
14	507	4	1.1E-11	5.8E-09	0.79%	6.88E+08
15	361	1	7.6E-13	2.8E-10	0.28%	3.62E+09
16	294	0	4.8E-14	1.4E-11	0.00%	0.00
17	166	2	2.8E-15	4.7E-13	1.20%	4.29E+12
18	131	0	1.6E-16	2.0E-14	0.00%	0.00
19	91	0	8.2E-18	7.5E-16	0.00%	0.00
20	112	0	4.1E-19	4.6E-17	0.00%	0.00
21	68	0	2.0E-20	1.3E-18	0.00%	0.00
22	50	0	8.9E-22	4.4E-20	0.00%	0.00
23	37	0	3.9E-23	1.4E-21	0.00%	0.00
24	47	0	1.6E-24	7.6E-23	0.00%	0.00
25	33	0	6.4E-26	2.1E-24	0.00%	0.00
26	49	0	2.5E-27	1.2E-25	0.00%	0.00
27	45	0	9.2E-29	4.1E-27	0.00%	0.00
28	47	0	3.3E-30	1.5E-28	0.00%	0.00
29	2365	2	1.1E-31	2.7E-28	0.08%	7.48E+27

Table 3.3: Perfectly ordered ballots in the Canterbury DHB election

Number of		Probability
candidates	Minimum	of finding
selected on	rs	such a
ballot		ballot
5	1.000	0.00833
6	0.943	0.00833
7	0.893	0.00615
8	0.833	0.00769
9	0.783	0.00861
10	0.745	0.00870
11	0.709	0.00910
12	0.678	0.00926
13	0.648	0.00971
14	0.626	0.00953
15	0.604	0.00973
16	0.582	0.00999
17	0.566	0.00983
18	0.550	0.00986
19	0.535	0.01
20	0.520	0.01
21	0.508	0.01
22	0.496	0.01
23	0.486	0.01
24	0.476	0.01
25	0.466	0.01
26	0.457	0.01
27	0.448	0.01

Number of candidates	Minimum	Probability of finding
selected on	rs	such a
ballot		ballot
28	0.440	0.01
29	0.433	0.01
30	0.425	0.01
31	0.418	0.01
32	0.412	0.01
33	0.405	0.01
34	0.399	0.01
35	0.394	0.01
36	0.388	0.01
37	0.383	0.01
38	0.378	0.01
39	0.373	0.01
40	0.368	0.01
41	0.364	0.01
42	0.359	0.01
43	0.355	0.01
44	0.351	0.01
45	0.347	0.01
46	0.343	0.01
47	0.340	0.01
48	0.336	0.01
49	0.333	0.01
50	0.329	0.01

Table 3.4: Critical values and probabilities for r_s

[6, 7]

		Canterbury DHB					Otago DHB		
Candidates in ballot (<i>n</i>)	Ballots	Expected highly ordered	Found highly ordered	Difference	Candidates in ballot (<i>n</i>)	Ballots	Expected highly ordered	Found highly ordered	Difference
1	5691				1	4323			
2	4977				2	4196			
3	8483				3	6047			
4	8030				4	5515			
5	8639	72.0	514	442.0	5	4573	38.1	157	118.9
6	5857	48.8	229	180.2	6	4100	34.2	86	51.8
7	51730	318.1	2972	2653.9	7	25389	156.1	831	674.9
8	3331	25.6	237	211.4	8	1905	14.6	115	100.4
9	2224	19.1	169	149.9	9	1112	9.6	62	52.4
10	2721	23.7	222	198.3	10	1470	12.8	90	77.2
11	1107	10.1	81	70.9	11	502	4.6	25	20.4
12	1170	10.8	78	67.2	12	577	5.3	26	20.
13	503	4.9	45	40.1	13	225	2.2	9	6.8
14	507	4.8	33	28.2	14	282	2.7	10	7.
15	361	3.5	30	26.5	15	148	1.4	7	5.
16	294	2.9	27	24.1	16	117	1.2	8	6.8
17	166	1.6	12	10.4	17	54	0.5	1	0.5
18	131	1.3	8	6.7	18	56	0.6	3	2.4
19	91	0.9	6	5.1	19	34	0.3	1	0.
20	112	1.1	4	2.9	20	56	0.6	2	1.
21	68	0.7	3	2.3	21	23	0.2	1	0.
22	50	0.5	5	4.5	22	26	0.3	0	-0.
23	37	0.4	2	1.6	23	11	0.1	1	0.
24	47	0.5	5	4.5	24	21	0.2	2	1.
25	33	0.3	4	3.7	25	81	0.8	1	0.
26	49	0.5	3	2.5	26	1530	15.3	79	63.
27	45	0.5	3	2.6					
28	47	0.5	0	-0.5					
29	2365	23.7	76	52.4					
	-	576.8	4768	4191.2		-	301.7	1517	1215.3
Fotal	108866	Rank i	ndifferent $n \ge 1$	3.8%	Total	62373	Rank i	ndifferent $n \ge 1$	1.9%
「otal <i>n</i> >= 5	81685	Rank i	ndifferent $n \ge 5$	5.1%	Total $n \ge 5$	42292	Rank i	ndifferent $n \ge 5$	2.9%

Table 3.5: Manual calculation of rank indifferent statistic

Further information and computer programs to automate the production of these statistics are available from the author on request.

7 References

- [1] Department of Internal Affairs, 'STV Information', 2004, URL on web site.
- [2] Personal communication with Christchurch City Council, 13 April 2005.
- [3] New Zealand Government, 'Part 4 Conduct of Elections and Polls using Single Transferable Voting Electoral System', *Local Electoral Amendment Regulations 2002.*
- [4] Jon A. Krosnik, Joanne M. Miller, and Michael P. Tichy, 'An unrecognized need for ballot reform', in Ann N. Crigler, Marion R. Just, and Edward J. McCaffery (eds.), *Rethinking the Vote : The Politics and Prospects of American Election Reform* (New York: Oxford University Press, 2004), pp. 52, 53, 63.
- [5] Susan Banducci, Michael Thrasher, Colin Rallings and Jeffrey A. Karp, 'Candidate appearance cues in low-information elections', 2003, Paper presented at the Annual Conference of the American Political Science Association (Philadelphia), URL on web site. p. 14.
- [6] Sidney Siegel and N. John Castellan, Jr., *Nonparametric statistics for the behavioral sciences*, second edition (New York: McGraw-Hill, 1988), pp. 242, 360-361.
- [7] Mark van de Wiel, Computer program spearman.c for calculating the 'Exact distribution of Spearman's rank statistic', available from URL on web site.